

Dipartimento di Fisica e Astronomia
"Galileo Galilei"

Artificial Intelligence-Based Class Assessment and Rare Sources Screening in *Fermi* Blazars

Supervisore: Prof. Denis Bastieri
Co-supervisore: Prof. Junhui Fan
Coordinator: Prof. Elisa Bernardini

Dottorando: Jingtian Zhu

Sommario

I Nuclei Galattici Attivi (AGN) sono tra le entità più affascinanti e dinamiche del cosmo. Queste meraviglie astronomiche, caratterizzate dalle loro intense emissioni elettromagnetiche in un ampio intervallo di energia (dalle bande radio ai raggi γ), sono centrali per la nostra comprensione dell'evoluzione galattica. All'interno delle diverse sottoclassi di AGN, i blazar emergono come un sottoinsieme particolarmente estremo e affascinante. I loro attributi distintivi, come la rapida variabilità e la marcata polarizzazione, li rendono indispensabili per le indagini astrofisiche.

Lo scopo principale di questa tesi è duplice: in primo luogo, sondare le complessità e le sfide associate alla classificazione binaria o al raggruppamento in blazar e, in secondo luogo, sfruttare queste tecniche per individuare i blazar che presentano proprietà uniche. Il capitolo di apertura pone le basi, fornendo una panoramica completa degli AGN, approfondendo i modelli di classificazione e unificazione degli AGN e mettendo in luce il processo di identificazione dei blazar con attributi fisici rari, contrassegnandoli come la sottoclasse AGN più estrema. Il capitolo 2 sposta l'attenzione sui telescopi e rivelatori all'avanguardia che danno forma all'astrofisica delle alte energie, con un' enfasi particolare sui ruoli chiave svolti da Fermi-LAT e dagli IACT terrestri come H.E.S.S. e MAGIC. Vengono inoltre discusse le potenzialità imminenti del CTAO, dei rivelatori tipo EAS e la precisione delle tecniche VLBI. Il capitolo 3 approfondisce le metodologie adottate, con un focus speciale sul potere trasformativo dell'intelligenza artificiale (AI) nella ricerca astrofisica. La capacità dell'intelligenza artificiale di elaborare vasti set di dati, l'impareggiabile precisione nel riconoscimento dei modelli e l'adattabilità all'evoluzione della ricerca ne sottolineano l'importanza nell'astrofisica moderna. I capitoli dal 4 al 6 offrono un'analisi approfondita di quattro studi innovativi che non solo contribuiscono all'ampio contesto della classificazione binaria o del clustering, ma permettono anche di comprendere meglio i blazar rari.

1. Il Capitolo 4 analizza il dibattito che persiste da decenni riguardo la dicotomia radio nei nuclei galattici attivi (AGN). Per risolvere questo problema abbiamo analizzato gli AGN nella banda ottica B e i dati nel radio a 6 cm. Il nostro studio sull'intensità nel radio, rappresentato come $\log R$, ha identificato una chiara divisione in $\log R = \langle 1.37 \pm 0.02 \rangle$, distinguendo gli AGN "loud" (RL) da quelli "quiet" (RQ), suggerendo una dicotomia RL/RQ. Proponiamo un duplice criterio per la classificazione degli AGN, integrando la luminosità e l'intensità radio. Questo approccio, utilizzando l'apprendimento automatizzato, è definito dalla relazione $\log L_{6\text{cm}} = -2.7 \log R + 44.3$. Al centro del dibattito RL/RQ c'è l'origine delle loro emissioni radio. I nostri risultati suggeriscono che RL e RQ condividono fonti di emissione simili, come jet e mini-jet, il che è confermato dalla correlazione tra luminosità radio 6 cm e banda B.
2. Il capitolo 5 approfondisce i jet dei blazar tramite le osservazioni di VLBI. La nostra ricerca su 407 VLBI blazar visti da *Fermi* (VFB) rivela una correlazione tra la luminosità dei raggi γ ($\log L_\gamma$) e la velocità apparente ($\log \beta_{\text{rmap}}$), suggerendo l'esistenza di un legame tra velocità del nodo nel jet e potenza del jet stesso. Correlazioni tra $\log \beta_{\text{app}}^{\text{max}}$, radiazione del getto ($\log L_{\text{rad}}$) e luminosità del getto a 5 GHz ($\log L_{5\text{GHz}}^{\text{ext}}$) suggeriscono il collegamento del movimento dei nodi all'intensità della radiazione e alla potenza cinetica. Tuttavia, questo è assente per i quasar radio a spettro piatto (FSRQ). Inoltre, la correlazione di $\log \beta_{\text{app}}^{\text{max}}$ con la luminosità del disco di accrescimento ($\log L_{\text{Disk}}$) enfatizza il ruolo dell'accrescimento nella dinamica dei nodi. Utilizzando modelli di miscela gaussiana, abbiamo identificato 228 potenziali candidati VFB utilizzando criteri quali $\log L_\gamma > 45.40$, $\alpha_{\text{ph}} > 224$ e $\log VI > 1.71$.
3. Il Capitolo 6 utilizza alcune tecniche di apprendimento automatizzato e supervisionato ed, in

particolare, la regressione logistica (LR), dove $LR(y) = \frac{1}{1+e^{-(\beta_0+\beta_1x)}}$, per identificare potenziali candidati blazar nel TeV. Applicando LR a cataloghi come 4FGL-DR2 / 4LAC-DR2, 3FHL, 3HSP e 2BIGB, abbiamo distinto blazar TeV dalle controparti non TeV in base ad alcune caratteristiche selezionate. Il modello LR assegna a ciascuna sorgente la probabilità di emettere nel TeV. Utilizzando una soglia dell'80%, abbiamo estratto dal 4FGL-DR2 / 4LAC-DR2 40 candidati TeV ad alta confidenza. L'analisi a banda larga delle distribuzioni spettrali di energia (SED) suggerisce che 7 di questi sono potenzialmente rivelabili dagli attuali osservatori IACT, ed uno potrebbe essere osservabile anche dagli EAS.

4. Il Capitolo 7 indaga le origini dei raggi cosmici di alta energia, approfondendo il modello che fa dei blazar potenziali emettitori di neutrini. Utilizzando il transfer learning (TL), questo studio analizza il catalogo *Fermi*-LAT Gamma-ray Source (4FGL-DR3) per identificare le specifiche che concorrono all'emissione di neutrini da parte dei blazar, superando l'apparente impasse dello squilibrio nella numerosità delle classi nel clustering. Tramite una rete neurale (ANN) abbiamo definito uno spazio parametrico specifico, distinguendo tra blazar che emettono neutrini (NB) ed entità non-NB, individuando i candidati NB entro un intervallo di confidenza all'99%. Questo approccio ha identificato con successo 273 candidati NB dal catalogo 4FGL-DR3, permettendoci di comprendere meglio l'emissione di neutrini da parte dei blazar e aiutandoci a rimodellare i meccanismi di produzione dei raggi cosmici.

Il capitolo 8 conclude la dissertazione, fornendo una sintesi dell'attuale utilizzo dell'intelligenza artificiale nell'ambito della ricerca astronomica, con un'enfasi particolare sulle sue implicazioni per gli studi sui blazar. Il capitolo postula inoltre le possibili strade per un'integrazione più profonda ed espansiva delle metodologie di intelligenza artificiale nel far progredire la nostra comprensione della fenomenologia dei blazar.

Abstract

Active Galactic Nuclei (AGN) are among the most captivating and dynamic entities in the cosmos. These astronomical marvels, marked by their intense electromagnetic emissions in broad energy range (from radio to γ -ray bands), are central to our comprehension of galactic evolution. Within the diverse subclasses of AGN, blazars emerge as an especially extreme and fascinating subset. Their distinctive attributes, such as rapid variability and pronounced polarization, render them indispensable for astrophysical inquiries.

The primary aim of this dissertation is twofold: first, to probe the intricacies and challenges associated with binary classification or clustering in blazars, and second, to leverage these techniques to pinpoint blazars exhibiting unique properties. The opening chapter lays the groundwork, providing a comprehensive overview of AGN, delving into the classification and unification models of AGN, and spotlighting the identification process for blazars with rare physical attributes, marking them as the most extreme AGN subclass. Ch. 2 is allocated to the comprehensive study of the *Fermi* Gamma-ray Space Telescope with its significant catalogs like 3FHL and 4FGL series. This chapter will elucidate the telescope's instrumentation, mission objectives, and the wealth of discoveries made possible by its observations. Ch. 3 delves into the methodologies adopted, with a special highlight on the transformative power of Artificial Intelligence (AI) in astrophysical research. AI's capability to process vast datasets, its unparalleled accuracy in pattern recognition, and its adaptability in evolving research landscapes underscore its significance in modern astrophysics. Ch. 4 through Ch. 7 offer an in-depth analysis of four groundbreaking studies. Each of these studies not only contributes to the broader narrative of binary classification or clustering but also advances our understanding of rare blazars.

1. Ch. 4 investigates the longstanding debate surrounding the radio dichotomy in AGNs that has persisted for decades. We analyzed AGNs with optical B band and radio 6 cm wavelength data to address this. Our study of radio loudness, represented as $\log R$, identified a clear division at $\log R = \langle 1.37 \pm 0.02 \rangle$, distinguishing radio-loud (RL) from radio-quiet (RQ) AGNs, hinting at an RL/RQ dichotomy. We propose a dual criterion for AGN classification, integrating radio luminosity and loudness. This approach, derived using machine learning, is defined by the relation $\log L_{6\text{cm}} = -2.7 \log R + 44.3$. Central to the RL/RQ debate is the origin of their radio emissions. Our findings suggest RLs and RQs share similar emission sources, such as jets and mini-jets, corroborated by correlating radio 6 cm and optical B-band luminosities.
2. Ch. 5 delves into blazar jets via the VLBI technique, our research on 407 VLBI detected *Fermi* blazars (VFBs) uncovers a correlation between γ -ray luminosity ($\log L_\gamma$) and apparent velocity ($\log \beta_{\text{app}}$), hinting at the tie of jet knot motion to jet power. Correlations between $\log \beta_{\text{app}}^{\text{max}}$, jet radiation ($\log L_{\text{rad}}$), and jet luminosity at 5 GHz ($\log L_{5\text{GHz}}^{\text{ext}}$) suggest knot motion's link to radiation and kinetic power. However, this is absent for flat-spectrum radio quasars (FSRQs). Additionally, the correlation of $\log \beta_{\text{app}}^{\text{max}}$ with accretion disc luminosity ($\log L_{\text{Disk}}$) emphasizes accretion's role in knot dynamics. Using Gaussian mixture models, criteria including $\log L_\gamma > 45.40$, $\alpha_{\text{ph}} > 2.24$, and $\log VI > 1.71$ identified 228 potential VFB candidates.
3. Ch. 6 utilizes a supervised machine learning technique, specifically Logistic Regression (LR), where $LR(y) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x)}}$, to identify potential TeV blazar candidates. By applying LR to catalogs such as 4FGL-DR2 / 4LAC-DR2, 3FHL, 3HSP, and 2BIGB, we aimed to distinguish TeV blazars from non-TeV counterparts based on selected features. The LR model assigns a probability to each source, indicating its likelihood of being a TeV blazar. Using a threshold of 80%, we identified 40 high-confidence TeV candidates from the 4FGL-DR2 / 4LAC-DR2

blazars. Broadband spectral energy distributions (SED) analysis of these candidates revealed that 7 are probable detections for current IACT observatories, with one suitable for EAS particle detector arrays.

4. Ch. 7 probes the origins of high-energy cosmic rays and has directed attention to blazars as potential neutrino emitters. Utilizing transfer learning (TL), this study delves into the *Fermi*-LAT Gamma-ray Source Catalog (4FGL-DR3) to identify blazars as neutrino emission candidates, addressing the challenge of class imbalance in clustering tasks. By employing an artificial neural network (ANN) to create a specific parameter space, we distinguish between neutrino blazar (NB) and non-NB entities, pinpointing NB candidates within a 99% confidence interval. This approach has successfully identified 273 NB candidates from the 4FGL-DR3 catalog, offering insights into the relationship between blazars and neutrinos and potentially reshaping our understanding of cosmic ray mechanisms.

In conclusion, Ch. 8 synthesizes the findings of this dissertation, drawing upon the transformative potential of AI approaches in elucidating the intricacies of blazar research. The chapter further postulates the burgeoning role of AI, anticipating its more expansive and profound implications in future explorations within the domain of astrophysics studies.

Contents

1	Introduction	1
1.1	Brief overview of AGN	1
1.2	Classification and Unified Model of AGN	2
1.2.1	Classification of AGN	2
1.2.2	Energy mechanism of AGN	5
1.2.3	Unified Model of AGN	8
1.3	Blazar	15
1.3.1	Types of Blazars	16
1.3.2	Blazar Sequence	17
1.3.3	Multi-band Radiation Characteristics of Blazars	18
1.3.4	SED Characteristics of Blazars	21
1.3.5	Radiative Models of Blazars	22
1.3.6	Jet of Blazars	26
2	<i>Fermi</i> telescope and <i>Fermi</i> source catalogs	31
2.1	Historical Context and Genesis of the <i>Fermi</i> Mission	31
2.2	<i>Fermi</i> instruments	32
2.2.1	<i>Fermi</i> -LAT	32
2.2.2	<i>Fermi</i> -GRB	36
2.3	<i>Fermi</i> Data Processing	37
2.3.1	Instrument Response Functions	37
2.3.2	Data Analysis Versions	38
2.3.3	Effective Area Calculation	38
2.3.4	Event Reconstruction Versions (Passes)	38
2.3.5	Evolution of Data Analysis	39
2.3.6	The Importance of IRFs	39
2.4	<i>Fermi</i> -LAT Gamma-Ray Source Catalogs	39
2.4.1	The First <i>Fermi</i> -LAT Source and AGN Catalog	39
2.4.2	The Second <i>Fermi</i> -LAT Source and AGN Catalogs	40
2.4.3	The Third <i>Fermi</i> -LAT Source and AGN Catalogs	41
2.4.4	The Fourth <i>Fermi</i> -LAT and AGN Catalogs	43
3	Methodology	47
3.1	Briefly overview of AI and Data Mining	48
3.2	Historical Context	49
3.3	Segmentation of AI	50

3.4	ML	51
3.4.1	Supporting Disciplines of ML	52
3.4.2	Classification of ML algorithms	53
3.4.3	ML flowchart	54
3.4.4	Basic framework of ML	54
3.4.5	SML algorithms	71
3.4.6	UML algorithms	79
3.4.7	Data Preprocessing	90
3.4.8	Splitting Dataset into Training, Validation, and Testing Sets	92
3.4.9	Data Reduction	93
3.5	ML/DL applications in astronomy	99
3.6	Class-imbalance problem	101
3.7	Sklearn and Pytorch	103
4	Radio Dualism in in AGN: A Detailed Study	105
4.1	Samples	106
4.2	Results	106
4.2.1	Distribution of Source Numbers vs. Redshift	106
4.2.2	Radio Loudness	106
4.2.3	Examining the Relationship Between Radio and Optical Luminosity	108
4.3	Discussion	109
4.3.1	Comparison of Source Distributions	109
4.3.2	Revisiting the Radio Dichotomy in Quasars: A Comprehensive Analysis	111
4.3.3	Refined Classification of Radio Loud and Radio Quiet Sources	112
4.4	Summary	113
5	The Jet Apparent Motion and Central Engine Dynamics in <i>Fermi</i> Blazars	115
5.1	Introduction	115
5.2	Data Collection Methodology	116
5.3	Results	117
5.3.1	Association Between Maximum Apparent Velocity and Luminosity, Photon Index, and Variability Index in <i>Fermi</i> GeV γ -ray Emission	117
5.3.2	Correlation Analysis: Maximum Apparent Velocity, Black Hole Mass, Accretion Disk Luminosity, and Normalized Accretion Luminosity	119
5.4	Discussion	123
5.4.1	Apparent Motion and Central Engine in Blazars: A Correlational Analysis	123
5.4.2	Criteria for Selecting VFB Candidates from <i>Fermi</i> Blazars	127
5.5	Conclusion	128
6	Investigating TeV blazar candidates of <i>Fermi</i> blazars	131
6.1	Introduction	131
6.2	SML	133
6.2.1	Data Preprocessing	133
6.2.2	Selection of the SML model	134
6.2.3	Evaluating performance	135
6.3	Optimization of Model Performance through Training and Tuning	135
6.3.1	Constructing Linear Boundaries	136

6.3.2	Predicting Labels	136
6.4	LR Model for Distinguishing TeVs from Non-TeV s	136
6.4.1	Sample Data Sources	136
6.4.2	TeV Source Identification	137
6.4.3	Analysis of LR in IdentifyingTBCs	137
6.5	Potential Targets for Ground-Based Cherenkov Detectors	138
6.5.1	SEDs Considering EBL Correction	138
6.5.2	Visibility Analysis ofTBCs for Ground-Based Telescopes	146
6.5.3	High-confidence TBCs	148
6.6	Discussions and conclusions	152
6.6.1	Discussions	152
6.6.2	Conclusions	154
7	Chasing the neutrino candidates of <i>Fermi</i> blazars	157
7.1	Introduction	157
7.2	Neutrino Detection in IceCube	158
7.3	Neutrino Emission from Blazars and Seyfert Galaxies	160
7.4	Methodology	164
7.5	Experiments	165
7.6	Blazars as Neutrino Emitters	165
7.6.1	Sample Selection	165
7.6.2	Feature engineering	166
7.6.3	Domains and Tasks	168
7.6.4	Transfer Learning Processing and Results	168
7.7	Discussion	171
7.7.1	Statistical Analysis	173
7.7.2	Single-Zone vs. Two-Zone Models	174
7.8	Conclusion	176
8	Conclusions and Prospect	177
8.1	Prospect	178
9	Appendix	179

List of Figures

1.1	A schematic representation of the SED of an AGN is shown by the black curve, divided into its primary physical components represented by colored curves, and compared with the SED of a star-forming galaxy depicted by the gray curve (Hickox & Alexander, 2018).	3
1.2	Emission line ratio diagnostic diagram by Baldwin et al. (1981) for SDSS emission-line galaxies. Star-forming galaxies are shown in green, Seyfert 2 in blue, and LINERs in red. The two different curves used to separate AGN and starbursts are indicated by the solid curve (Kauffmann separator) and the dashed curve (Kewley separator) (Groves et al., 2006).	6
1.3	A schematic representation of AGN classifications (Errando Trias, 2009)	6
1.4	Top: Combined jet structure of M87 from 2013 to 2018, observed every two years at a frequency of 43 GHz. The corresponding year is displayed in the top left corner of each subfigure. White arrows indicate the position angle of the jet in each subfigure. Bottom: Best fit results based on images combined annually from 2000 to 2022. Green and blue dots represent data from observation frequencies of 22 GHz and 43 GHz, respectively. The red line represents the best fit based on the precession model (Cui et al., 2023).	8
1.5	Schematic representation of the AGN Unified Model (Marin, 2016). Type 1 AGN are visible at inclinations of 0° – 60° , while Type 2 AGN are visible at approximately 60° – 90° . Color code: Central SMBH in black, surrounding X-ray corona in purple, multi-temperature accretion disk in rainbow pattern, BLR in red and light brown, torus of dust around the nucleus in dark brown, polar ionized winds in dark green, and the final extension of the NLR in yellow-green. A double-sided, kilo-parsec jet is added to account for radio-loud AGN.	9
1.6	Types of active galaxies (Dermer & Giebels, 2016). Cartoon illustration of AGN taxonomy, following the unified scheme for radio-loud and radio-quiet galaxies. In this two-parameter model that includes the observational angle, any given AGN is either radio-quiet or radio-loud, speculated here to depend on black hole spin a/M (where M is black hole mass and a is its angular momentum), and either low-power or high-power, determined by the mass accretion rate. Misaligned AGN sources are radio galaxies, including low-luminosity FR1 and high-luminosity FR2 radio galaxies corresponding to BL Lac objects, differentiated by direction into narrow-line and broad-line radio galaxies.	10

- 1.7 Types of active galaxies (Tadhunter, 2008). Cartoon illustration of AGN taxonomy, following the unified scheme for radio-loud and radio-quiet galaxies. In this model that includes the observational angle as one of the parameters, any given AGN is either radio-quiet or radio-loud, speculated here to depend on black hole spin a/M (where M is black hole mass and a is its angular momentum), and either low-power or high-power, determined by the mass accretion rate. Misaligned AGN sources are radio galaxies, including low-luminosity FR1 and high-luminosity FR2 radio galaxies corresponding to BL Lac objects, differentiated by direction into narrow-line and broad-line radio galaxies. 11
- 1.8 A conceptual illustration depicts the arrangement of the accretion flow across various spectral states of black holes, contingent upon the mass accretion rate. The hot accretion flow is denoted by red triangles, while the conventional thin disk is represented by robust black horizontal lines. As the mass accretion rate escalates, the transition radius R_{tr} , at which the thin disk truncates, diminishes (Yuan & Narayan, 2014). . . . 13
- 1.9 Fossati et al. (1998) shows the synthetic SED of PKS 2155-304 showing the highest and lowest simultaneous states of the night, as well as historical data (shown in gray). The hard X-ray data (butterfly) corresponds to the RXTE-HEXTE spectrum of the high state in 1996. The right axis provides the luminosity scale in erg/s. Highest state (blue triangles): T300-high spectrum scaled to the highest VHE flux in the 4-minute light curve. Lowest state (red squares): T300-low spectrum, scaled to the lowest VHE flux in the 4-minute light curve. The X-ray state during the corresponding period is equal to the T300-Xmax and T400-Xmin spectra, so it is plotted here. The dashed line shows the single-zone SSC fit of the multi-wavelength activity in 2003 (black circles). 19
- 1.10 Broadband SED of 1ES 1959+650 with data from MJD 56064 (black) and MJD 56067 (red). The VHE spectrum for the low state is represented by the average spectrum measured over the two dark runs, excluding the state on MJD 56067. These data are explored with an SSC representation, where the black line corresponds to the low state and the red dashed and dotted lines correspond to the high γ -ray state observed on MJD 56067 (Aliu et al., 2014). 22
- 1.11 Schematic representation of the primary hadronic and leptonic radiative processes in the emission region (Mastichiadis, 2016). 23
- 1.12 The renowned jet of quasar 3C 273 is depicted through multi-wavelength observations: (a) the radio frequency image at 1.6 GHz sourced from the MERLIN archive; (b) the optical representation captured by the Hubble Space Telescope at a wavelength of 6000 Å; and (c) the X-ray visualization derived from a series of observations by the Chandra X-ray Observatory, processed to highlight the energy range of 0.3–6 keV and subsequently subjected to minor smoothing for clarity (Boettcher et al., 2012). . . 27
- 1.13 Lobe and knot of Cygnus A. Figure courtesy of <http://www.cv.nrao.edu/abridle/images/knot.pdf>. 28
- 1.14 Schematic representation of superluminal motion (Ghisellini, 2013). A sub-source moves at a speed of βc . In the comoving frame, it moves from point A to point B in time Δt_e , emitting photons. The time interval between photon receptions at point D is $\Delta t_a = \Delta t_e / \delta$ 29
- 1.15 Multi-epoch KaVA 22 GHz images of the M87 jet. All images are convolved with a 1 mas circular beam. The green line represents a constant velocity at the speed of light (Hada, 2017). 30

2.1	The γ -ray sky in the view of <i>FGRST</i> beyond 1 GeV within 12 years. Figure courtesy of https://space.mit.edu/HETG/Reports/HETG_Report_SciJun02.html	32
2.2	The FGST and its two instruments. The LAT images the sky in the energy band from about 20 MeV to more than 300 GeV while the GBM complements the LAT for the study of GRBs and transients, providing spectral coverage from 8 keV to about 40 MeV (Michelson et al., 2010).	33
2.3	Schematic representation of the LAT instrument's cross-section, illustrating the annihilation of incident photons into e^+ and e^- . Figure courtesy of https://fermi.gsfc.nasa.gov/ssc/data/analysis/do	
2.4	The clean sample's source locations are depicted as follows: FSRQs are marked with red circles, BL Lacs with blue circles, radio galaxies with magenta stars, and AGNs whose types are not identified are indicated by green triangles. (Abdo et al., 2010a). .	40
2.5	The positions of sources in the 2LAC clean sample. Red circles: FSRQs; Blue circles: BL Lac objects; Magenta circles: non-blazar AGNs; Green triangles: AGNs of unknown type (Ackermann et al., 2011).	41
2.6	The distribution of sources in the 3LAC clean sample. Red circles: FSRQs; Blue circles: BL Lac objects; Magenta circles: non-blazar AGNs; Green triangles: AGNs of unknown type (Ackermann et al., 2015a).	42
2.7	Survey locations of the 4FGL sources (upper part) and their positions on the galactic plane (lower three parts categorized by source type). All AGNs are marked in blue, well-defined associations of other classes in red, and unassociated sources or those with unknown properties in black (Abdollahi et al., 2020).	43
2.8	Locations of 4LAC sources projected on galactic coordinates (upper part) and J2000 equatorial coordinates (lower part) (Ajello et al., 2020).	45
3.1	The inclusion relationship between AI, ML, and DL (Yang et al., 2019).	47
3.2	Four sets of definitions of AI (Russell & Norvig, 2010, 2020).	48
3.3	Simple reflex agent diagram (Russell & Norvig, 2010, 2020).	48
3.4	Classification of ML algorithms (Peng et al., 2021).	53
3.5	SML flowchart.	55
3.6	Hypothesis space and version space (Hsieh & Wang, 2018).	56
3.7	Generative vs. Discriminative Models. Source: https://www.blog.dailydoseofds.com/p/intuitive-guide-to-generative	58
3.8	Bias-variance trade-off (Rashidi et al., 2019).	66
3.9	L2 norm regularization (left) and L1 norm regularization (right) (Santosh et al., 2022). .	67
3.10	K-Fold Cross-Validation https://www.analyticsvidhya.com/blog/2022/02/k-fold-cross-validation-technique-and-its-essentials/	68
3.11	Stratified K-Fold Cross-Validation (Duan, 2023).	69
3.12	Confusion Matrix (Duan, 2023).	70
3.13	ROC Curves (Raschka, 2015a).	71
3.14	The algorithm diagram of the perceptron and Adaline (Raschka, 2015a).	73
3.15	Sigmoid function (Raschka, 2015a).	74
3.16	Comparison of activation functions in Adaline and Logistic Regression.	75
3.17	Kernel trick diagram (Hachimi et al., 2020).	79
3.18	Schematic diagram of K-means algorithm (Raschka, 2015a).	80
3.19	human brain neurons, artificial neurons, and ANN (Bre et al., 2017).	86
3.20	TL process diagram https://medium.com/modern-nlp/transfer-learning-in-nlp-f5035cc3f62f	90

3.21	Three types of methods for Feature selection: filter, wrapper, and embedded method (Xie et al., 2020).	94
3.22	PCA versus LDA. Figure courtesy of https://sebastianraschka.com/Articles/2014_python_lda.html	98
3.23	PCA versus Kernel PCA. Figure courtesy of https://ml-lectures.org/docs/structuring_data/ml_without_neural_networks.html	98
4.1	Redshift distribution of source numbers. Quasars (red dots), BL Lacs (blue triangles), and other AGNs (green squares) are shown. Dashed curves represent Gaussian fits.	107
4.2	Radio loudness distribution. Quasars (red), BL Lacs (blue), and other AGNs (green) are represented.	107
4.3	Density probability distribution of radio loudness fitted with two Gaussian components.	109
4.4	Radio luminosity versus B-band luminosity. RLs (circles and solid lines) and RQs (triangles and dashed lines) are shown.	110
4.5	Construct a diagram illustrating the relationship between radio luminosity and radio loudness. In this diagram, represent RLs with circles and RQs with empty triangles. Include a black line in the diagram to demarcate the dividing line between these two categories.	113
5.1	the correlations of $\log L_\gamma$, α_{ph} , and $\log VI$ with $\log \beta_{\text{app}}^{\text{max}}$ across different subclasses of blazars. Here, FSRQs are represented by red markers, BL Lacs by blue, and BCUs by green.	120
5.2	the correlations among various logarithmic parameters: the black hole mass ($\log \left(\frac{M_{\text{BH}}}{M_\odot} \right)$), the disk luminosity ($\log L_{\text{Disk}}$), and the ratio of disk luminosity to Eddington luminosity ($\log \left(\frac{L_{\text{Disk}}}{L_{\text{Edd}}} \right)$). These parameters are analyzed against the maximum apparent velocity ($\log \beta_{\text{app}}^{\text{max}}$) for VFBs. FSRQs are denoted in red, BL Lacs in blue, and BCUs in green.	122
5.3	The correlations between the intrinsic logarithmic luminosities, denoted as $\log L_\gamma^{\text{in}}$, and the maximum apparent velocities, $\log \beta_{\text{app}}^{\text{max}}$, for the VFBs. The FSRQs are represented using red markers, whereas the BL Lacs are depicted with blue markers.	124
5.4	the relationship between the logarithmic values of radiative power ($\log P_{\text{rad}}$) and extended luminosity at 5 GHz ($\log L_{5\text{GHz}}^{\text{ext}}$) with the logarithmic maximal apparent beta ($\log \beta_{\text{app}}^{\text{max}}$) of VFBs. FSRQs represented in red, BL Lacs in blue, and BCUs in green.	126
5.5	Distributions for the parameters: γ -ray luminosity (denoted as $\log L_\gamma$), photon index (represented by α_{ph}), and the variability index (expressed as $\log VI$), were modeled GMM. The VFBs are depicted with blue bars, while the remaining <i>Fermi</i> blazars are illustrated using orange bars.	129
6.1	SFS analysis for Four Learning Sets. The graphs are structured with the number of parameters on the X-axis and the concentration of AUC on the Y-axis. Each learning set is represented in a separate panel, with five dotted lines in various colors indicating the training iterations from Training 1 to Training 5. <i>Top Left Panel</i> : This panel illustrates the results for the 4FGL-DR2 / 4LAC-DR2 data set; <i>Top Right Panel</i> : This panel shows the SFS analysis for the 3FHL data set; <i>Bottom Left Panel</i> : In this panel, the results for the 3HSP data set are displayed; <i>Bottom Right Panel</i> : This panel presents the analysis for the 2BIGB data set.	141

6.2 The AUC values for four distinct learning sets are presented, each characterized by blue histograms for training sets and orange histograms for testing sets. *Top Left Panel: 4FGL-DR2 / 4LAC-DR2* - This panel likely represents data or results associated with the 4FGL-DR2 and 4LAC-DR2 datasets; *Top Right Panel: 3FHL* - This panel is dedicated to the 3FHL dataset; *Bottom Left Panel: 3HSP* - This panel focuses on the 3HSP dataset; *Bottom Right Panel: 2BIGB* - This panel displays information related to the 2BIGB dataset. 142

6.3 The analysis involves the ROC curve for four distinct prediction datasets, arranged as follows: The 4FGL-DR2 and 4LAC-DR2 datasets are displayed in the *top left panel*; The 3FHL dataset is shown in the *top right panel*; The 3HSP dataset is featured in the *bottom left panel*; The 2BIGB dataset is presented in the *bottom right panel*. 143

6.4 Describe the classification and prediction data sets for 3FHL, 3HSP, and 2BIGB. In these data sets, red triangles represent TeV blazars, blue circles indicate non-TeV blazars and the green line or plane is the classification boundary. The arrangement is as follows: the *top left panel* shows 3FHL, the *top right panel* displays 3HSP, the *bottom left panel* presents 2BIGB from the perspective of non-TeV blazars, and the *bottom right panel* illustrates 2BIGB as viewed along the edge of the green classification plane. 144

6.5 For each of the four learning sets, the confusion matrix is presented with the following four panels. *Top Left Panel: 4FGL-DR2 / 4LAC-DR2*; *Top Right Panel: 3FHL*; *Bottom Left Panel: 3HSP*; *Bottom Right Panel: 2BIGB*. In the context of coordinate values, ‘T’ represents ‘True’, and ‘F’ signifies ‘False’. 145

6.6 SEDs for candidates with energies around 40 TeV are presented, with only two items displayed for brevity, the complete version can be found at(Zhu et al., 2023). A comprehensive version of this figure is accessible in the online publication. Data points are denoted by grey error bars. It is important to note that the photons in the TeV band have not been certified by IACTs and EAS arrays. The optimal-fitting synchrotron and IC bumps are represented by red and solid blue curves, respectively, while the black dash-dot curves illustrate the IC bumps absorbed by EBL. Three vertical dash-dot lines correspond to $\log \nu$ (Hz) at 1 TeV, 10 TeV, and 100 TeV. The sensitivities of various observatories are indicated by different colored dotted curves: green for CTAO north (north site, $0^\circ \leq Z \leq 20^\circ$), magenta for CTAO south (south site, $0^\circ \leq Z \leq 20^\circ$), olive for H.E.S.S zenith ($0^\circ \leq Z \leq 5^\circ$), tan for H.E.S.S low ($12^\circ \leq Z \leq 22^\circ$), blue for MAGIC low ($0^\circ \leq Z \leq 30^\circ$), cyan for MAGIC medium ($30^\circ \leq Z \leq 45^\circ$), blue-violet for VERITAS ($0^\circ \leq Z \leq 20^\circ$), red for LHAASO, and dark-orange for HAWC. Additionally, the redshifts (z) for each source are also provided. Furthermore, data points with flux errors exceeding flux upper limits are marked with pentagons, and crosses represent data in the low radio energy range ($\log \nu(\text{Hz}) < 9$). 147

6.7 In 2023, four sources including 4FGL J0123.7-2311 and 4FGL J0622.3-2605 will be observable by H.E.S.S. near its zenith. Visibility details for other IACTs and EAS arrays are in the online supplementary data. The H.E.S.S. online tool generated these visibility charts, where white indicates daylight, grey shades represent different twilights, yellow marks moon presence or twilight, and blue shows when the object is above certain altitudes. The RA/Dec coordinates are for the current epoch, not the J2000 standard. 149

6.8	Two confusion matrices for SML and HSPs. The <i>left panel</i> uses data from the 4FGL-DR2 / 4LAD-DR2 catalog with five FOFs: Γ , V_F , $\log f_7^{\text{ph}}$, z , $\log \nu_p^s$, while the <i>right panel</i> focuses on sources with $\log \nu_p^s \geq 15.3$ Hz. Both matrices display true positives (TP), false negatives (FN), false positives (FP), and true negatives (TN) in their respective quadrants.	153
6.9	Two confusion matrices for SML and HSPs. The <i>left panel</i> uses data from the 3FHL catalog with five FOFs: $\log f_2^{\text{ph}}$, z , while the <i>right panel</i> focuses on sources with $\log \nu_p^s \geq 15.3$ Hz. Both matrices display true positives (TP), false negatives (FN), false positives (FP), and true negatives (TN) in their respective quadrants.	153
7.1	Schematic of the IceCube detector. Image source: https://icecube.wisc.edu/science/icecube/ .	159
7.2	Principles of detecting different flavors of neutrinos in IceCube. Image source: chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj/https://www2.physics.ox.ac.uk/sites/default/files/2012-03-27/antonin_vacheret_pdf_18644.pdf	160
7.3	Neutrino energy spectrum observed by IceCube in its first eight years for muon neutrinos above 200 TeV. Image source: (Haack et al., 2017a).	161
7.4	Reconstructed arrival directions of muon neutrinos with energies above 200 TeV observed by IceCube over eight years. Different colors represent different energies. The gray solid line indicates the Galactic plane and the black dashed line represents the supergalactic plane (Haack et al., 2017a).	162
7.5	Light curves of TXS 0506+056 in various bands around the time of IceCube 170922A detection (IceCube Collaboration et al., 2018a).	163
7.6	Neutrino burst from TXS 0506+056 during 2014-2015 (IceCube Collaboration et al., 2018b).	163
7.7	The datasets used to train or evaluate the ANN model.	169
7.8	Training logic utilizing the triplet loss function.	170
7.9	The clustering results are displayed across four panels, each representing a different dataset. <i>Top Left Panel</i> : Clustering on the training set; <i>Top Right Panel</i> : Clustering on the validation set; <i>Bottom Left Panel</i> : Clustering on the testing set; <i>Bottom Right Panel</i> : Clustering on the <i>blazar samples</i> . In these panels, the <i>NB samples</i> are represented by red circles, while the <i>non-blazar samples</i> are denoted by blue triangles. The 20 core samples in the densest cluster of the <i>NB samples</i> in the bottom right panel are highlighted by purple circles. Additionally, the <i>non-NB samples</i> are illustrated by green rectangles.	172

List of Tables

4.1	The radio sources of our sample.*	114
5.1	The Selected Ensemble of Blazars from the <i>Fermi</i> Catalog	118
5.2	The correlation of $\log P_{\text{rad}}$ and $\log L_{5\text{GHz}}^{\text{ext}}$ against $\log \beta_{\text{app}}^{\text{max}}$	125
5.3	Potential Candidates of Superluminal Blazars Detected by <i>Fermi</i>	130
6.1	Extraction of Features from Four Catalogs for Learning Sets.	139
6.2	LR model and its performance on the 4FGL-DR2 / 4LAC-DR2 learning dataset	139
6.3	LR model and its performance on the 3FHL learning dataset	140
6.4	LR model and its performance on the 3HSP learning dataset	140
6.5	LR model and its performance on the 2BIGB learning dataset	140
6.6	Results of SED fitting in the observed frame for candidates with an energy of 40 TeV.*	146
6.7	Observation periods for IACTs and EAS arrays.*	149
6.8	Forecasting outcomes for 40 potential TeV blazar candidates*	150
6.9	Upper redshift boundaries for the 20 blazars as determined for each IACT and EAS array.*	151
6.10	Blazars at 81 TeV exhibit $\tau(E, z)$ when E equals 1 TeV.*	154
7.1	35 NB samples in 4FGL-DR3.	167
7.2	The 14 selected features for TL.	168
7.3	The structure of ANN model and the hyper-parameters of the training process.	170
7.4	273 NBCs*.	171
7.5	Statistical results	173
9.1	Potential Candidates of Superluminal Blazars Detected by <i>Fermi</i> *.	179
9.2	273 NBCs*.	185

Chapter 1

Introduction

1.1 Brief overview of AGN

Fath (1909) made a groundbreaking observation at the Lick Observatory, noting that the spectrum of NGC 1068, previously classified as a spiral nebulae, exhibited six pronounced emission lines. These lines starkly contrasted with the absorption spectra typical of stars. The seminal work of Hubble (1926) redefined our understanding of NGC 1068 and two other spiral nebulae, identifying them as extragalactic entities. This revelation ushered extragalactic objects into the forefront of astronomical research. Seyfert (1943) had identified a subset of spiral galaxies characterized by intense optical emission lines and luminous nuclei. These galaxies were subsequently termed "Seyfert galaxies." The post-World War II era, spanning the 1940s to the 1950s, witnessed a surge in radio astronomy, facilitated by the repurposing of wartime radio antennas for astronomical research. This period saw the emergence of radio surveys, notably the 3C radio survey MacDonalD et al. (1968), which introduced a plethora of radio sources to the astronomical community.

In 1959, the *Third Cambridge Catalog of Radio Sources* cataloged several luminous radio sources with optical counterparts resembling point-like stars. These entities were aptly named "quasars" (quasi-stellar objects or QSOs) Matthews et al. (1964). Woltjer (1959) posited that the radiation from extragalactic galaxies was concentrated within a mere 1pc of their centers, suggesting central masses on the order of $10^8 M_{\odot}$. The observation of Schmidt (1963) 3C 273 using the Palomar Observatory's 5-meter telescope revealed a redshift of 0.158, implying luminosities surpassing the Milky Way by two orders of magnitude. Subsequent observations of quasars like 3C 48 (Greenstein & Schmidt, 1964), with even greater redshifts as 0.37, underscored their extraordinary luminosities and atypical spectral properties, suggesting non-stellar energy sources.

Emerging insights from X-ray astronomy led Salpeter (1964) and Lynden-Bell (1969) to propose that the accretion of supermassive black holes (SMBHs, ranging from $10^6 - 10^{10} M_{\odot}$) could account for the immense luminosities observed in quasars. This realization bridged the understanding of Seyfert galaxies and quasars, categorizing them under the umbrella of active galaxies. Based on preliminary estimates, active galaxies constitute approximately 1% - 10% of the total galactic population (Ormes & for High Energy Astrophysics, Goddard Space Flight Center). In both observational and theoretical realms, the primary focus is on the properties of the nuclear region, termed the "Active Galactic Nucleus (AGN)". Typically, there is no strict distinction made between AGN and active galaxies in the literature.

Historically, the nascent stages of AGN research, constrained by observational capabilities were centered on individual sources, delving into their structure, radiation mechanisms, and central energy

dynamics. The prevailing consensus postulated that gas accretion onto the central SMBH could liberate approximately 10% of the rest-mass energy of the accreted material. This period also saw the development of continuum models rooted in accretion disk dynamics, photoionization emission line models, and dust torus-centric infrared radiation models.

The advent of large-scale spectral surveys, such as the Sloan Digital Sky Survey (SDSS), and multi-band deep field surveys, including the Hubble Deep Field and XMM-Newton Deep Field, exponentially expanded the catalog of known AGNs. For instance, the SDSS alone has increased the optically-certified AGN samples to approximately 100,000. This proliferation in data has shifted the research paradigm from individual source studies to large-sample statistical analyses, incorporating advanced artificial intelligence techniques to probe various facets of AGNs, from accretion gas sources to black hole seed origins.

As one of the quartet of astronomical marvels unearthed in the 1960s – alongside interstellar molecules, cosmic microwave background radiation, and pulsars – AGNs have captivated astronomers with their myriad of extreme observational attributes. Their compact scales, high redshifts, luminosity fluctuations, broad-spectrum electromagnetic radiation, and other phenomena have solidified their status as astronomical enigmas, continually driving research in the field.

1.2 Classification and Unified Model of AGN

1.2.1 Classification of AGN

A comprehensive understanding of the physical nature of AGN remains elusive. While there exists a qualitative grasp of the central engine of AGN, the intricate details of its physical structure, radiation processes, and population evolution are still under exploration. Consequently, the classification of AGN appears somewhat convoluted. With the advancement of observational techniques, classifications have become increasingly refined, with numerous subcategories emerging from the primary classes. Simultaneously, new types have been identified, and overlaps and mergers between various categories have blurred the demarcation lines. Due to the diverse observational bands and objectives in AGN research, historical naming and classification conventions persist. Netzer (2015) defines an AGN as a galaxy whose center contains a supermassive black hole with a mass greater than $10^5 M_{\odot}$, which is actively accreting material, and has an Eddington ratio $L_{\text{bol}}/L_{\text{Edd}} > 10^{-5}$. However, it is very difficult to strictly define AGN. It has no strict boundaries with normal galaxies and is usually determined by its observational characteristics: (Peterson, 1997; Krolik, 1998):

- **Luminous Compact Nucleus.** AGN radiation is concentrated in a compact region, typically less than 1 pc. However, their luminosity can rival or even surpass that of the entire galaxy. AGN luminosities range from 10^{42} to 10^{48} erg/s, while typical galaxy luminosities are around 10^{44} erg/s.
- **Broad-band Spectrum.** Unlike regular galaxies that primarily radiate in the optical band, AGN also exhibits strong X-ray emissions. Most AGN also displays radio emissions, and jet radiation in AGN spans almost the entire electromagnetic spectrum, from radio to Very High Energy (VHE) γ -rays (Fig. 1.1).
- **Ultraviolet Excess.** AGN exhibits strong ultraviolet emissions, distinct from regular stars. The optical-ultraviolet spectrum displays a prominent bump, termed the Big Blue Bump (BBB). This feature often serves as a primary criterion for AGN identification in surveys. The theoretical peak of the BBB is expected in the extreme ultraviolet (EUV) regime. As the mass of the

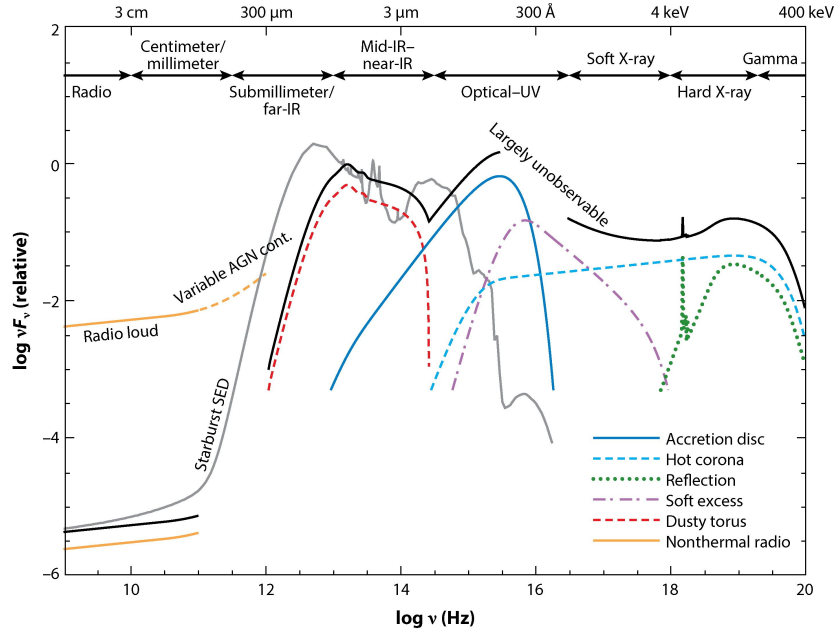


Figure 1.1: A schematic representation of the SED of an AGN is shown by the black curve, divided into its primary physical components represented by colored curves, and compared with the SED of a star-forming galaxy depicted by the grey curve (Hickox & Alexander, 2018).

central black hole increases, the theoretical expectation is that the accretion disk temperature decreases, leading to a softer EUV spectrum. Observationally, brighter quasars, which imply larger black hole masses, exhibit relatively weaker emission lines. This phenomenon can be explained by a softer EUV spectrum and is known as the Baldwin effect (Pogge & Peterson, 1992). This observation appears to be consistent with classical accretion disk theory. However, using ground-based SDSS and space-based GALEX observational data, and after accounting for the incompleteness in EUV detection, it was found that the average EUV spectrum of quasars is independent of their intrinsic luminosity. This not only suggests that intrinsic luminosity variations cannot explain the Baldwin effect but also poses a significant challenge to the predictions of standard accretion disk theory. The study proposes a potential new physical origin for the Baldwin effect: brighter quasars have smaller thermal fluctuations in their accretion disks, leading to fewer emission-line clouds. After correcting for the effects of intergalactic medium absorption, the research also revealed that the average EUV spectrum of quasars is softer than previously reported in other studies, further challenging the standard accretion disk model. This ultra-soft, luminosity-independent EUV spectrum aligns well with predictions from accretion disk models that incorporate disk winds, suggesting the widespread presence of disk winds (Murray et al., 1995) in quasars.

- **X-ray Continuum.** This can originate either from the jet or from the accretion disk's hot corona through scattering processes. In both scenarios, it exhibits a power-law spectrum. Some radio-quiet AGNs also display an excess of soft X-ray emissions, the origin of which remains uncertain. When X-ray continua irradiates cold heavy elements, it results in the emission of X-ray fluorescence lines, with the most notable being the iron feature around 6.4 keV. This line can be either narrow or broad; relativistically broadened iron lines can be instrumental in

studying the dynamics of the accretion disk very close to the nucleus, thereby shedding light on the nature of the central black hole.

- **Strong Emission Lines.** Unlike stellar and galaxy spectra, AGNs exhibit strong emission lines with Equivalent Widths (EW) of approximately 100 \AA . The Full Width at Half Maximum (FWHM) of these lines can be very broad, exceeding 10^4 km/s . Broad lines encompass permitted and semi-forbidden lines, but not forbidden lines. Narrow emission line components have widths on the order of a few hundred km/s and can include both forbidden and permitted lines. A typical AGN spectrum in the optical-ultraviolet range showcases the BBB and several primary emission lines. Additionally, near the Mg II line, there's a smaller bump, often referred to as the "small blue bump," primarily constituted by the Fe II emission line system mixed with the Balmer continuum (Vanden Berk et al., 2001).
- **Variability.** AGNs display significant variability in both continuum and emission line intensities. The variability amplitude and timescale are typically stochastic. Blazars, a specific AGN subclass, can exhibit flux variations spanning several orders of magnitude, with variability timescales ranging from years to minutes. Some sources even exhibit minute-scale variability (Aharonian et al., 2007a; Fan et al., 2021a; Pandey & Stalin, 2022).

It's worth noting that not all AGNs exhibit all the aforementioned characteristics. For instance, Type II AGN optical continua are predominantly governed by their host galaxies, with minimal ultraviolet excess and limited variability studies. Some subclasses may lack or exhibit weak emission lines. Additionally, many subclasses possess unique observational phenomena, some of which will be discussed in the subsequent sections on AGN classification. Despite the complex classification of AGNs, they are essentially categorized based on luminosity, optical spectra, radio intensity, and radio morphology (Tadhunter, 2008).

AGN classification, though seemingly intricate, fundamentally hinges on luminosity, optical spectra, radio intensity, radio morphology, and energy output modes (Tadhunter, 2008). Instead of enumerating various AGN types, the following discussion will elucidate AGN classification based on these characteristics.

Luminosity. Historically, quasars and Seyfert galaxies in AGN were named separately, with the sole distinction being their luminosity. Schmidt & Green (1983) proposed a criterion, where a nuclear luminosity in the B-band of $MB > -21.5 + 5 \log(h\nu)$ differentiates quasars from Seyfert galaxies.

Optical Spectra. Based on optical spectra, AGNs are bifurcated into Type I and Type II. Type I AGNs exhibit both broad and narrow emission line components in their spectra, whereas Type II AGNs display only the narrow component, with the optical continuum dominated by the host galaxy's stellar contribution. This classification further refines quasars and Seyfert galaxies. The distinction between Type I and Type II is not absolute; the intensity ratio of broad to narrow components might vary continuously. Transitional AGNs, primarily Seyfert galaxies, can be further subdivided based on this ratio, e.g., Seyfert 1.1, 1.3, 1.6 (Osterbrock, 1981). Additionally, some AGNs exhibit spectral transitions from Type I to Type II or vice versa, termed changing-look AGNs (LaMassa et al., 2015).

Using the ratio of narrow emission lines, Baldwin et al. (1981) devised a method to distinguish H II regions, resulting from star-forming activities, from the nuclear active regions in AGNs. As illustrated in Fig. 1.2, AGN can produce stronger high-ionization lines, such as $[\text{O III}] \lambda 5007/\text{H}\beta$, $[\text{N II}]/\text{H}\alpha$ $\lambda 6548, 6583$, indicating that the ionizing photons do not originate from O or B-type stars. Based on this methodology, AGNs with lower luminosities can be identified. Depending on the strength of high and low-ionization lines, these AGNs are further classified into Seyfert galaxies and Low-Ionization Nuclear Emission-line Regions (LINERs).

Beyond emission lines, absorption features are also present in AGN spectra. Studies reveal that approximately 10%–20% of quasar spectra exhibit broad absorption lines of high-ionization species such as CIV $\lambda 1549$, OVI $\lambda 1035$, SiIV $\lambda 1397$, NV $\lambda 1240$, and Ly α . Among these, around 15% also detect broad absorption lines of low-ionization species like MgII, FeII, and FeIII. Based on these absorption features, these broad absorption line quasars (BAL QSOs) are further categorized into High-ionization BAL QSOs (HiBALs) and Low-ionization BAL QSOs (LoBALs) (Reichard et al., 2003).

Radio Intensity. AGNs are typically classified into Radio Loud and Radio Quiet categories based on their radio loudness, defined as the radio-to-optical flux ratio $R = f_{5GHz}/f_{4400\text{\AA}}$ (Kellermann et al., 1989). AGNs with $R \geq 10$ are termed Radio Loud, while those with $R < 10$ are Radio Quiet. Miller et al. (1990) also suggested using radio luminosity L_{5GHz} , with a threshold of 10^{25}WHz^{-1} for differentiation. Radio Quiet AGNs typically exhibit radio emissions weaker by factors of 100 to 1000 compared to Radio Loud AGNs. Padovani (2016) argued that the distinction between Radio Quiet and Radio Loud AGNs is outdated and misleading. The primary physical difference between the two is the presence of strong relativistic jets. He proposed a classification based on jetted and non-jetted AGNs. Using machine learning, Ch. 4 revisits the specific value of radio loudness used to differentiate Radio Loud and Radio Quiet categories and reclassifies them based on optical B-band and radio 6 cm wavelength data.

Radio Morphology. Based on radio morphology, Radio Loud AGNs can be classified into compact (unresolved) and extended (resolved) sources. Among the radio-strong extended sources, they can be further categorized based on the flux and morphology of their core and lobes into core-dominated (FR I type), lobe-dominated (FR II type), and other morphologies (e.g., head-tail type). Additionally, with the improvement in the resolution of radio telescopes, Baldi et al. (2015) introduced a new subclass of radio galaxies—FR 0 type radio galaxies. These sources exhibit properties of FR I type radio galaxies, but their extended radiation is weaker by about 100 times compared to typical FR I type galaxies with comparable optical emission lines or radio core luminosities. Fig. 1.3 provides a systematic overview of AGN classification. For more detailed insights into AGN classification, refer to Padovani et al. (2016b).

1.2.2 Energy mechanism of AGN

To unify the various subclasses of AGN, it is imperative to first discern the energy mechanism of AGN. Historically, the energy source of celestial bodies was believed to be gravitational in the 19th century. This perspective shifted to nuclear reactions in the 20th century, up until the discovery of quasars in the 1960s. As previously mentioned, the typical luminosity of AGN, ranging from 10^{42} to 10^{48} erg/s, significantly surpasses that of normal galaxies. The variability timescale of AGN spans from a few hours to several years, suggesting that the maximum scale of AGN lies between 10^{-4} and 10 pc, which is considerably smaller than the scale of a regular galaxy. Matthews & Sandage (1963) estimated the radiative region of AGN to be less than a few light weeks based on optical variability timescales. Lynden-Bell (1969) discussed the energy source of radio-loud quasars, considering their high luminosity and small scale, and concluded that nuclear reactions could not be the primary energy source for quasars. Instead, gravitational contraction was deemed the most significant energy source. In fact, if the energy efficiency of nuclear reactions in stars is $\eta = E/Mc^2 \approx 0.7\%$, and if quasars also rely on nuclear reactions for energy, then they would require the conversion of approximately $250 M_{\odot}$ of hydrogen annually. Given that a typical galaxy has a mass of about $10^{11} M_{\odot}$, it could sustain such reactions for at most 10^9 years, which is significantly less than the age of the universe. Thus, nuclear reactions cannot be the primary energy source for quasars. The mechanism of energy production

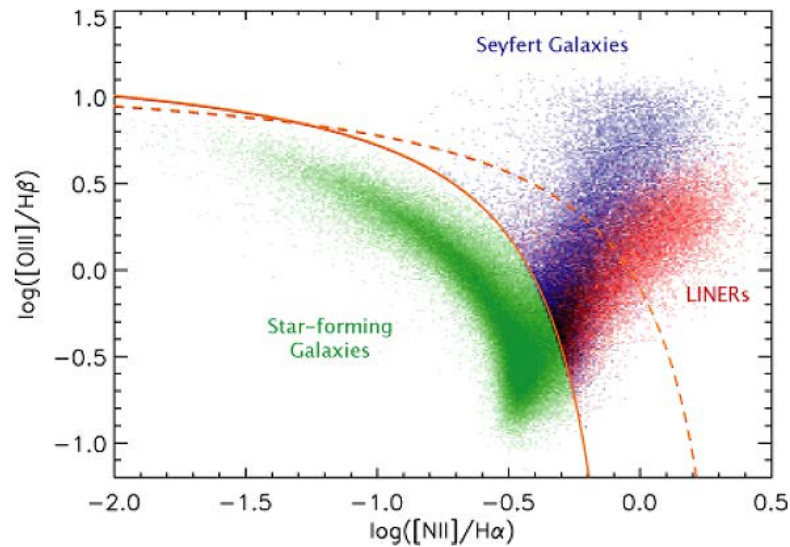


Figure 1.2: Emission line ratio diagnostic diagram by Baldwin et al. (1981) for SDSS emission-line galaxies. Star-forming galaxies are shown in green, Seyfert 2 in blue, and LINERs in red. The two different curves used to separate AGN and starbursts are indicated by the solid curve (Kauffmann separator) and the dashed curve (Kewley separator) (Groves et al., 2006).

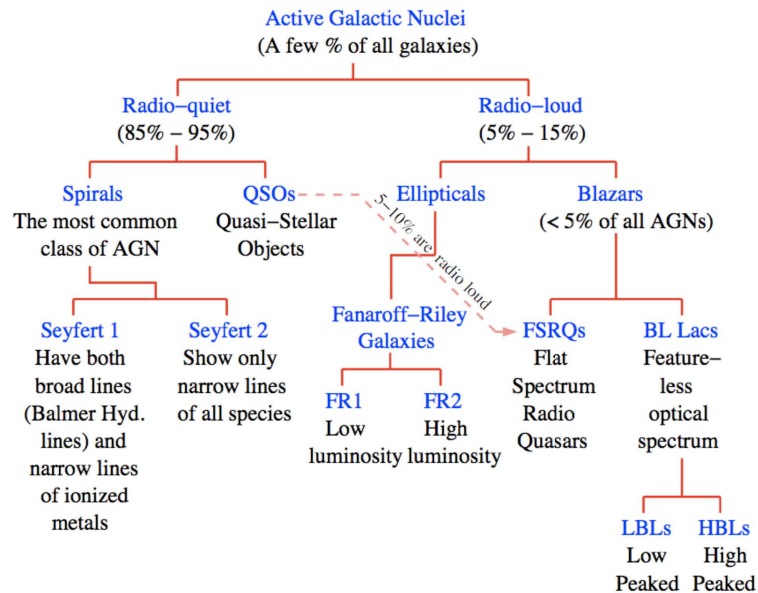


Figure 1.3: A schematic representation of AGN classifications (Errando Trias, 2009)

through the accretion of surrounding matter by a central dense body, releasing gravitational potential energy, began to gain traction.

Accretion refers to the process wherein a celestial body attracts and accumulates surrounding gas, dust, and other matter due to gravitational forces. The result of accretion is the conversion of gravitational potential energy into kinetic energy, which is eventually released as electromagnetic radiation. This energy conversion method is termed as accretion luminosity or energy production. The quantitative estimation of energy release from accretion by dense bodies has been widely accepted. For a central body with mass M and radius R , the gravitational potential energy released when the matter of mass m is accreted onto its surface is given by (Frank et al., 2002): $\Delta E_{\text{acc}} = GMm/R$, where G is the gravitational constant. If the central body is a neutron star with a mass comparable to the sun, i.e., $M \approx M_{\odot}$, and a typical neutron star radius $R \approx 10$ km, then the energy released per gram of accreted matter is $\Delta E_{\text{acc}} \approx 10^{20}$ erg. For matter of mass m , the energy released through nuclear fusion is $\Delta E_{\text{nuc}} = 0.007mc^2$, where c is the speed of light. Calculations indicate that the energy released per gram of matter through nuclear fusion is only 6×10^{18} ergs. To characterize the amount of energy released per gram of matter, Lynden-Bell (1969) introduced the energy conversion efficiency and defined it as: $\eta = \Delta E/mc^2$. The accretion luminosity efficiency is given by: $\eta = GMm/r/mc^2 = G/c^2 \times M/r$. It is evident that when M/r is large, implying that the central body is very dense, the efficiency of accretion luminosity is high. Conversely, when M/r is small, the efficiency is low. The theory of energy release through accretion does not necessitate that the center of an AGN is a black hole; it only requires that M/r is sufficiently large. Several models have been proposed, such as the giant pulsar model and the dense star cluster model. The giant pulsar model posits that the center of an AGN is a rotating giant pulsar, while the dense star cluster model suggests that the center of an AGN is a cluster of dense stars, such as neutron stars. However, these models have been refuted based on observational and theoretical evidence. If the central body is a star or neutron star, accreted matter can reach the surface of the body, and r is the radius of the body. For a typical neutron star, η is approximately 15%. When the central body is a black hole, r is the radius of the innermost stable circular orbit of the accretion disk. For a Schwarzschild black hole, $\eta = 6\%$; for a maximally spinning Kerr black hole, η can reach up to 42% (Frank et al., 2002). When the mass of the accreting body is sufficiently large, such an energy conversion efficiency can produce the observed luminosity of an AGN. Therefore, extracting energy from the accretion process around an SMBH provides a natural explanation for powering AGN. Rees (1984) proposed potential pathways for the formation of SMBH, and numerous observational evidences suggest the presence of SMBH at the centers of AGN (Peterson, 1993; Miyoshi et al., 1995; Beckmann & Shrader, 2012). During the accretion process, due to the transfer of angular momentum, an accretion disk is formed. Thus, the SMBH and the accretion disk together constitute the central engine of an active galactic nucleus. Whether the central black hole in an AGN is a Schwarzschild black hole or a Kerr black hole, and whether different types of black holes correspond to different types of AGN, are challenging questions to answer precisely. Analyses by Dabrowski et al. (1997); Wilms et al. (2001); Elvis et al. (2002) suggest that the majority of central black holes in AGN are likely Kerr black holes. Bian & Zhao (2003) conducted a study on a sample of AGN with known central black hole masses and determined the accretion efficiency. All radio-quiet AGN in the sample had an accretion efficiency of less than 5.6%, suggesting that their central black holes are likely Schwarzschild black holes. All radio-loud AGN had an accretion efficiency greater than 5.6%, indicating that their central black holes are likely Kerr black holes. Notably, Cui et al. (2023) analyzed observational data of the M87 galaxy from 2000 to 2022 and found that the jet from the central black hole of the M87 galaxy exhibits periodic oscillations, with a period of approximately 11 years and an amplitude of about 20 degrees. This phenomenon can be explained by the precession of the jet due to the Lense-Thirring effect, providing evidence that the M87 is a Kerr black hole.

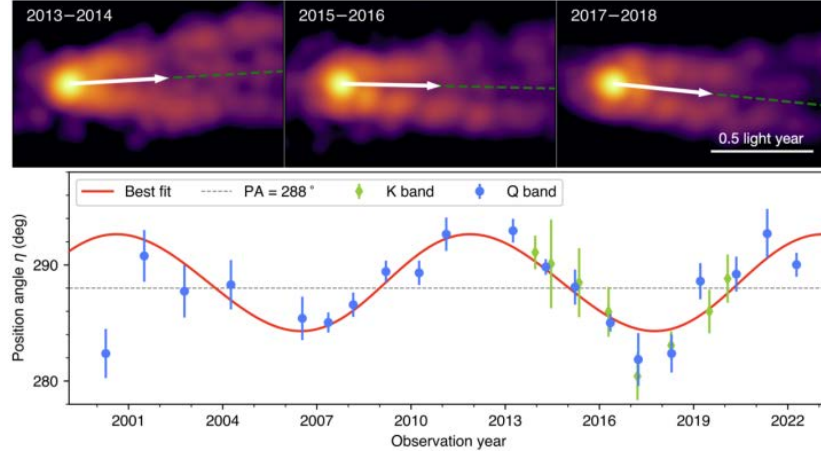


Figure 1.4: Top: Combined jet structure of M87 from 2013 to 2018, observed every two years at a frequency of 43 GHz. The corresponding year is displayed in the top left corner of each subfigure. White arrows indicate the position angle of the jet in each subfigure. Bottom: Best fit results based on images combined annually from 2000 to 2022. Green and blue dots represent data from observation frequencies of 22 GHz and 43 GHz, respectively. The red line represents the best fit based on the precession model (Cui et al., 2023).

1.2.3 Unified Model of AGN

As observations became more refined, it became evident that the mere presence of a black hole and an accretion disk could not account for several observed phenomena, such as broad emission lines, narrow emission lines, and large-scale radio emissions in AGN spectra. Moreover, the intrinsic properties of AGN are likely similar, and the observed differences might be attributed to varying observational angles. Based on these insights, the "Unified Model" was developed. The old unified scheme (Antonucci, 1993; Urry & Padovani, 1995; Urry, 2003) was a bold attempt to attribute the differences between AGN subclasses to varying observational angles (see Fig. 1.5). This model successfully explained the observed differences between Type I and Type II AGN. However, as observational research progressed, it became clear that the diversity in observational features of AGN subclasses could not be solely attributed to observational angles. A dynamic "weak" unified model, incorporating more intrinsic parameters, became necessary. For instance, adding rotation parameters and accretion rate/luminosity to the observational angle, as depicted in Fig. 1.6. Tadhunter (2008) proposed unifying AGN subclasses in three-dimensional parameter space (see Fig. 1.7), considering the presence of broad permitted lines in the spectrum, radio loudness, and luminosity.

Regardless of the unified model, it is widely accepted that AGN, from the inside out, comprises the following fundamental structures:

- **SMBH.** As previously mentioned, the central massive black hole is a fundamental element common to all AGN. The black hole mass ranges from 10^6 to $10^{10} M_{\odot}$, with a scale of approximately 10^{-7} to 10^{-3} pc. In astrophysical environments, the physical characteristics of a black hole are often described solely by its mass and spin. While the impact of black hole spin on AGN structure and radiation spectrum is challenging to confirm observationally, theoretical studies suggest that rapidly rotating black holes have accretion disks closer to the black hole, higher radiation efficiency, and produce bluer radiation spectra. Additionally, numerical simulations indicate that rapidly rotating black holes are more likely to drive jet emissions and the

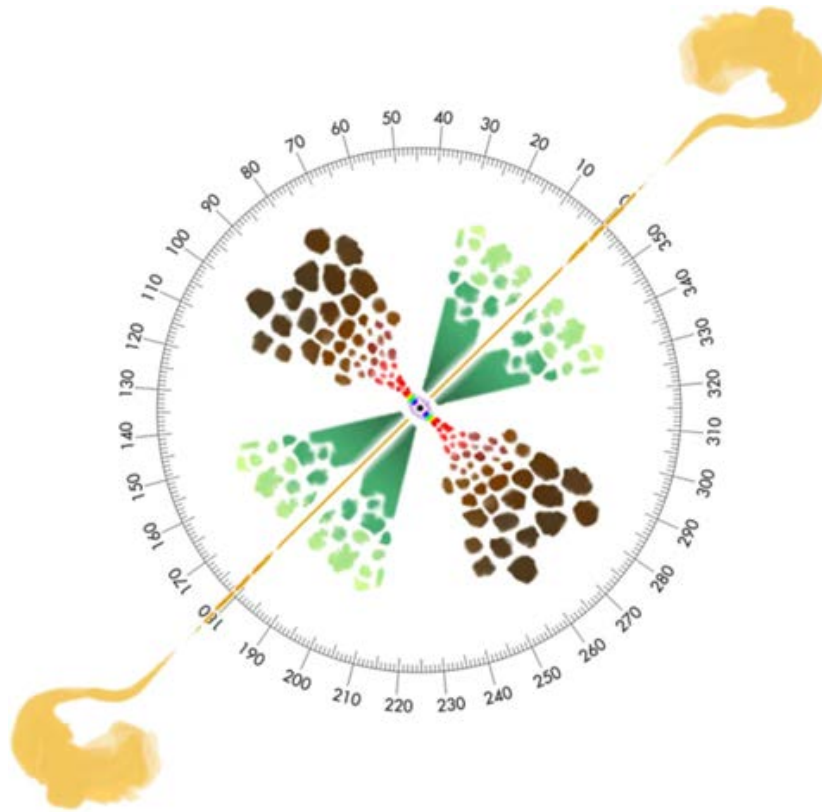


Figure 1.5: Schematic representation of the AGN Unified Model (Marin, 2016). Type 1 AGN are visible at inclinations of 0° – 60° , while Type 2 AGN are visible at approximately 60° – 90° . Color code: Central SMBH in black, surrounding X-ray corona in purple, multi-temperature accretion disk in rainbow pattern, BLR in red and light brown, torus of dust around the nucleus in dark brown, polar ionized winds in dark green, and the final extension of the NLR in yellow-green. A double-sided, kilo-parsec jet is added to account for radio-loud AGN.

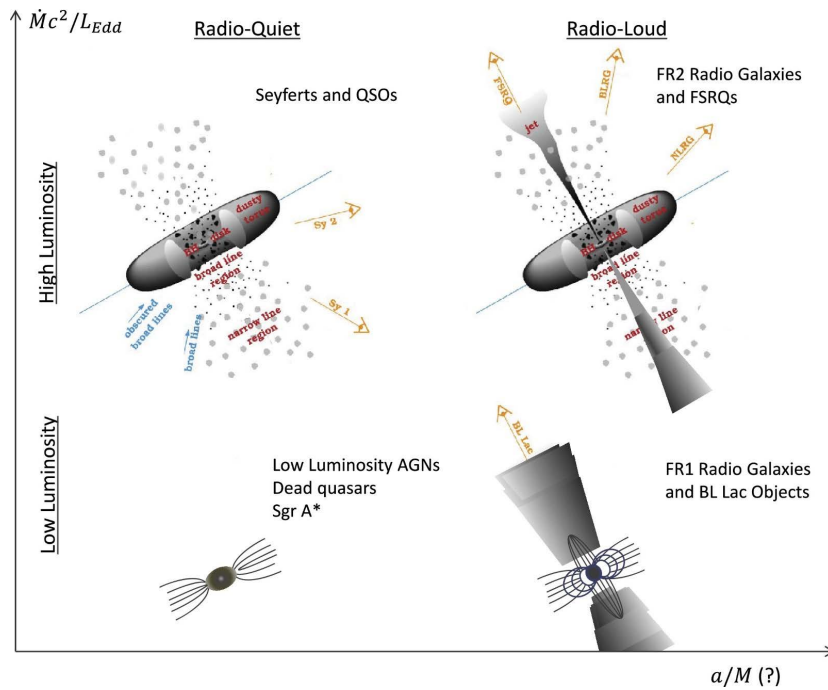


Figure 1.6: Types of active galaxies (Dermer & Giebels, 2016). Cartoon illustration of AGN taxonomy, following the unified scheme for radio-loud and radio-quiet galaxies. In this two-parameter model that includes the observational angle, any given AGN is either radio-quiet or radio-loud, speculated here to depend on black hole spin a/M (where M is black hole mass and a is its angular momentum), and either low-power or high-power, determined by the mass accretion rate. Misaligned AGN sources are radio galaxies, including low-luminosity FR1 and high-luminosity FR2 radio galaxies corresponding to BL Lac objects, differentiated by direction into narrow-line and broad-line radio galaxies.

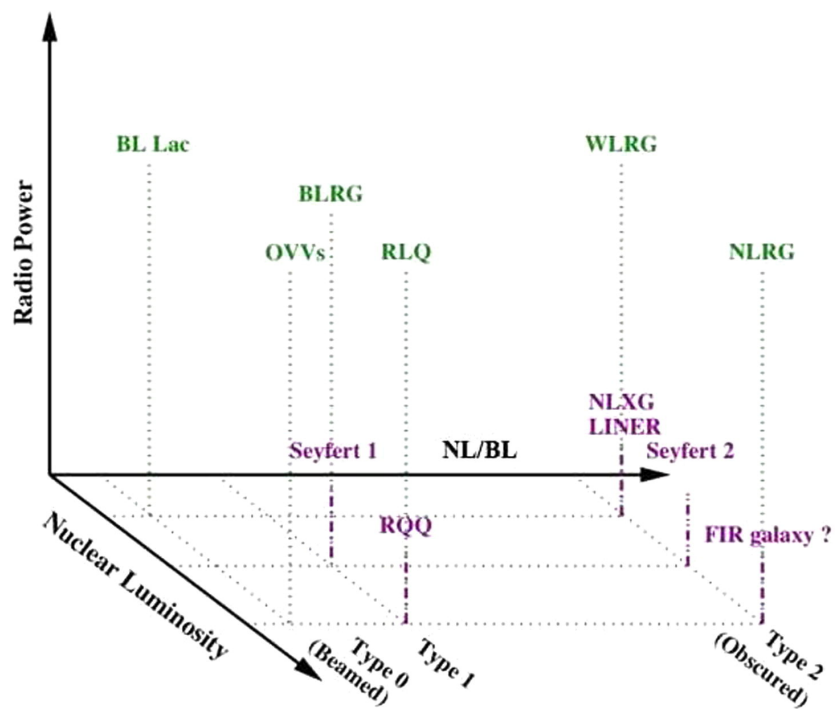


Figure 1.7: Types of active galaxies (Tadhunter, 2008). Cartoon illustration of AGN taxonomy, following the unified scheme for radio-loud and radio-quiet galaxies. In this model that includes the observational angle as one of the parameters, any given AGN is either radio-quiet or radio-loud, speculated here to depend on black hole spin a/M (where M is black hole mass and a is its angular momentum), and either low-power or high-power, determined by the mass accretion rate. Misaligned AGN sources are radio galaxies, including low-luminosity FR1 and high-luminosity FR2 radio galaxies corresponding to BL Lac objects, differentiated by direction into narrow-line and broad-line radio galaxies.

formation of disk winds (Tchekhovskoy et al., 2011; Narayan et al., 2014).

- **Accretion Disk (Disk).** SMBHs accrete surrounding matter, forming an accretion disk or flow with a scale ranging from $\sim 10^{-7}$ to 1 pc. The gas within the accretion disk releases gravitational potential energy through viscous processes, serving as the primary energy source for AGNs. During accretion, a fraction of the energy is radiated electromagnetically, while the remainder is carried away either by radial processes or by disk winds induced by radiative pressure. Both theoretical and observational evidence suggest that the structure of the accretion disk varies with the accretion rate (Narayan & Yi, 1994; Vasudevan & Fabian, 2009; Yuan & Narayan, 2014). Based on accretion rates, disk models can be categorized into Slim disk, standard thin disk (SSD), and advection-dominated accretion flows (ADAF) (Yuan & Narayan, 2014; Kat, 1998).

Specifically, a significant challenge in the development of accretion disk theory has been the lack of understanding of viscous processes. These processes lead to the majority of matter being accreted onto the central body, with a minor fraction carrying most of the angular momentum outward. This is a crucial physical process in accretion disk theory. Shakura & Sunyaev (1973) introduced the renowned α -model to describe viscosity, sidestepping detailed studies of the viscous process. This geometrically thin but optically thick accretion disk is also termed the standard thin disk or SSD. Another variant, the SLE disk (Shapiro et al., 1976), is dominated by gas radiative pressure but is optically thin. Slim disks, or optically thick ADAFs, were proposed by Abramowicz et al. (1998) as geometrically thick, radiatively dominated disks. These are thermally and viscously stable, transonic, and physically self-consistent, making them suitable for super-Eddington accretion scenarios. The presence of photon trapping or capture effects in slim disks leads to a saturation luminosity (Wang & Zhou, 1999; Wang et al., 2013), which has been proposed as a cosmological standard candle (Wang et al., 2013, 2014). ADAFs, or optically thin variants, were introduced by Narayan & Yi (1995) and Abramowicz et al. (1995). Some theories suggest a coexistence of multiple accretion modes around black holes, known as the truncated disk model (Done et al., 2012; Yuan & Narayan, 2014; Bambi, 2016).

- **Corona.** To explain the X-ray radiation of AGN, it is widely believed that a high-temperature (up to 10^9 K) hot plasma corona exists above the accretion disk near the black hole. Plasma in the corona produces intrinsic X-ray radiation by inverse Compton scattering of ultraviolet photons radiated by the accretion disk. The origin, heating mechanism, and geometric structure of the corona remain unclear. The Comptonization ability of the corona to soft photons depends on the electron temperature (T_e) and optical depth (τ_e), typically measured by the Compton factor $y = \max(\tau_e, \tau_e^2)(4kT_e/m_e c^2)$. When $y > 1$, the average energy gain of scattered photons is $A(y) = \exp(y)$. When $y \gg 1$, the average energy of scattered photons reaches the energy of hot electrons, at which point Compton scattering saturates. The spectral index of the scattered photon power spectrum is related to the Compton factor y as $\Gamma = -1/2 + \sqrt{9/4 + 4/y}$. When $y \approx 1$, $\Gamma \approx 2$, which is the intrinsic X-ray spectral index observed in typical AGN (Rybicki & Lightman, 1986).
- **Broad Line Region (BLR).** To explain the broad emission lines in AGN spectra, the unified model posits the existence of rapidly rotating (typical Keplerian speeds of approximately 3000 km s^{-1}) independent cloudlets outside the accretion disk. These cloudlets produce broad emission lines when ionized by radiation from the AGN center. Due to the absence of broad forbidden lines, the electron number density in the BLR is relatively high, typically not less than 10^8 cm^{-3} . Long-term spectral observations of AGN reveal that when the AGN continuum

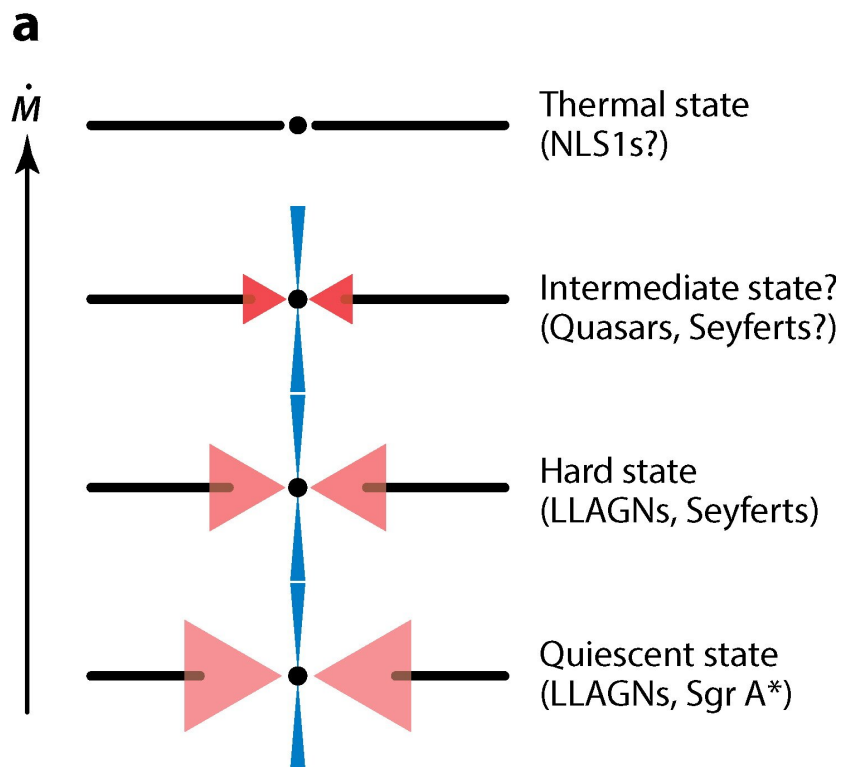


Figure 1.8: A conceptual illustration depicts the arrangement of the accretion flow across various spectral states of black holes, contingent upon the mass accretion rate. The hot accretion flow is denoted by red triangles, while the conventional thin disk is represented by robust black horizontal lines. As the mass accretion rate escalates, the transition radius R_{tr} , at which the thin disk truncates, diminishes (Yuan & Narayan, 2014).

undergoes light variations, the broad emission lines change accordingly, with time delays ranging from several weeks to several months (Wandel et al., 1999; Kaspi et al., 2000a, 2005; Kim et al., 2019). This suggests a BLR scale of approximately 0.01 to 1 pc, with the scale size correlating with luminosity. Assuming Keplerian motion of BLR cloudlets, the black hole mass can be estimated from the full width at half maximum (FWHM) of a single spectral observation (Netzer, 2013, e.g., using $H\beta$):

$$M_{\text{BH}}(H\beta) = 1.05 \times 10^8 \left(\frac{L_{5100}}{10^{46} \text{ erg s}^{-1}} \right)^{0.65} \left(\frac{\text{FWHM}(H\beta)}{10^3 \text{ km s}^{-1}} \right)^2 M_{\odot}. \quad (1.1)$$

- **Narrow Line Region (NLR).** AGN spectra feature forbidden emission lines, such as [O III] $\lambda\lambda 4959, 5007$, [N II] $\lambda\lambda 6548, 6583$, and [S II] $\lambda\lambda 6716, 6731$ (also known as nebula lines), believed to originate from ionized cloudlets outside the BLR with lower electron densities (10^3 to 10^6 cm^{-3}) and velocities ranging from 300 to 1000 km s^{-1} . With the development of integral field spectroscopy (IFU), the two-dimensional spatial structure of the NLR in nearby AGN can now be mapped. IFU images of the [Si VI] line in the Type I AGN NGC 4151 reveal a complex dynamical structure in the NLR, including rotation and outflows (Müller-Sánchez et al., 2011). Additionally, spatially resolved spectra have shown that the NLR in some AGN extends up to 100 kpc or even higher, with these regions termed extended narrow-line regions (ENLR).
- **Dust Torus.** To explain the observational differences between Type I and Type II AGN, the unified model suggests the existence of a toroidal structure composed of gas cloudlets and dust with a high column density ($>10^{22} \text{ cm}^{-2}$) in the outer regions of AGN. For Type II AGN, the dust torus obscures radiation from the central accretion disk and BLR, resulting in observational differences in the optical-ultraviolet bands compared to Type I AGN. The inner boundary of the dust torus depends on the sublimation temperature of the dust, with brighter AGN having a more distant inner boundary and a smaller covering angle for the dust torus. The structure of the dust torus is not continuous but consists of independent clumps (Elitzur & Ho, 2009).
- **Jets and Outflows** Black holes, during the accretion of matter, often give rise to outflows or jets (Blandford & Begelman, 1999). In the context of hot accretion flows, ADAF self-similar solutions indicate a positive Bernoulli parameter, suggesting strong outflows in the ADAF model. Observationally, most low-luminosity AGNs exhibit radio emissions. However, systems with geometrically thin cold disks typically exhibit weaker radio jet components. The reasons for this disparity remain elusive, but three potential factors are the geometric thickness of ADAFs facilitating magnetic field advection into the black hole, the larger Bernoulli parameter in hot accretion flows, and the role of outflow winds in jet collimation and stability (Yuan & Narayan, 2014).

AGNs exhibit well-collimated plasma and energy flows, observable across radio, optical, and X-ray wavelengths, spanning galactic to cluster scales. The relativistic motion of plasma clumps within these jets, due to light travel time effects, can appear superluminal. Interactions between the jet terminus and interstellar medium can lead to radio lobe or "cocoon" structures, with scales ranging from 1 to 1000 kpc and velocities typically less than 0.2 c . Various structures, such as knots and bends, are also observed in these jets. The formation, acceleration, and collimation mechanisms of these jets remain open questions. Two prerequisites for jet formation are believed to be rotation and magnetic fields. Large-scale magnetic fields are thought to be ubiquitous in the universe. Under certain conditions, the standard accretion disk in AGNs can

produce magnetically dominated outflows that transfer angular momentum, allowing for the advection of strong magnetic fields near the black hole (Cao & Lai, 2019).

Jets can be broadly categorized into two types based on morphology (Fender & Belloni, 2004): continuous and intermittent. Continuous jets appear temporally and spatially continuous, while intermittent jets manifest as discrete clumps. Several models have been proposed for the formation of continuous jets:

1. The BZ mechanism, proposed by Blandford & Znajek (1977), where large-scale magnetic fields extract the rotational energy of the black hole, leading to Poynting flux-dominated jets.
2. The BP mechanism, introduced by Blandford & Payne (1982), where the energy is extracted from the accretion disk's rotation, typically resulting in matter-dominated jets.
3. The Magnetic Tower (MT) model (Lynden-Bell, 2003), where large-scale magnetic fields anchored to the accretion disk generate axial magnetic field components, driving jet formation.

Non-relativistic outflows, termed disk winds, have slower velocities but may occupy a larger solid angle. Despite their lower energy compared to jets, disk winds can have significant mass loss rates. For ADAF models, early studies by Stone et al. (1997) utilized two-dimensional hydrodynamic simulations to compute inflow, outflow, and net accretion rates. Recent observations suggest that outflows in some quasar samples can originate from distances up to 10 pc from the galactic center (He et al., 2019), exceeding the predicted scales of disk wind models. These outflows may correspond to the scales of dust tori, with kinematics consistent with dust-driven outflow models (He et al., 2022), indicating diverse origins for AGN outflows.

1.3 Blazar

Blazars, as one of the most extreme subclasses of active galactic nuclei (AGN), are characterized by their non-thermal radiation predominantly emitted from relativistic plasma within their jets (Sambruna et al., 1996). All blazars are radio-loud (Peterson, 1997). Due to the small angle between our line of sight and the direction of jet motion, blazars only constitute a minor fraction of all AGNs. Consequently, our current samples of blazars remain relatively small. Most of our understanding of blazar phenomena is based on a limited number of luminous sources, suggesting that we are only seeing the tip of the iceberg. Compared to other subclasses of AGN, blazars typically exhibit the following distinct observational features (Wills et al., 1992; Urry & Padovani, 1995; Fan, 2002; Villata et al., 2006; Fan et al., 2014; Xiao et al., 2015; Gupta et al., 2016; Xiao et al., 2019; Abdollahi et al., 2020; Xiao et al., 2020; Fan et al., 2021b):

1. **Radio Loudness.** Dominant and compact emission sources.
2. **Rapid and Significant Variability.** Due to the relativistic beaming effect in the laboratory frame, blazars exhibit rapid and significant variability across all observable wavelengths.
3. **High Polarization.** Presence of high and variable linear polarization in both radio and optical bands. Optical polarization exceeds 3%, while radio linear polarization ranges between 1% to 2%.

4. **Non-thermal Continuous Spectrum.** A continuous non-thermal radiation spectrum is observed across all detectable wavelengths, with strong radiative luminosity from radio to X-ray bands.
5. **Superluminal Motion in VLBI Components.** Indicative of relativistic jets.
6. **Strong γ -ray Sources.** All extragalactic γ -ray sources detected by the Energetic Gamma Ray Experiment Telescope (EGRET) are blazars, with a majority of identified TeV sources being blazars.

From these observational characteristics, we can infer various intrinsic properties or physical processes of the jet. For instance, the absence of thermal radiation from the accretion disk, as observed in other galaxies, suggests that the dominant non-thermal radiation from the jet overshadows the thermal radiation from the inner black hole-accretion disk system. Superluminal motion hints at the relativistic motion of the primary radiation zone of the jet. The relativistic motion of this primary radiation zone, combined with the jet's motion direction nearly pointing towards the observer, results in a strong relativistic beaming effect. This effect inevitably leads to faster and more intense observed variability. Based on radio and optical polarization, it is evident that the radiation from radio to ultraviolet, and even soft X-rays, predominantly arises from synchrotron radiation of extremely relativistic electrons within the jet. If the intense gamma radiation is interpreted through the leptonic origin model, it implies the presence of ultra-relativistic electrons with $\gamma > 10^6$. However, synchrotron radiation would rapidly cool such high-energy electrons. To sustain continuous high-energy radiation, lower-energy electrons must be continuously accelerated to maintain a sufficient number of ultra-high-energy electrons. This acceleration/cooling process might further contribute to the rapid and significant variability observed.

1.3.1 Types of Blazars

Blazars were initially grouped as Optically Violent Variable (OVV) and BL Lac object (Peterson, 1997). OVV is the optical counterpart of FSRQ. Similarly, based on the mode of discovery, FSRQs have also been referred to as Highly Polarized Quasars (HPQs) and Core-Dominated Quasars (CDQ) (Sambruna et al., 1996). The more prevalent classification now combines FSRQ and BL Lac into the blazar category.

A classic division of blazars is based on the equivalent width (EW) of their emission lines. Flat Spectrum Radio Quasars (FSRQs) have an emission line $EW \geq 0.5$ nm, while BL Lacertae objects either lack emission lines or have an $EW \leq 0.5$ nm (Stickel et al., 1991). While this criterion is convenient, it has been argued that it doesn't distinctly separate FSRQs from BL Lacs. The main reason is the absence of a bimodal distribution in EW values, which can vary depending on the flux state of the source. Some blazars, despite having small observed EWs, inherently possess emission line luminosities comparable to FSRQs. Their broad-band SEDs also resemble those of FSRQs. The weak emission lines are a result of being overwhelmed by the non-thermal radiation from the jet. Thus, using EW as a distinguishing factor might lead to misclassifications (Ghisellini et al., 2011; Giommi et al., 2012).

The difference in emission line strength between BL Lac objects and FSRQs, which is attributed to photoionization by the accretion disk continuum, naturally raises the question of whether their accretion modes differ. Observational studies have indeed confirmed the presence of two distinct accretion modes in blazars (Marchesini et al., 2004; Ghisellini et al., 2009).

Originally, BL Lacs were further categorized into Radio-selected BL Lacs (RBL) and X-ray-selected BL Lacs (XBL) based on their discovery in radio and X-ray surveys, respectively (Giommi et al., 1995). Over time, it was realized that many RBLs are radio-strong and X-ray-weak due to their synchrotron emission peaking at relatively lower frequencies (infrared to optical), while many XBLs are radio-weak and X-ray-strong, with their synchrotron peak in the UV to X-ray range. Thus, a more accurate terminology emerged: Low-frequency-peaked BL Lacs (LBL) and High-frequency-peaked BL Lacs (HBL), depending on whether the radio/X-ray spectral index a_{rx} is greater or less than 0.75 (Urry & Padovani, 1995) or 0.8 (Sambruna et al., 1996).

the original classification of BL Lac objects by Padovani & Giommi (1995) delineated two subclasses: LSP BL Lac or LBL and HSP BL Lac or HBL. This classification was based on whether their broadband radio-to-X-ray spectral index was greater or less than 0.75 (Urry & Padovani, 1995) or 0.8 (Sambruna et al., 1996). This categorization was later expanded to encompass all blazars by Abdo et al. (2010b), who proposed a threefold subdivision based on the synchrotron peak frequency: LSP, where $\log \nu_p^s \leq 14.0$; ISP (intermediate synchrotron peaked sources), with $14 < \log \nu_p^s \leq 15$; and HSP, for $\log \nu_p^s > 15$. Subsequently, Fan et al. (2016) suggested slightly modified criteria, defining ISPs as having $14 < \log \nu_p^s \leq 15.3$ and HSPs as those with $\log \nu_p^s > 15.3$. A similar classification scheme was also recently presented by Yang et al. (2022a).

Furthermore, Costamante et al. (2001) introduced a fourth subclass, termed extreme HBLs (EHBLs), characterized by BL Lacs with $\log \nu_p^s > 17$. According to Foffano et al. (2019), EHBLs can be further divided into three subclasses based on the peak of their inverse Compton (IC) bump. The subclass with an IC bump peak between 0.1 TeV and 1 TeV is considered a continuation of HSP, whereas the subclass with an IC bump peak > 10 TeV, referred to as “hard-TeV blazars”, constitutes a distinct category. These hard-TeV blazars are notable for their high power, remarkably stable flux, and hard-TeV spectral behavior at TeV energies, presenting challenges to the synchrotron self-Compton (SSC) model. The remaining subclass, with an IC bump peaking at a few TeV, acts as a transitional class with a flat TeV spectral slope.

The distinction between BL Lacs and FSRQs based on their SEDs, the presence or absence of a blue bump (attributed to accretion disk radiation), and the orientation of the electric vector concerning the jet direction (Lister & Marscher, 1999) further underscores the complex nature of blazar classification and the underlying physics.

1.3.2 Blazar Sequence

In the study conducted by Fossati et al. (1998), three comprehensive blazar samples were analyzed. The findings revealed an inverse correlation between the luminosity of the synchrotron peak, the luminosity at 5GHz, the dominance of γ -rays, and the frequency of the synchrotron peak. A sequence was identified where sources with higher thermal luminosity had a lower synchrotron peak frequency, and conversely, sources with a higher peak frequency had lower thermal luminosity. This sequence illustrated an evolutionary scenario from FSRQs to LBLs to HBLs. Based on these findings, a unified model for the spectral distribution of blazars termed the “blazar sequence”, was proposed (see Fig. 1.9). Since the introduction of the blazar sequence, its authenticity has been a subject of continuous debate. Initial skepticism revolved around the existence of blazars with high peak frequencies and high luminosities. While some candidates were identified, subsequent observations refuted these claims (Padovani et al., 2002, 2003, 2004; Padovani, 2007). Ghisellini et al. (1998) provided a theoretical explanation using a single-zone, uniform, and steady-state electron model. A decrease in the synchrotron peak frequency corresponded to an intensification of the external photon field (originating from the accretion disk or broad-line region), which in turn enhanced the cooling capability of

relativistic electrons, leading to a decrease in γ_{peak} as Compton dominance increased.

Differentiating blazars based on their energy extraction methods, Cavaliere & D’Elia (2002) proposed an evolutionary sequence for blazars. They posited that as the accretion rate decreases from FSRQs to BL Lacs, the relativistic jet transitions from being powered by accretion to being powered by the Blandford-Znajek (BZ) mechanism. Böttcher & Chiang (2002) argued that the blazar sequence could not be solely explained by orientation effects. Through analytical radiation models, they estimated significant physical parameters in FSRQs and proposed an evolutionary sequence for blazars based on the temporal evolution of the accretion rate. They suggested that FSRQs existed in the early universe, and as time progressed, the accretion rate of the black hole decreased, leading to a reduction in the energy density of surrounding matter and external soft photon fields. This resulted in an evolutionary sequence from FSRQs to LBLs to HBLs. They further proposed that the accretion rate is directly proportional to the scattering optical depth of the BLR, using a single physical parameter, radial Thomson depth $\tau_{\text{T,BLR}}$, to describe the blazar sequence.

Ghisellini & Tavecchio (2008) systematically pointed out that the intrinsic physical reasons for the blazar sequence are due to differences in accretion modes (accretion rates) and black hole masses among different blazar subclasses. Based on observations, they assumed that the jet power is proportional to the accretion rate, the distance of the jet dissipation region from the black hole is proportional to the black hole mass, and the luminosity of the accretion disk cannot effectively ionize the broad-line region below a certain threshold. This led to the “theoretical blazar sequence”, explaining the blazar sequence and two types of exceptions (red low-power FSRQs and blue high-power FSRQs). They also predicted the existence of sources with high peak frequencies and high luminosities and sources with low peak frequencies and low luminosities. Giommi et al. (2012) used Monte Carlo simulations to analyze the effects of peak frequency position, host galaxy components, accretion modes, and beaming effects on blazar emission line intensities. They concluded that the observed blazar sequence is due to a selection effect caused by the absence of redshift constraints for most high peak frequency and high luminosity sources. Subsequently, Padovani et al. (2012) identified four sources with high peak frequencies and high luminosities using photometric redshift results. Meyer et al. (2012) proposed a possible explanation for the blazar sequence based on the properties of blazar jet structures using a radio-selected AGN sample. They suggested that the jets of BL Lacs might exhibit velocity gradients and are decelerating, while the jets of FSRQs do not have velocity gradients and can be described by a single Lorentz factor. Several observational evidences support the possibility of different Lorentz factors for BL Lacs and FRIIs, and the lack of γ -ray detections for FRIIs suggests that FRIIs/FSRQs might not exhibit such jet properties. In addition to the selection effect caused by redshift, the blazar sequence might also be a result of beaming effects. Nieppola et al. (2008) analyzed the correlation between synchrotron peak frequency and luminosity after correcting for beaming effects and found that they are no longer inversely correlated but have a certain positive correlation. Finke (2013) found a correlation between Compton dominance and synchrotron peak frequency, supporting the blazar sequence. After correcting for beaming effects, they found a positive correlation between intrinsic luminosity and frequency. Therefore, they suggested that this inverse correlation is caused by beaming effects or selection effects.

1.3.3 Multi-band Radiation Characteristics of Blazars

Radio band. Morphologically, based on the location where the radio radiation is mainly concentrated, the radio band radiation is divided into core-dominant and lobe-dominant types. In addition, a small number are in between the two. Most radio-loud AGNs have lobe-dominant radio radiation, and the radio radiation is generally considered to have a power-law spectrum: $\nu F_{\nu} \propto \nu^{-\alpha}$, where νF_{ν}

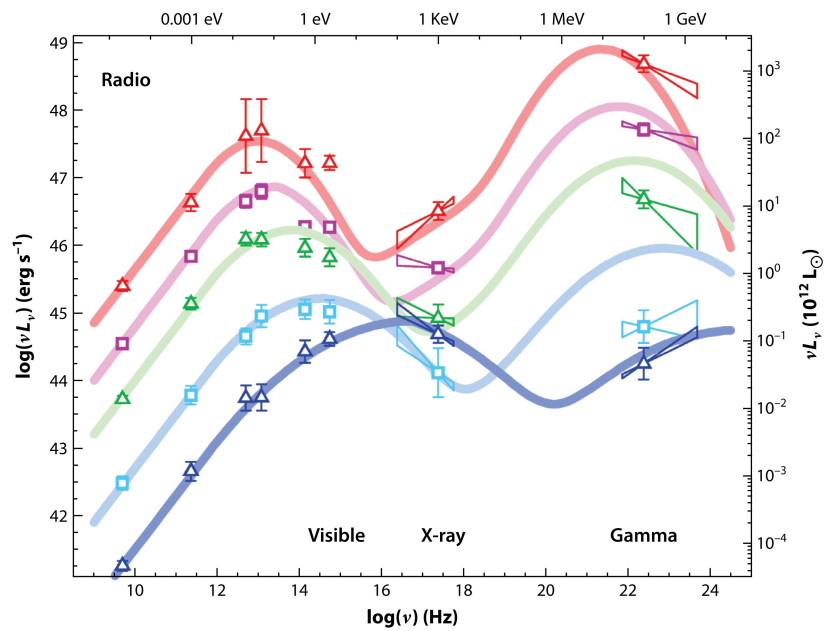


Figure 1.9: Fossati et al. (1998) shows the synthetic SED of PKS 2155-304 showing the highest and lowest simultaneous states of the night, as well as historical data (shown in gray). The hard X-ray data (butterfly) corresponds to the RXTE-HEXTE spectrum of the high state in 1996. The right axis provides the luminosity scale in erg/s. Highest state (blue triangles): T300-high spectrum scaled to the highest VHE flux in the 4-minute light curve. Lowest state (red squares): T300-low spectrum, scaled to the lowest VHE flux in the 4-minute light curve. The X-ray state during the corresponding period is equal to the T300-Xmax and T400-Xmin spectra, so it is plotted here. The dashed line shows the single-zone SSC fit of the multi-wavelength activity in 2003 (black circles).

is the flux density and α is the spectral index (Perrone, 2023). In the standard model (Blandford & Königl, 1979; Falcke & Biermann, 1995), the longitudinal pressure gradient theoretically leads to a moderate acceleration along the jet. Previously, this acceleration was ignored, and the velocity was initially proposed by Falcke (1995) to explain the inversion observed in the radio core of M81. Proper acceleration increases the Lorentz factor in the low-frequency direction, so the distance from the jet base increases, making the observed flux dimmer relative to the high-frequency direction. In summary, considering the influence of the longitudinal pressure gradient, a more inverse spectrum can be naturally explained. In addition, large-scale synchrotron radiation losses and adiabatic losses, as well as external pressures that may limit jet expansion, may also play a role.

UV-Optical-Infrared band. The continuous spectrum of Blazars from UV to infrared can be described by a power-law spectrum, with a spectral index between 0 and 1, and successfully explains the BBB; however, unlike general AGNs, blazars generally do not show BBB. In addition, blazars also have SBB, a dip at 1 μ m, and an infrared “bump”. The infrared radiation of blazars has both thermal and non-thermal origins (Sanders et al., 1989; Paltani et al., 1998). The thermal radiation origin theory believes that it may be the outer accretion disk of the black hole that absorbs surrounding matter and releases energy, but when fitting the spectrum with thermal radiation, the line of sight problem should be considered; but for the non-thermal origin of infrared radiation, the fast variable non-thermal radiation may come from synchrotron radiation inside the jet (Paltani et al., 1998). There is some evidence supporting the non-thermal radiation of infrared radiation: first, the power-law spectrum, which is considered to be a feature of non-thermal radiation; second, the absorption and emission of dust particles; third, light changes, which are considered to be strong evidence of non-thermal radiation origin; and finally, infrared radiation may be related to other radiation bands (such as hard X-rays). Blazars and quasars have a common feature of light changes, but they are different. For example, the light changes of quasars are small and have no periodicity, while some blazars can detect quasi-periodicity.

X-ray band. Blazars are radio-loud sources that emit strong X-ray radiation. Generally speaking, the X-ray radiation luminosity accounts for 5%-40% of the total thermal luminosity. Blazars have core-dominant and lobe-dominant types of radio-loud sources. The radio-loud sources of the two are different from X-ray radiation, of course, they are also different from radio-quiet quasars. Lobe-dominant radio sources account for the majority, and only a few are core-dominant radio sources. However, whether it is core-dominant or lobe-dominant sources, there are two points in common. First, the X-ray spectrum can be regarded as a power spectrum; second, there is no bump in the X-ray spectrum near 20KeV. X-rays can observe more violent light changes than other bands. Although the changes in X-rays are the most violent compared to other bands, rapid light changes in the X-ray band have not formed a stable period. In recent years, through the study of radio jets of radio-loud quasars, a batch of X-ray jets with various morphologies has been detected (Schwartz et al., 2000; Wilson et al., 2001; Marshall et al., 2001), indicating the complex structure of the jet, implying that the X-ray radiation mechanism should be applied to different parts of the X-ray jet.

γ -ray band. For blazars, they are strong γ -ray sources, and most of the radiation energy is mainly in the energy $E_{\gamma} > 100\text{MeV}$ band. A simple rate spectrum can be used to describe the γ -ray spectrum of the γ -ray band, with an average photon spectral index of 2.2, and there will be no significant difference between different subtypes; whether it is FSRQs or BL Lac objects, the gamma spectral index is not related to redshift. In terms of light changes, the γ -ray flux changes frequently, and the proportion of FSRQs is higher than that of BL Lac objects. In general, when the γ -ray flux increases, the spectrum becomes flat, and no periodicity of γ -ray flux has been found so far. The radiation of X-rays and high-energy γ -rays in blazars is very complicated, and many radiation mechanisms may work together. Such as blackbody radiation, synchrotron radiation (Maraschi et al., 1992a; Bloom

& Marscher, 1996a), inverse Compton scattering (Hardcastle et al., 2001), etc. The evolutionary stage of blazars may determine which model dominates the radiation in the X-ray band: synchrotron self-Compton (SSC) scattering or external Compton (EC) scattering. It is generally believed that for FSRQs, the EC model dominates the radiation in the X-ray band; while for BL Lacs, the SSC model dominates the radiation in the X-ray band.

Very High Energy γ -ray. Very High Energy (VHE) γ -rays, characterized by radiation energies exceeding 100 GeV, serve as one of the most significant messengers from the non-thermal universe. The primary impetus behind VHE γ -ray astronomy lies in the quest to identify the origins of high-energy cosmic rays. The Crab Nebula was the first VHE source detected by the Imaging Atmospheric Cherenkov Telescope (IACT) in 1989 (Weekes et al., 1989), marking a milestone in the field. Subsequently, an increasing number of VHE sources have been discovered across various celestial objects, with AGN being a prominent category.

Predominantly, VHE γ -ray emissions have been observed from blazars and a minority of radio galaxies. Notably, Mrk 421 was the first extragalactic VHE γ -ray source detected by IACT (Punch et al., 1992). Since the initial detection of Mrk 421, numerous extragalactic sources have been identified in the TeV band (Punch et al., 1992; Abeysekara et al., 2017a). Utilizing the 10-meter Gran Telescopio Canarias (GTC), Paiano et al. (2017) detected 22 objects in the constellation Scorpius, 15 of which were confirmed at TeV energy levels, with the remaining seven posited as TeV candidates. Rapid variability on the scale of hours in the TeV band for 3C 279 was reported by Paliya et al. (2015). Additionally, VHE γ -ray emissions from the flat-spectrum radio quasar PKS 1222+21 (4C 21.35, $z=0.432$) were detected by the MAGIC telescope (Aleksić et al., 2011; Zha et al., 2019).

In Ch. 6, this dissertation endeavors to identify further candidates of blazars emitting in the TeV range, expanding upon the current compendium of VHE sources.

1.3.4 SED Characteristics of Blazars

Observations across the entire electromagnetic spectrum provide us with the SED of blazars. In astronomy, multi-wavelength SEDs are typically represented in a $\log\nu$ - $\log\nu F_\nu$ space. A $\log\nu$ - $\log\nu F_\nu$ plot can intuitively reflect the radiative flux of different bands and the relative flux between them. The radiation from blazars primarily originates from non-thermal emission in the jet. Blazar multi-wavelength SEDs exhibit a distinct double-peaked structure, as shown in Fig. 1.10. Meanwhile, blazar SEDs are typically characterized by two well-separated bumps: a low-energy bump in the infrared to soft X-ray range due to synchrotron emission, and a high-energy bump between hard X-ray to γ -ray associated with inverse Compton (IC) radiation in the leptonic model (IC; e.g., Sikora et al. 1994a). The seed photons for IC scattering could either come from the same electron population that produces the synchrotron bump in the self-Compton (SSC) model (Ghisellini et al., 1985; Maraschi et al., 1992b; Bloom & Marscher, 1996b), e.g., 1ES 0347-121, 1ES 0229+200 (Costamante et al., 2018; Aharonian et al., 2007b,c), or from external regions in the External Compton (EC) model, e.g., from the accretion disk (Dermer & Schlickeiser, 1993a), broad line region (Sikora et al., 1994b), and dust torus (Błażejowski et al., 2000a). Alternatively, the hadronic process can also contribute to the high-energy bump through high-energy cosmic rays via photohadronic reactions (Dimitrakoudis et al., 2012; Tavecchio, 2014). Blazars are often observed to exhibit rapid and significant variability, with minimum variability timescales ranging from minutes to days. During blazar flaring episodes, changes might be observed across all spectral bands (Aharonian et al., 2009a). However, recent observations have revealed that flares in certain blazars, such as 1ES 1959 + 650 (Aliu et al., 2014) and Mkn 421, do not occur simultaneously across all bands but are restricted to specific ones. Fig. 1.10 presents the multi-wavelength SEDs of 1ES 1959 + 650 in both quiescent and flaring states. While a significant

outburst is observed in the VHE TeV γ -ray band for 1ES 1959 + 650, almost no variation is detected in other bands (optical, X-ray). Such unique flares are termed “orphan TeV flares” and are typically explained using the hadronic model.

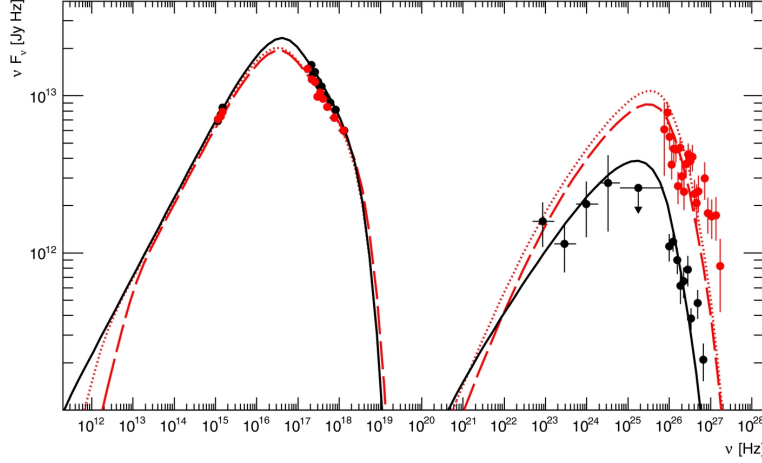


Figure 1.10: Broadband SED of 1ES 1959+650 with data from MJD 56064 (black) and MJD 56067 (red). The VHE spectrum for the low state is represented by the average spectrum measured over the two dark runs, excluding the state on MJD 56067. These data are explored with an SSC representation, where the black line corresponds to the low state and the red dashed and dotted lines correspond to the high γ -ray state observed on MJD 56067 (Aliu et al., 2014).

1.3.5 Radiative Models of Blazars

The high apparent thermal luminosity, rapid variability, and superluminal motion provide credible evidence for the existence of relativistic jets. The radiative models for blazars generally assume that the radiation is produced in a roughly spherical region. This region propagates along the jet with a speed of $\beta_T c$, corresponding to a bulk Lorentz factor of $\Gamma = 1/\sqrt{1-\beta^2}$. If the jet forms an angle θ with the line of sight, it will undergo Doppler boosting, with the Doppler factor given by $\delta_D = [\Gamma(1 - \beta_T \cos \theta)]^{-1}$. Doppler boosting and cosmological effects will influence the observed properties, mainly reflected in the following aspects, where primed quantities refer to the comoving frame and unprimed quantities refer to the observer’s frame: (1) Flux enhancement, $F_{\nu_{obs}}^{obs} = \delta_D^3 F'_{\nu'}$. (2) Frequency blue-shift, $\nu_{obs} = \delta_D \nu' / (1 + z)$. (3) Time variability shortening, $t_{var}^{obs} = t'_{var} (1 + z) / \delta_D$.

Apart from the relativistic boosting effects and cosmological effects mentioned above, the mechanism responsible for the low-energy peak of the blazar spectrum is generally believed to be syn-

chrotron radiation from relativistic electrons in a magnetic field. As for the γ -rays, their production mechanism remains a topic of debate. Current research on their radiative mechanism is mainly divided into two categories: leptonic models, which mainly involve inverse Compton scattering by relativistic electrons in the jet (the soft photons could either be from the jet's synchrotron radiation, from external regions, or cosmic microwave background photons), and hadronic models, which involve interactions between high-energy relativistic protons in the jet and external soft photons, leading to the production of secondary electron-positron pairs and low-energy *gamma*-rays. The secondary pairs can then undergo inverse Compton scattering or synchrotron radiation. Proton synchrotron radiation could also be a source of high-energy *gamma*-rays. Böttcher (2000) and Rachen (2000) provide overviews of these two models. Additionally, Böttcher et al. (2013) introduced a hybrid leptonic-hadronic model to explain the VHE spectra of TeV blazars. Below, we briefly introduce the leptonic and hadronic models (see Fig. 1.11).

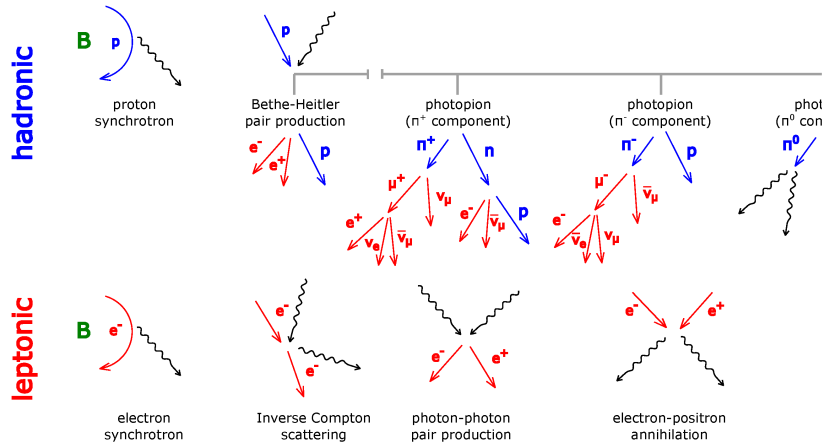


Figure 1.11: Schematic representation of the primary hadronic and leptonic radiative processes in the emission region (Mastichiadis, 2016).

Leptonic Model

In scenarios where protons are not accelerated to sufficiently high energies to reach the threshold energy for p - γ interactions to produce π mesons, high-energy radiation is dominated by Compton scattering of relativistic electrons. The leptonic model posits that the low-energy component of blazars' SEDs arises from synchrotron radiation of relativistic electrons, while the high-energy component is due to inverse Compton scattering of relativistic electrons with soft photons. These soft photons can either be synchrotron photons from within the jet, i.e., the SSC model (Bloom & Marscher, 1996a; Maraschi et al., 1992a; Zhang et al., 2012a), or external photons, i.e., the EC model (Georganopoulos et al., 2002). The external soft photon field could be radiation from the accretion disk (Dermer & Schlickeiser, 1993b), radiation scattered from the broad-line region (Sikora et al., 1994c), near-infrared radiation from the dust torus, or synchrotron radiation from other radiative zones (Błażejowski et al., 2000b). Both internal and external γ - γ absorption can modify the observed radiation spectrum. Internal γ - γ absorption can lead to the injection of new electron-positron pairs. In a self-consistent blazar leptonic model, synchrotron self-absorption must also be considered. The standard SSC model successfully explains the SEDs of most HBLs. However, it does not adequately explain the SEDs of FSRQs and ISP BL Lacs, which typically have dominant γ -ray emission. SSC

model fits suggest that the energy densities of electrons and magnetic fields are far from equipartition. To address this, a separate high-energy component has been introduced to explain the γ -ray emission of FSRQs. The EC model has more free parameters than the SSC model, and the γ -ray emission of FSRQs may be contributed by multiple EC components. However, the strength of each EC component depends on the distance of the radiative zone from the central black hole. Since the location of the γ -ray emitting zone in FSRQs is still uncertain, the physical parameters in the FSRQ jet are not well-constrained. To better constrain the physical parameters in the FSRQ jet, Dermer et al. (2014) proposed fitting FSRQs' SEDs using an equipartition approach. Fits to FSRQs' SEDs suggest that they are generally close to equipartition, while BL Lacs deviate slightly from equipartition ($U_e/U_B \sim 10 - 100$). The equipartition approach may explain the blazar sequence.

Steady-state leptonic models typically assume a known electron spectrum, either a power-law or a broken power-law. Based on this assumption, leptonic models can fit most blazars' SEDs well (Ghisellini et al., 2010). However, this approach lacks a self-consistent treatment of the electron spectrum. A more realistic approach considers the temporal evolution of the electron spectrum, including a self-consistent treatment of particle injection, acceleration, cooling, and escape. To explain the SEDs and variability patterns of blazars, it is essential to solve the time-dependent electron dynamics and radiative transfer problems. Time-dependent Synchrotron Self-Compton (SSC) models have successfully explained many blazars' SEDs and variability patterns (Kataoka et al., 2000; Li & Kusunose, 2000; Mastichiadis & Kirk, 1997; Sokolov et al., 2004). External radiation fields have also been included in this manner (Böttcher & Dermer, 2002; Sikora et al., 2001; Sokolov & Marscher, 2005). Homogeneous leptonic models explain many blazars' SEDs and correlate multi-wavelength variability well. However, some new observational results challenge the homogeneous leptonic model, prompting the development of inhomogeneous leptonic models. For example, in HBL 1ES1959+650 (Krawczynski et al., 2004) and MrK 412 (Błażejowski et al., 2005), X-ray and γ -ray emissions are uncorrelated. In PKS 2155304, X-rays and TeV γ -rays are correlated, but optical and X-rays are not, along with minute-scale γ -ray variability (Costamante, 2008). To overcome these challenges, several inhomogeneous leptonic models have been proposed, including the spine-sheath model (Ghisellini et al., 2005), the jet deceleration model (Georganopoulos & Kazanas, 2003), the internal shock scenario (Böttcher & Dermer, 2010; Joshi & Böttcher, 2011; Spada et al., 2001), and the turbulent multi-zone model (Marscher, 2014).

Hadronic Model

The hadronic model can be divided into proton-proton (p-p) interactions and proton-photon (p- γ) interactions. If protons are accelerated to extremely high energies to reach the threshold energy for p- γ or p-p interactions to produce π mesons, high-energy protons will interact with surrounding matter or background photons to produce high-energy *gamma*-rays. The core idea of the hadronic model is that high-energy *gamma*-rays are produced by proton-induced cascades. The p-p interaction has been used to explain the high-energy radiation of blazars. In this model, it is generally assumed that high-energy protons interact with clouds in the broad-line region to produce π^0 , which decays to produce the observed *gamma*-rays. The p-p interaction generally requires a relatively high neutral hydrogen number density ($n_H > 10^{11} \text{cm}^{-3}$). However, such a high number density is challenging to achieve in the environment of blazars. Therefore, in the environment of blazars, the efficiency of the p-p interaction will be much lower than that of the p- γ interaction, with the p- γ interaction being dominant. For BL Lacs, since there is no strong external photon field around them, high-energy protons mainly interact with synchrotron radiation photons. For FSRQs, which typically have a strong external photon field around them, high-energy protons can interact with both synchrotron radiation

photons within the jet and external soft photons to produce high-energy *gamma*-rays. The hadronic model generally believes that the low-energy component of the SEDs still comes from the primary electron synchrotron radiation, while the high-energy component comes from proton synchrotron radiation, photons from π^0 decay, and synchrotron and inverse Compton radiation from secondary particles. To accelerate protons to extremely high energies, the hadronic model generally requires a relatively high magnetic field, ranging from several tens to hundreds of Gauss.

1. Proton Synchrotron Model (SPB): Shock waves accelerate both electrons and protons simultaneously. The low-energy component of the SEDs is produced by primary electron synchrotron radiation, while the high-energy component is produced by proton synchrotron radiation. The SPB model requires a relatively high magnetic field, ranging from several tens to hundreds of Gauss, to accelerate protons to extremely high energies. The SPB model can explain the SEDs of some blazars, especially those with a dominant high-energy component. However, the SPB model requires a relatively high magnetic field, which is difficult to achieve in the environment of blazars. Therefore, the SPB model is not widely accepted.
2. Proton-Photon Model (PPB): The PPB model posits that high-energy protons interact with background photons to produce π^0 , which decays to produce the observed *gamma*-rays. The PPB model can explain the SEDs of some blazars, especially those with a dominant high-energy component. However, the PPB model requires a relatively high neutral hydrogen number density ($n_{HH} > 10^{11} \text{cm}^{-3}$), which is difficult to achieve in the environment of blazars. Therefore, the PPB model is not widely accepted.
3. Proton-Photon Model (PGB): The PGB model posits that high-energy protons interact with background photons to produce π^0 , which decays to produce the observed *gamma*-rays. The PGB model can explain the SEDs of some blazars, especially those with a dominant high-energy component. However, the PGB model requires a relatively high neutral hydrogen number density ($n_{HH} > 10^{11} \text{cm}^{-3}$), which is difficult to achieve in the environment of blazars. Therefore, the PGB model is not widely accepted.
4. Proton-Induced Cascade Cluster Radiation Model (PIC): Since the p-p interaction does not need to be considered, the PIC model mainly involves the p- γ interaction. The gyroradius of the proton is $R_L = 3.3 \times 10^{15} B_1^{-1} E_{19} \text{cm}$, where $B_{10} = 10 B_1 \text{G}$ and $E_{19} = 10^{19} E_1 \text{eV}$. The light variation time scale indicates that the typical scale of the radiation zone is $R_b \leq 10^{16} \text{cm}$. The condition $R_L < R_b$ indicates that the maximum energy that protons can be accelerated to is $E_p^{max} \geq 10^{19} \text{eV}$. When the energy of the proton reaches the threshold energy for π meson production through the p- γ interaction, the proton-induced cascade process will begin. The processes of π meson production in the p- γ interaction are:

$$p + \gamma \rightarrow p + \pi^0 \quad (1.2)$$

$$p + \gamma \rightarrow n + \pi^+ \quad (1.3)$$

The decay processes of the π meson are:

$$\pi^+ \rightarrow \mu^+ + \nu_\mu, \quad (1.4)$$

$$\mu^+ \rightarrow e^+ + \nu_e + \bar{\nu}_\mu, \quad (1.5)$$

$$\pi^0 \rightarrow 2\gamma, \quad (1.6)$$

$$\pi^0 \rightarrow 3\gamma \quad (\text{less common}). \quad (1.7)$$

Lepto-Hadronic Model

The lepto-hadronic model is a combination of the leptonic and hadronic models. In this model, both electrons and protons are accelerated to relativistic energies, and both contribute to the observed radiation. The lepto-hadronic model can explain the SEDs of some blazars, especially those with a dominant high-energy component. However, the lepto-hadronic model requires a relatively high magnetic field, which is difficult to achieve in the environment of blazars. Therefore, the lepto-hadronic model is not widely accepted.

The pathways leading to the production of high-energy γ -rays have been subject to various interpretations by different authors. One perspective, exemplified by Cao & Wang (2014), posits that the origin of high-energy γ -rays can be attributed to the synchrotron self-Compton (SSC) mechanism involving electrons, as well as directly from the decay of π mesons into secondary gamma photons. Another hypothesis, proposed by Cerruti et al. (2015a), suggests that all particles resulting from the decay of π mesons contribute to the radiation in the emission region.

A further viewpoint considers that some high-energy photons may originate from high-energy protons escaping the emission region and interacting with the cosmic microwave background (CMB) photons through p - γ interactions. This initiates a new cascade of gamma photons, which may persist or be considered only for the first-generation secondary electrons that undergo inverse Compton scattering with CMB photons (Yan & Zhang, 2015; Zheng et al., 2016).

The hadronic processes imply that blazars could be sources of neutrino emissions. On September 22, 2017, the IceCube Neutrino Observatory detected for the first time a high-confidence (3σ) neutrino event associated with the flaring activity of the blazar TXS 0506+056 in both time and space across multiple wavelengths (IceCube Collaboration et al., 2018c). This finding substantiates the role of blazars as one of the primary extragalactic sources of high-energy neutrinos. Ch. 7 of this dissertation will discuss in detail the methodologies employed to identify additional potential neutrino-emitting blazars.

1.3.6 Jet of Blazars

Blazars are a distinct subclass of AGN characterized by jets with extreme relativistic velocities. These jets, with small opening angles, are oriented either directly or nearly directly towards the observer. Due to the relativistic beaming effect, the radiation from the jet is significantly amplified, dominating the overall emission. The jet's luminosity often surpasses the luminosity of the entire host galaxy, with velocities sometimes exceeding $0.999c$. Consequently, observations and studies of blazars provide valuable insights into the characteristics of these jets.

Particle Acceleration Mechanisms in Jets

As relativistic shock waves traverse the jet, they accelerate particles within. Depending on the orientation of the background magnetic field concerning the shock normal, these shocks can be classified as parallel or oblique. Particles can be accelerated through two primary Fermi acceleration mechanisms. The first-order Fermi acceleration mechanism posits that particles gain energy by repeatedly crossing the shock front between the downstream and upstream regions of the shock. The energy gain is proportional to the particle's velocity, i.e., $\Delta E/E \propto u/c$. The second-order Fermi acceleration mechanism, a stochastic process associated with plasma turbulence, suggests that particles gain energy through reflections by moving magnetic mirrors. The energy gain in this case is proportional to the square of the mirror's velocity, i.e., $\Delta E/E \propto (u/c)^2$. Relativistic parallel shocks produce a power-law injection spectrum $N(\gamma) \propto \gamma^{-q}$ with an injection index $q \approx 2.2 - 2.3$ (Achterberg et al., 2001).

Oblique shocks yield a softer spectral index $q > 2.3$ (Ostrowski & Bednarz, 2002). The second-order Fermi acceleration might produce a harder spectral index, $q \approx 1$ or slightly greater (Virtanen & Vainio, 2005). The spectral indices q adopted in fitting blazar SEDs and light curves may reveal the dominant acceleration mechanism within the blazar jet.

Direct X-ray and high-energy γ -ray observations indicate that particles within blazar jets can be accelerated to very high energies. However, the location of particle injection into the radiation zone and the subsequent acceleration mechanisms remain elusive. Several theories regarding particle injection/acceleration have been proposed, including impulsive injection near the base of the jet (Dermer & Schlickeiser, 1993b; Dermer et al., 1997), isolated shocks propagating along the jet (Kirk et al., 1998; Marscher & Gear, 1985; Sikora et al., 2001; Sokolov et al., 2004), internal shocks produced by collisions between multi-shell materials within the jet (Spada et al., 2001; Böttcher & Dermer, 2010; Joshi & Böttcher, 2011), random acceleration in relativistic shear layers (Rieger & Duffy, 2004), and magnetic reconnection in Poynting-flux dominated jets (Sikora et al., 2005).

Observations of Jets

The morphology of AGN jets varies across different observational wavelengths, as illustrated in Fig. 1.12. Structurally, jets comprise distinct features such as a bright core, jet body, lobes, and hot spots. Some jets, as depicted in Fig. 1.13, also exhibit distinct knots or nodes in radio observations. While jets are generally considered to be highly collimated, recent studies suggest the presence of a pronounced conical structure (Kishimoto et al., 2022). Additionally, some research indicates that jets may possess a parabolic structure (Nokhrina et al., 2022).

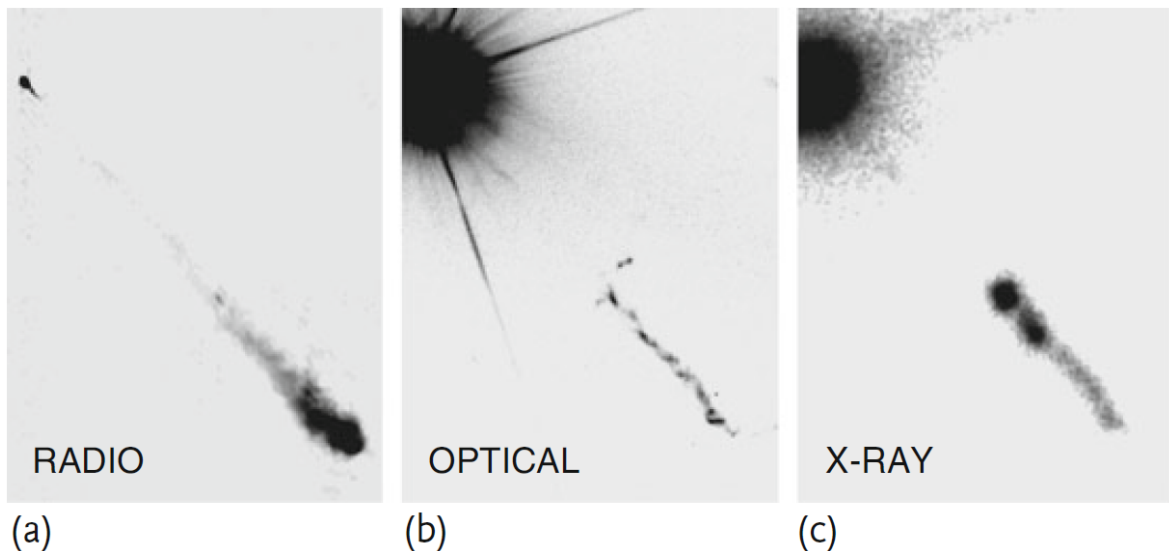


Figure 1.12: The renowned jet of quasar 3C 273 is depicted through multi-wavelength observations: (a) the radio frequency image at 1.6 GHz sourced from the MERLIN archive; (b) the optical representation captured by the Hubble Space Telescope at a wavelength of 6000 \AA ; and (c) the X-ray visualization derived from a series of observations by the Chandra X-ray Observatory, processed to highlight the energy range of 0.3–6 keV and subsequently subjected to minor smoothing for clarity (Boettcher et al., 2012).

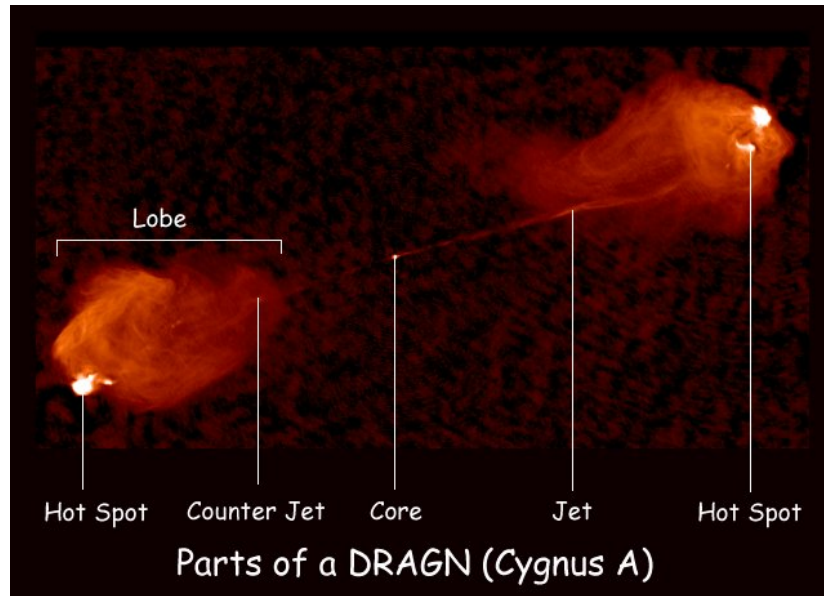


Figure 1.13: Lobe and knot of Cygnus A. Figure courtesy of <http://www.cv.nrao.edu/abridle/images/knot.pdf>.

Superluminal Motion in Jets

Radio VLBI observations have revealed that the structure of radio jets is not continuous but consists of multiple sub-sources or knots. These sub-sources are not stationary but exhibit motion. By comparing the distances between various knots and their propagation times in jet images taken at different epochs, some jets exhibit apparent superluminal motion, a visual illusion caused by special relativistic effects. Using radio spectra taken at different times, the angular displacement $\Delta\theta$ of a sub-source can be calculated. Knowing the distance d , the angular distance can be converted to a linear distance: $\Delta\theta = d\Delta\theta$. Given the observational time interval Δt_a , the apparent velocity is: $v_{app} = d\Delta\theta/\Delta t_a$. For some AGN, this velocity exceeds the speed of light c , manifesting as superluminal motion. Fig. 1.14 provides a schematic representation explaining superluminal motion. This topic will be discussed in detail in Ch. 5.

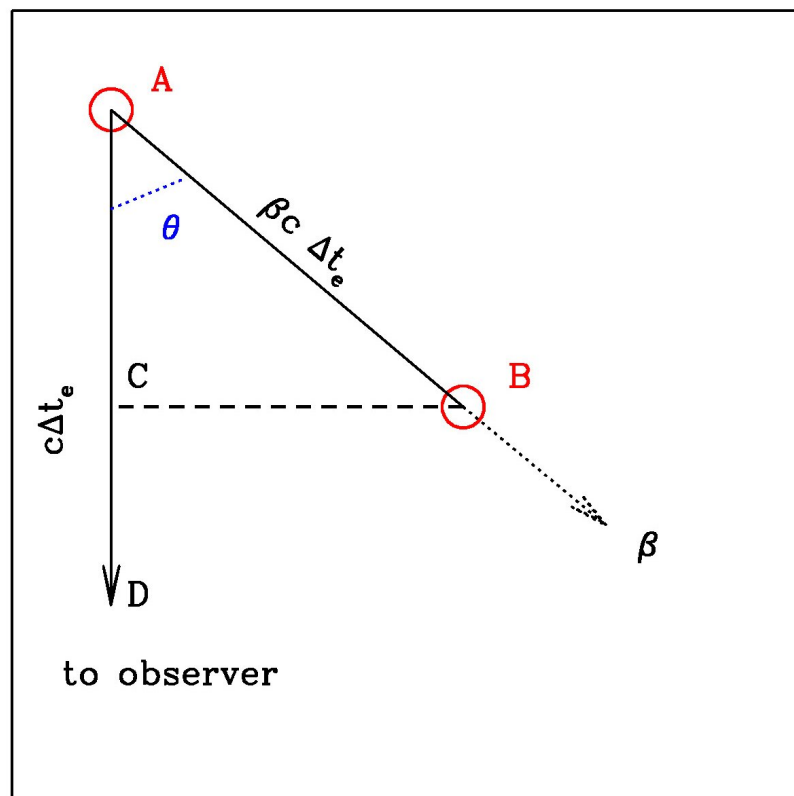


Figure 1.14: Schematic representation of superluminal motion (Ghisellini, 2013). A sub-source moves at a speed of βc . In the comoving frame, it moves from point A to point B in time Δt_e , emitting photons. The time interval between photon receptions at point D is $\Delta t_a = \Delta t_e / \delta$.

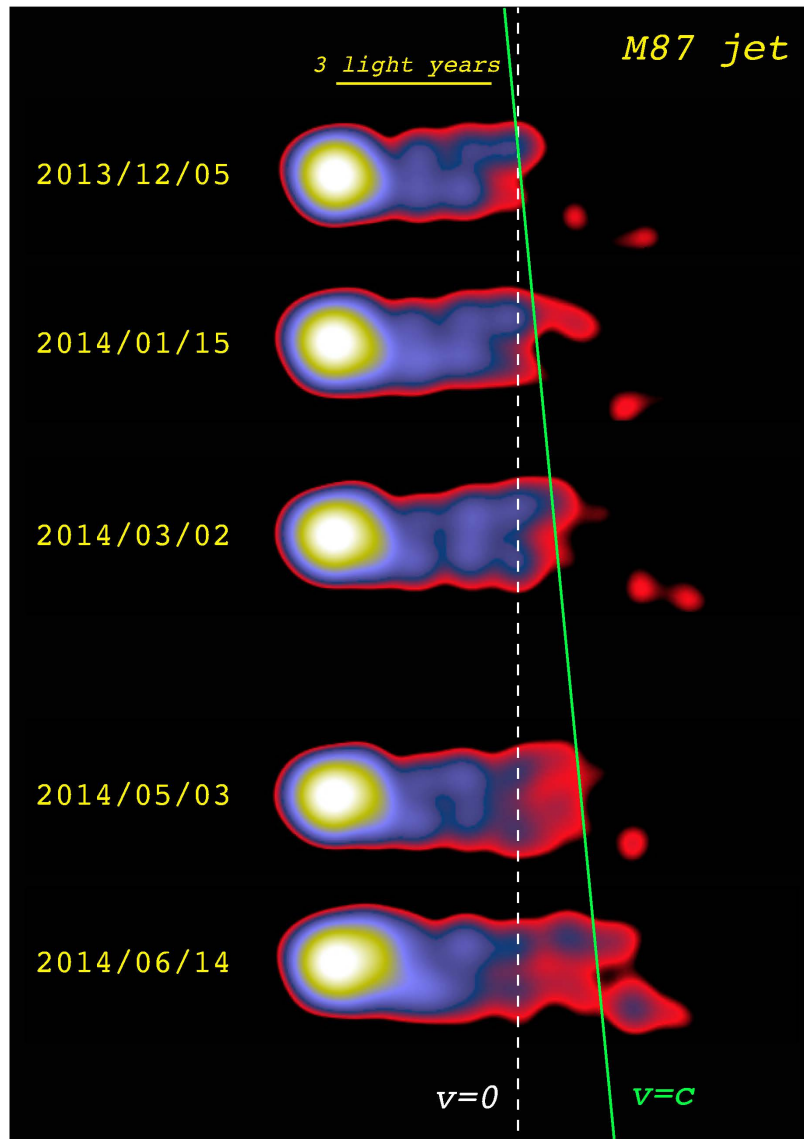


Figure 1.15: Multi-epoch KaVA 22 GHz images of the M87 jet. All images are convolved with a 1 mas circular beam. The green line represents a constant velocity at the speed of light (Hada, 2017).

Chapter 2

Fermi telescope and *Fermi* source catalogs

2.1 Historical Context and Genesis of the *Fermi* Mission

- **Overview of the *Fermi* γ -ray Space Telescope.** The *Fermi* γ -ray Space Telescope (*FGST*¹), formerly the γ -ray Large Area Space Telescope (*GLAST*), is a premier observatory for γ -ray astronomy. Launched on June 11, 2008, *Fermi* has been surveying the sky in γ -rays, which are the most energetic form of light. The telescope's observations cover an energy range from 8 keV to over 300 GeV, revealing a universe that is vastly different from the one visible to the naked eye. The Large Area Telescope, its primary instrument, has been instrumental in advancing our understanding of cosmic phenomena, from the behavior of supermassive black holes in active galaxies to the origins of cosmic rays and the properties of dark matter (Atwood et al., 2009a; Ajello et al., 2020; Abdo et al., 2009; Ackermann et al., 2015b). Fig. 2.1 shows the γ -ray sky seen by *FGRST*
- **Historical context of the *Fermi* mission.** The *Fermi* mission, named after the renowned physicist Enrico Fermi, is a collaborative effort involving NASA, the United States Department of Energy, and international partners. It follows in the footsteps of previous γ -ray observatories, offering unprecedented sensitivity and resolution. The mission's scientific objectives are broad, encompassing the study of particle acceleration mechanisms, the resolution of the γ -ray sky, and the exploration of fundamental physics questions (Atwood et al., 2009a; Abdo et al., 2009; Ackermann et al., 2015b; Carr et al., 2010). *Fermi*'s Large Area Telescope (LAT) has provided critical data on blazars, γ -ray bursts, and the Milky Way, contributing to a deeper understanding of these complex phenomena (Ajello et al., 2020; Lott et al., 2020; Abdo et al., 2009; Berenji & Siegal-Gaskins, 2014).
- **Genesis of the *Fermi* mission.** The *Fermi* telescope's journey began with its assembly and prelaunch activities at the Astrotech payload processing facility in Titusville, Florida². After a successful launch, *Fermi* entered a low-Earth orbit, ready to embark on its mission to map the γ -ray sky. The mission has not been without its challenges, including a close encounter with space debris, but through careful management and software upgrades, such as the Pass 8 data analysis enhancement³, *Fermi* has continued to operate effectively. Its ongoing mission promises to further our understanding of the high-energy universe, with the potential for new

¹<https://fgst.slac.stanford.edu/>

²<https://www3.nasa.gov/centers/kennedy/news/releases/2008/release-20080304.html>

³https://fermi.gsfc.nasa.gov/ssc/data/analysis/documentation/Pass8_usage.html

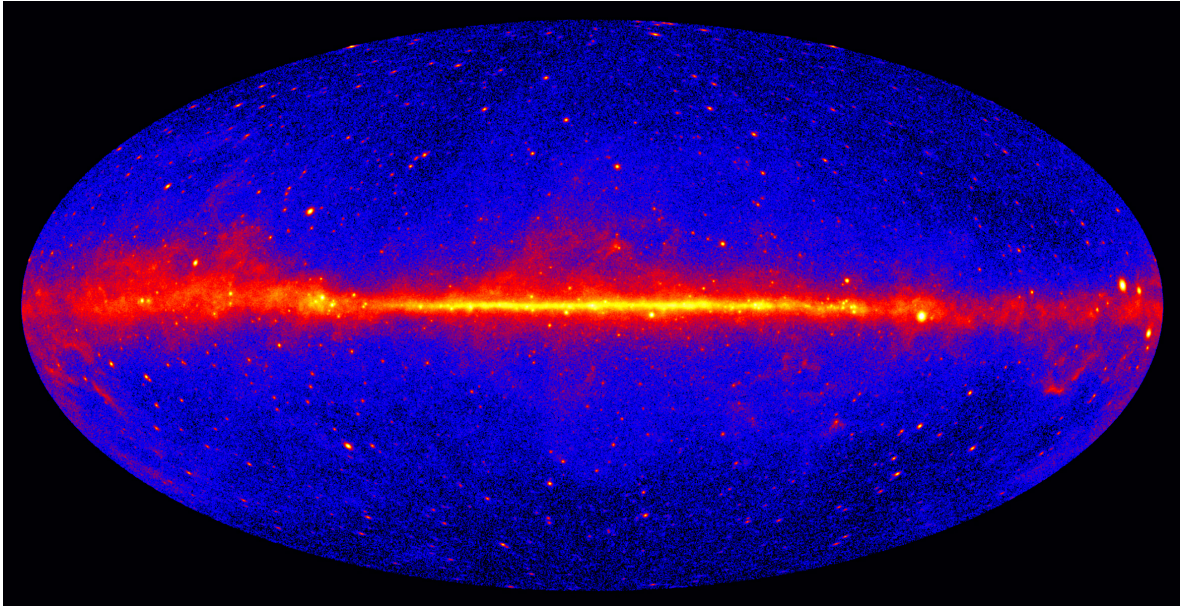


Figure 2.1: The γ -ray sky in the view of *FGST* beyond 1 GeV within 12 years. Figure courtesy of https://space.mit.edu/HETG/Reports/HETG_Report_SciJun02.html.

and unexpected discoveries (Acero et al., 2015; Ajello et al., 2017; Ajello et al., 2020; Lott et al., 2020; Abdollahi et al., 2020; Ballet et al., 2020; Abdollahi et al., 2022; Ballet et al., 2023).

2.2 *Fermi* instruments

The FGST is equipped with a pair of scientific instruments: the Large Area Telescope (LAT, Atwood et al. 2009a), to perform a comprehensive survey of the entire sky in high-energy γ -rays, studying a wide range of cosmic phenomena, from black holes to dark matter. and the Gamma-ray Burst Monitor (GBM, Meegan et al. 2009), to detect and analyze γ -ray bursts and other transient events, providing rapid notifications and detailed observations of these fleeting high-energy occurrences. Fig. 2.2 is a schematic diagram of *Fermi*-LAT and *Fermi*-GRB.

2.2.1 *Fermi*-LAT

The *Fermi*-LAT is the principal scientific instrument on the FGST, which was launched into orbit on June 11, 2008. As a powerful γ -ray observatory, the LAT's mission is to survey the high-energy sky, mapping and studying γ -ray sources, which include not only γ -ray bursts but also other phenomena such as active galactic nuclei, pulsars, and diffuse γ -ray emissions.

- **Specific requirements.**

1. Due to the presence of numerous variable sources of γ -rays radiation in the cosmos, LAT must have a large field of view, exceeding 2 square degrees (approximately one-fifth of the entire sky).
2. To precisely identify and study sources, LAT must be capable of measuring the positions of sources within 1 arcminute (approximately 1/30th the diameter of the full moon).

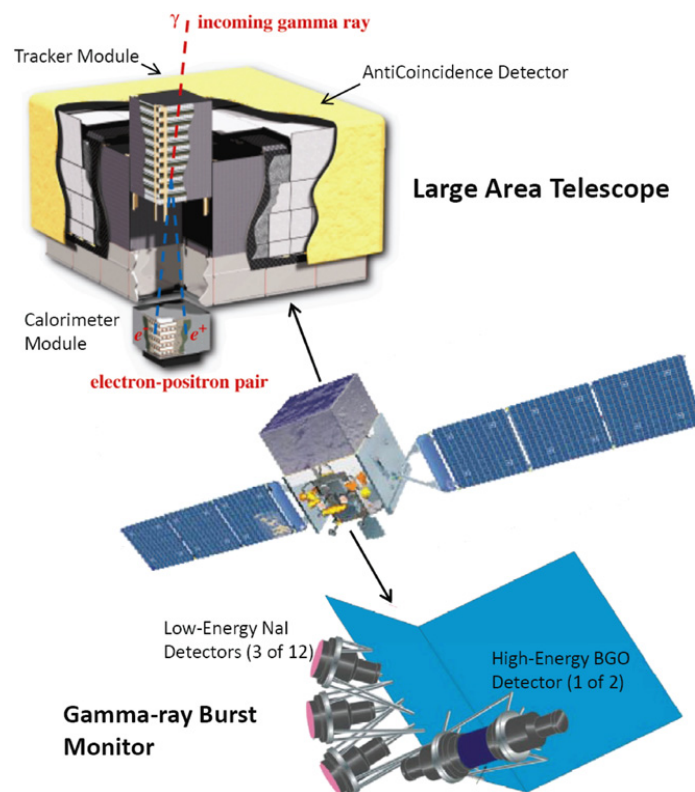


Figure 2.2: The FGST and its two instruments. The LAT images the sky in the energy band from about 20 MeV to more than 300 GeV while the GBM complements the LAT for the study of GRBs and transients, providing spectral coverage from 8 keV to about 40 MeV (Michelson et al., 2010).

3. The study of gamma rays spans a wide energy range; therefore, LAT's energy coverage must extend from 8 keV to 300 GeV. Additionally, LAT's sensitivity will extend beyond 10 GeV, as very little is known about cosmic objects at these energies.
4. γ -ray bursts (GRBs) can release a significant amount of gamma rays within fractions of a second. Consequently, LAT must be able to measure γ -rays emissions within short time intervals.

- **Instrumentation and Capabilities.** The *Fermi*-LAT is a marvel of modern astrophysics instrumentation, featuring a modular design of a 4×4 array of identical towers. Each tower measures $40 \times 40 \text{ cm}^2$ and is a composite of a tracker, a calorimeter, and a data acquisition module. The tracker, consisting of 18 xy layers of silicon strip detectors, boasts a long and successful history in high-energy physics experiments. It achieves high detection efficiency ($\approx 99\%$), exceptional position resolution ($< 60 \mu\text{m}$), a strong signal-to-noise ratio ($> 20:1$), minimal cross-talk, and efficient triggering and readout processes without the need for consumables.

The calorimeter, integral to each tower, comprises eight layers of cesium iodide (CsI) bars arranged hodoscopically and read out by photodiodes, amounting to a total thickness of 10 radiation lengths. This hodoscopic configuration allows for the three-dimensional reconstruction of particle showers, aiding in energy leakage correction and enhancing the discrimination against hadronic cosmic rays. The anticoincidence shield, enveloping the array of towers, utilizes segmented scintillator tiles read out by wavelength-shifting fibers and miniature phototubes, ensuring that the LAT is sensitive primarily to γ -rays and not charged particles.

The *Fermi*-LAT instrument comprises three main subsystems (see Fig. 2.3): the Anti-coincidence Detector (ACD), the Tracker (TKR), and the Calorimeter (CAL). Additionally, it is equipped with a Data Acquisition System (DAQ) for data collection and processing.

1. *Anti-coincidence Detector.* The Anti-coincidence Detector (ACD) serves as the first line of defense against cosmic-ray-induced false events in the LAT (Moiseev et al., 2007). Its primary purpose is to tag incoming particles before they enter the telescope. The ACD consists of 89 plastic scintillator tiles, each 1 cm thick, which produce flashes of light when charged cosmic-ray particles (but not electrically neutral gamma rays) strike them. This allows the LAT to distinguish between charged cosmic rays and passing gamma rays. The ACD is segmented into different sizes as it approaches the calorimeter, with the top surface composed of a 5×5 array of $32 \text{ cm} \times 32 \text{ cm}$ plastic scintillator tiles. The side sections include 4×4 arrays of scintillator tiles, with the third row reducing to $15 \text{ cm} \times 32 \text{ cm}$, and the bottom row consisting of a single large tile measuring $170 \text{ cm} \times 170 \text{ cm}$. Each scintillator tile is read out by two photomultiplier tubes (PMTs) located at the bottom of the ACD, providing high detection efficiency for gamma rays (0.9997).
2. *Tracker.* The Tracker (TKR) plays a crucial role in determining the direction of γ -ray photons and accurately measuring the paths of electrons and positrons produced by initial γ -ray interactions (Belli et al., 2007). It consists of 18 x, y planes (a total of 36 layers) of silicon-strip tracking detectors (SSDs), with the first 16 layers interleaved with tungsten converter foils. The TKR is physically divided into 19 trays, with the top and bottom trays containing parallel SSD layers in the y-direction. The middle 17 trays consist of two parallel SSD planes, alternating vertically between trays, providing x and y position measurements for particles traversing the detector. The top 12 trays (front section) contain thin tungsten converter foils located just above the uppermost detector layers, with

a thickness of $0.035 X_0$ (radiation lengths). The back 4 trays consist of thicker tungsten converter foils ($0.18 X_0$), and the bottom 3 trays do not contain any converter material. Under normal incidence, the detector's total depth is $1.5 X_0$, allowing for the detection of electromagnetic particles cascading through.

3. *Calorimeter.* The Calorimeter (CAL) measures the energy of gamma photons and is designed to provide precise energy measurements and reject cosmic rays, which deposit energy differently from gamma rays. The calorimeter consists of 8 layers of thallium-doped cesium iodide (CsI) crystals (Moskalenko et al., 2007). Each crystal is optically isolated and read out at both ends by two PIN photomultiplier tubes (PMTs), generating fluorescence. The 8 layers of crystals in CAL are aligned alternately in the x and y directions, forming a star-shaped array. The calorimeter has a depth of $8.6 X_0$ for vertically incident particles, allowing it to capture γ -ray energies up to 1 TeV. CAL provides three spatial coordinates: the x and y coordinates of the crystal and the position of the photon crossing the crystal axis. The latter is determined by measuring the intensity ratio of light received at both ends of the crystal using two PMTs with significantly different gains, enabling position measurements accurate to less than a millimeter for gamma photons with energies exceeding 1 GeV. The LAT instrument consists of a total of 1536 crystals, with each module measuring 1080 cm in its active range. CAL achieves an energy resolution of less than 10% at 10 GeV and less than 6% at 300 GeV.
4. *Data Acquisition System.* The Data Acquisition System (DAQ) is the central nervous system behind LAT and includes the Global-trigger/ACD-module/Signal-distribution Unit (GASU), the Global Trigger Module (GTM), the Event Builder Module (EBM), and the Command Response Unit (CRU). The Tower Electronics Module (TEM) facilitates triggering and data readout for each module in the Tracker and Calorimeter. Two primary Event Processing Units (EPUs) and one primary Spacecraft Interface Unit (SIU) handle data collection from the Tracker, Calorimeter, and Anti-coincidence Detector, providing preliminary discrimination between unwanted signals from cosmic rays and genuine γ -ray signals to determine which signals should be transmitted to the ground. The DAQ system is composed of dedicated electronics and microprocessors.

The Fermi LAT instrument operates cohesively, utilizing these subsystems and the DAQ system to capture high-energy γ -ray events, enabling the exploration of the high-energy universe.

- **Scientific Objectives.** *Gamma-ray Source Cataloging.* It aims to create a comprehensive catalog of γ -ray sources, contributing to our understanding of the high-energy universe.

Dark Matter Research. The LAT searches for indirect signals of dark matter, which could be revealed through excess γ -rays from annihilations or decays of dark matter particles.

Fundamental Physics. It tests models of fundamental physics, including searches for violations of Lorentz invariance or changes in the nature of space-time at small scales.

Transient Phenomena. The LAT monitors the sky for transient events, such as γ -ray bursts and flares from blazars, providing crucial data for multi-wavelength and multi-messenger astronomy.

- **Operational Aspects.** The LAT operates autonomously, with events causing detector hits in three planes automatically triggering readouts of each tower and the anticoincidence system. Data analysis is a complex process involving the calibration of detectors, signal extraction, and

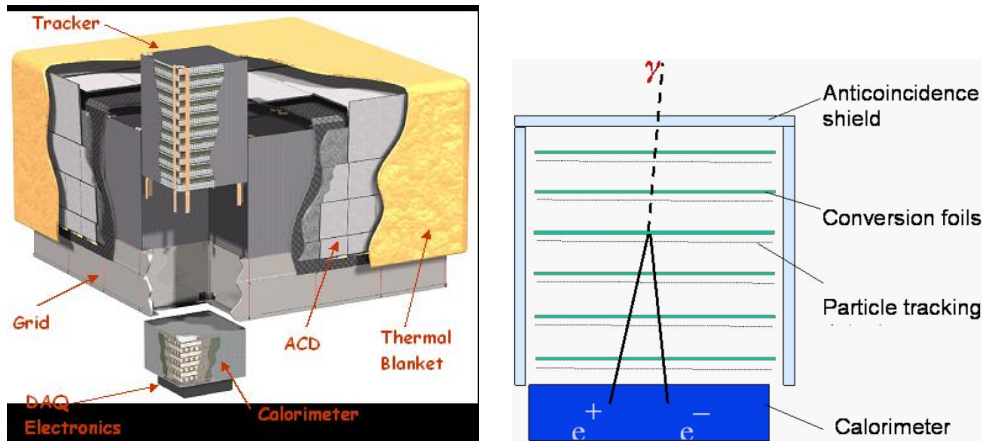


Figure 2.3: Schematic representation of the LAT instrument's cross-section, illustrating the annihilation of incident photons into e^+ and e^- . Figure courtesy of https://fermi.gsfc.nasa.gov/ssc/data/analysis/documentation/Cicerone/Cicerone_Introduction/LAT_overview.html.

interpretation of the results. The LAT team collaborates with other missions and observatories for coordinated observations and has been instrumental in numerous discoveries and advancements in the field of high-energy astrophysics.

The LAT's design and capabilities have been validated through detailed computer simulations and tests with prototype towers at particle accelerators. Its data and the resulting scientific insights have significantly advanced our understanding of the γ -ray sky, providing an unprecedented view of the energetic universe.

2.2.2 *Fermi*-GRB

The *Fermi*-GBM is a pivotal instrument aboard the Fermi Gamma-ray Space Telescope, to observe the cosmos in γ -rays light. The GBM's primary role is to work in concert with the Large Area Telescope (LAT) to detect and analyze γ -rays bursts (GRBs)—the universe's most luminous and energetic events that occur during cataclysmic occurrences such as the collapse of massive stars or the collision of dense objects like neutron stars or black holes.

- **Specific requirements.**

1. GRBs emanate from arbitrary directions in the sky; thus, GBM must observe the entire sky as much as possible at all times.
2. To gather more information about GRBs, GBM should be capable of measuring photon energies over a wide range, from as low as 8 keV to energies overlapping with LAT's range.

3. Since the durations of GRBs can vary from microseconds to several thousand seconds, GBM must be able to detect GRBs over a broad time range.

- **Instrumentation and Capabilities.** The GBM's detection suite includes 12 Sodium Iodide (NaI) scintillation detectors and 2 Bismuth Germanate Oxide (BGO) scintillation detectors. The NaI detectors are attuned to γ -rays in the lower energy spectrum, ranging from about 8 keV to 1 MeV, providing burst triggers and localization. The BGO detectors extend the energy coverage from approximately 150 keV to 30 MeV. This arrangement not only ensures a seamless overlap between the NaI and BGO detectors but also with the LAT for higher energy γ -rays. The GBM's detectors are strategically mounted to avoid obstructing the LAT's field of view or the spacecraft's solar panels, fitting snugly between the LAT and the spacecraft's shroud envelope. The BGO detectors are placed on opposite sides of the spacecraft, while the NaI detectors are arranged in four banks of three, allowing for comprehensive coverage of various azimuth and elevation angles. This configuration mirrors the capabilities of the BATSE large area and spectroscopy detectors but with an expanded energy range and a more compact collection area.
- **Scientific Objectives.** *GRB Detection.* Its foremost objective is to detect GRBs and ascertain their characteristics, such as duration, spectrum, and temporal profile.

Transient Events. The GBM is also vigilant for other transient occurrences, including solar flares, terrestrial γ -rays flashes, and magnetar bursts.

GRB Alerts. It provides real-time alerts for GRBs, facilitating immediate multi-wavelength follow-up observations that are crucial for understanding the bursts' progenitors and physics.

Cross-Correlation. In synergy with the *Fermi*-LAT, the GBM enables cross-correlation of low-energy and high-energy γ -rays observations, offering a comprehensive view of GRBs and other transient high-energy phenomena.

Operational Aspects. The GBM's operational framework involves meticulous data analysis, where the team calibrates the detectors, extracts γ -rays signals, and interprets the findings within the ambit of theoretical models. Collaboration is a cornerstone of the GBM's operation, as it partners with various missions and observatories for swift response and extensive monitoring. This collaborative approach is vital for multi-messenger astronomy, allowing for an integrated study of GRBs across the electromagnetic spectrum and, potentially, alongside gravitational wave and neutrino observations.

Through its sophisticated instrumentation and strategic design, the GBM has significantly propelled our comprehension of GRBs, unveiling novel burst attributes and deepening our insight into the extreme physical processes that unfold during these explosive events.

2.3 Fermi Data Processing

2.3.1 Instrument Response Functions

On August 4, 2008, the *Fermi*-LAT began its scientific operations. Approximately one year later, LAT data became publicly available through the Fermi Science Support Center (FSSC). The standard astronomical data analysis relies on a set of specialized scientific analysis tools developed, maintained, and made publicly accessible by the LAT team and FSSC, collectively referred to as ScienceTools. One critical component of these tools is the Instrument Response Functions (IRFs). The performance

of the detector is described using IRFs, which are influenced by factors such as the energy of incident photons, the incident angle, and the location of electron-positron pair conversion.

The LAT team interprets three key aspects of IRFs:

1. **Effective area**, $A_{\text{eff}}(s, E, \hat{v})$. This represents the probability of γ -ray conversion in a given set cross-sectional area for a given event selection s , photon energy E , and velocity direction \hat{v} .
2. **Point Spread Function (PSF)**, $P(\hat{v}'; E, \hat{v}, s)$. It denotes the probability density of reconstructing the velocity direction for a given event selection.
3. **Energy dispersion**, $D(E'; E, v, s)$. This indicates the probability density of measuring energy for a given event selection.

Instrument Response Functions are used in maximum likelihood analysis to predict the observed γ -ray spectrum $M(E', \hat{P}', s)$ based on the infrared spectra:

$$\iiint S(E, \hat{p}') A_{\text{eff}}(E, \hat{v}(t, \hat{p}), s) P(\hat{v}'(t, \hat{p}), s) E(\hat{v}(t, \hat{p}), s) dE dt d\Omega \quad (2.1)$$

The integration range includes the required time range for analysis, the solid angle in the LAT reference frame, and the energy range of LAT. It's important to note that IRFs can significantly vary within the LAT's Field of View (FOV).

2.3.2 Data Analysis Versions

Throughout Fermi-LAT's mission, as our understanding of LAT's performance and its in-orbit environment improved, the analysis software for γ -ray events has undergone several updates. There have been four major versions (Passes) of LAT's database: Pass 6, Pass 7, Pass 7 Reprocessed, and Pass 8. These updates primarily focused on reducing systematic errors in the IRFs. The latest Pass 8 version represents a comprehensive correction to the entire data analysis chain, resulting in increased effective area, improved PSF, and reduced background contamination. Another significant parameter is the observation time distribution in the LAT reference frame, closely related to exposure.

2.3.3 Effective Area Calculation

The effective area for a given energy and direction in the sky (E, \hat{P}, s) is defined as the integral of the instrument's effective area over the entire time range of interest:

$$\epsilon(E, \hat{p}, s) = \int A_{\text{eff}}(E, \hat{v}(t, \hat{p}, s)) dt \quad (2.2)$$

2.3.4 Event Reconstruction Versions (Passes)

Changes in event reconstruction or parameterization for event classes are referred to as "Passes," and they correspond to improvements in the instrument's response. For example, P7_SOURCE_V6 refers to the use of Pass 7 reconstruction for source-class data with version 6 of the response functions. These version changes affect the instrument's response functions but do not impact the basic data reconstruction or event types.

2.3.5 Evolution of Data Analysis

Since 2008, LAT data's event reconstruction and the choice of instrument response functions have undergone three versions: Pass 6 in 2008, Pass 7 in August 2011, and Pass 7 Reprocessed, which improved the γ -ray Galactic diffuse model, IRFs, and direction reconstruction for energies exceeding 3 GeV. Subsequently, in June 2015, Pass 8 (Ackermann et al. 2016) was released. Pass 8 significantly improved LAT's performance compared to Pass 7, including complete data reconstruction, reduced γ -ray background, increased effective area, improved PSF, more accurate directional measurements, and the ability to analyze data down to 60 MeV. However, in Pass 8 (version P8R2), only the ULTRACLEANVETO event class is fully isotropic, meaning that there are no further concerns about source-class anisotropy in LAT data reconstruction and selection. To address this issue, the Fermi group reclassified all LAT data and produced a new Pass 8 version (P8R3_V3, Ackermann et al. 2015c; Bruel et al. 2018) with improvements in source acceptance and reduced background, particularly with no significant isotropic background. These changes resulted in an improved PSF for γ -ray point sources.

2.3.6 The Importance of IRFs

It's crucial to emphasize that the correct interpretation of *Fermi*-LAT data relies heavily on a precise knowledge of the instrument's response to gamma rays. Any uncertainties or inaccuracies in the IRFs can lead to systematic errors in the analysis. Therefore, the Fermi group continually refines and updates the IRFs to improve the accuracy of data interpretation and the reliability of scientific results.

2.4 *Fermi*-LAT Gamma-Ray Source Catalogs

The *Fermi*-LAT has been a remarkable instrument in the field of γ -ray astronomy since its launch in 2008. One of its significant contributions to the scientific community is the creation of *Fermi*-LAT catalogs, which systematically document and classify γ -ray sources detected by the telescope. These catalogs have played a pivotal role in advancing our understanding of high-energy astrophysics, and in this section, we will explore the evolution of these catalogs from the first 1FGL to the latest release, 4FGL-DR4.

The *Fermi*-LAT has been a remarkable instrument in the field of γ -ray astronomy since its launch in 2008. One of its significant contributions to the scientific community is the creation of *Fermi*-LAT catalogs, which systematically document and classify γ -ray sources detected by the telescope. These catalogs have played a pivotal role in advancing our understanding of high-energy astrophysics, and in this section, we will explore the evolution of these catalogs from the first to the latest release, 4FGL-DR4.

2.4.1 The First *Fermi*-LAT Source and AGN Catalog

1. **The First *Fermi*-LAT Source Catalog (1FGL).** The journey of *Fermi*-LAT catalogs began with the 1FGL (Abdo et al., 2010c), which was released in 2010. This catalog marked a significant milestone as it provided the first comprehensive list of γ -ray sources detected by *Fermi*-LAT. The 1FGL catalog contained 1451 sources (see Fig. 2.4) in the energy range from 100 MeV to 100 GeV, including 685 blazar candidates. Each source was characterized by its position in the sky, spectrum, and variability information. The 1FGL catalog was invaluable for researchers as it identified a wide variety of γ -ray emitters, including AGN, pulsars, supernova remnants,

and more. It allowed astronomers to systematically study these objects in the γ -ray regime, providing a wealth of data for multi-wavelength and multi-messenger studies.

2. **The First *Fermi*-LAT AGN Catalog (1LAC).** Before the release of the 1LAC catalog, Abdo et al. (2009) studied the first batch of bright AGNs detected by *Fermi*-LAT in the initial three months of its survey (the LAT Bright AGN Sample, LABS), which included 58 FSRQs, 42 BL Lacs, 2 radio galaxy nuclei, and 4 AGNs of unknown type. Thus, the first AGN catalog detected by *Fermi*-LAT, the 1LAC, collected data over 11 months. However, 1LAC is also a subset of 1FGL. It comprises 671 γ -ray sources located at high Galactic latitudes ($|b| > 10^\circ$), each with a Test Statistic value greater than 25. Some LAT sources are associated with multiple AGNs, so the actual catalog includes 709 AGNs, encompassing 300 BL Lac objects, 296 FSRQs, 41 AGNs of other types, and 72 AGNs of unknown type (as shown in Fig. ??). 1LAC provides a detailed introduction to the 1LAC sample and summarizes its characteristics, including the distribution of γ -ray flux, γ -ray photon spectral index, γ -ray variability, redshift distribution, and γ -ray luminosity. It also describes the multi-wavelength properties of 1LAC AGNs from radio to TeV energies. For more details, see Abdo et al. (2010a).

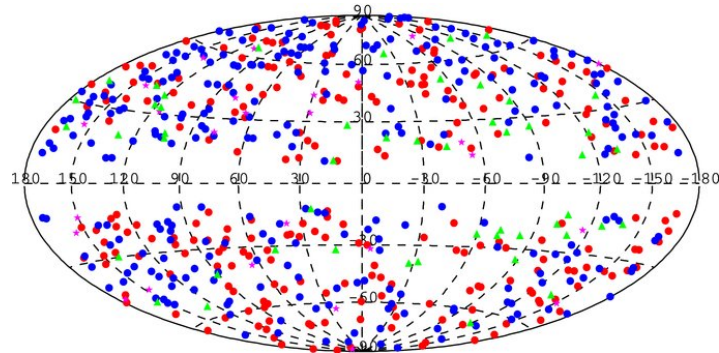


Figure 2.4: The clean sample's source locations are depicted as follows: FSRQs are marked with red circles, BL Lacs with blue circles, radio galaxies with magenta stars, and AGNs whose types are not identified are indicated by green triangles. (Abdo et al., 2010a).

2.4.2 The Second *Fermi*-LAT Source and AGN Catalogs

1. **The Second *Fermi*-LAT Source Catalog (2FGL).** Nearly a year after the release of the first clean sample (1LAC), Tosti et al. (2011); Nolan et al. (2012a) published the 2FGL data. The 2FGL data is based on the first 24 months of operations starting from August 4, 2008. A detailed comparison between 2FGL and 1FGL is provided. The isotropic galactic diffuse model used in the 2FGL analysis was improved compared to the 1FGL detection period. Based on the average flux detected over the first 24 months, 2FGL includes 1873 sources detected or characterized in

the 100 MeV to 100 GeV range, with 127 sources confirmed and the remaining 1170 associated or correlated with known or potential γ -ray emitters. The catalog was updated in April 2012, for more details see Nolan et al. (2012a).

2. **The Second *Fermi*-LAT AGN Catalog (2LAC).** Four months after the release of 2FGL, 2LAC published a clean sample based on 2FGL data, including 886 AGNs, comprising 395 BL Lacs, 310 FSRQs, 157 unclassified blazar candidates (i.e., with broadband blazar characteristics but without observed optical spectral features), 8 misaligned AGNs, 4 narrow-line Seyfert 1s (NLS1s), 10 AGNs of other types, and 2 starburst galaxies (see Fig. 2.5). Almost all FSRQs have a synchrotron peak frequency $<10^{14}$ Hz, but about half of the BL Lacs have a synchrotron peak frequency $>10^{15}$ Hz. Compared to the 1LAC catalog, the number of associated sources in 2LAC increased by 52%. Therefore, a comprehensive description of the newly detected sources will require more broadband data. The general trends observed in 1LAC were confirmed (Ackermann et al., 2011).

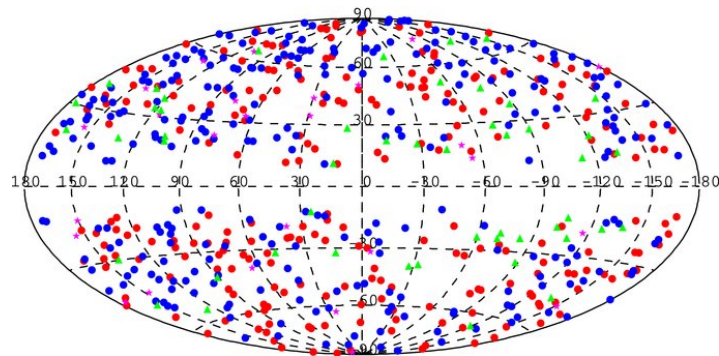


Figure 2.5: The positions of sources in the 2LAC clean sample. Red circles: FSRQs; Blue circles: BL Lac objects; Magenta circles: non-blazar AGNs; Green triangles: AGNs of unknown type (Ackermann et al., 2011).

2.4.3 The Third *Fermi*-LAT Source and AGN Catalogs

1. **The Third *Fermi*-Large Area Telescope Source Catalog (3FGL).** The *Fermi*-LAT Collaboration 2015 (Acero et al., 2015) first published the 3FGL catalog, later updating it in April and June with two versions correcting the original. The 3FGL catalog includes 3033 sources with detection confidence levels higher than 4σ , detailing each source's location, spectral properties, and variability curves. Over 1100 confirmed or associated sources are blazars, including several other non-blazar AGN categories (Acero et al., 2015).
2. **The Third *Fermi*-LAT AGN Catalog (3LAC).** Ackermann et al. (2015a) introduced the third generation *Fermi*-LAT source clean catalog (3LAC Clean Sample), based on observations from

August 4, 2008, to July 31, 2012, in the 100 MeV to 300 GeV energy range with Test Statistic (TS) data greater than 25. 3LAC includes 1591 AGNs at high Galactic latitudes ($|\text{b}| > 10^\circ$), a 71% increase over the second catalog based on two years of data. There are 28 sources with multiple associations, and out of 2193 high-latitude γ -ray sources, 1563 are AGNs (see Fig. 2.6). Most of them (98%) are blazars. About half of the newly detected blazars are of unknown type, lacking sufficient spectral information to determine their emission lines' intensity. Based on their γ -ray spectral characteristics, these sources are evenly divided into FSRQs and BL Lacs. The most detected BL Lacs are of the high synchrotron peak (HSP) type. About 50% of BL Lacs have no measured redshift. Some new rare extreme types (HSP FSRQs and high-luminosity HSP BL Lacs) are recorded. The general properties of the 3LAC sample confirm the findings of previous catalogs. The latest discovery is that even the faintest blazars listed in the BZCAT blazar sample, covering flux distributions from radio, optical, and X-ray bands, still represent a significant proportion of the 3LAC blazar population, suggesting that even the weakest known blazars may eventually emit gamma rays at LAT detection levels. Also, the 3LAC catalog includes 25 non-blazar AGNs, a significant increase over the 2LAC catalog (Ackermann et al., 2015a).

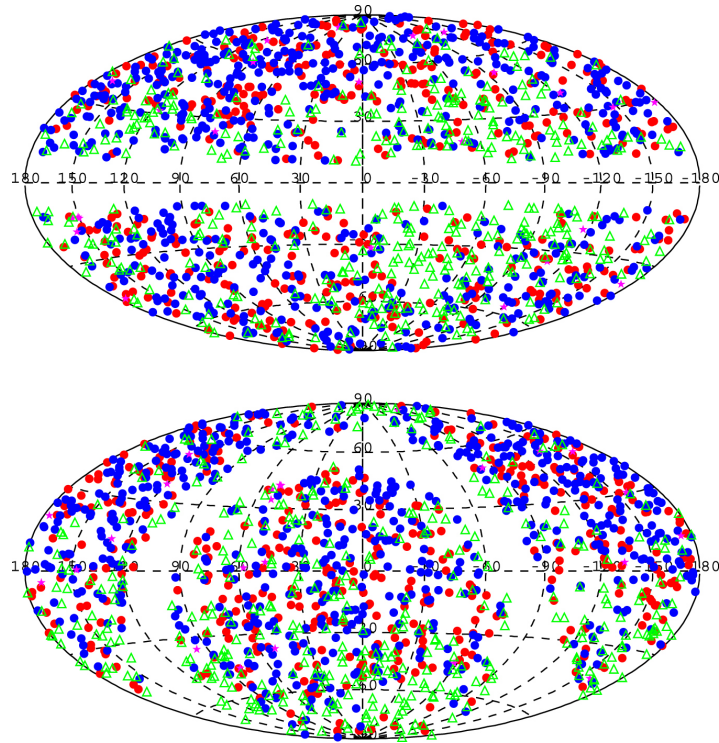


Figure 2.6: The distribution of sources in the 3LAC clean sample. Red circles: FSRQs; Blue circles: BL Lac objects; Magenta circles: non-blazar AGNs; Green triangles: AGNs of unknown type (Ackermann et al., 2015a).

3. **The Third *Fermi*-LAT High-Energy Catalog (3FHL).** The *Fermi*-LAT catalogs continued to evolve with the release of the Third *Fermi*-LAT High-Energy Catalog (3FHL) in 2017. Focusing on the high-energy γ -ray sky, 3FHL extended the energy range from 10 GeV to 2 TeV, marking a significant advancement in the study of the most energetic γ -ray sources. The 3FHL catalog included 1556 sources, of which 1207 were identified as blazars, making it an invaluable resource

for researchers exploring extreme astrophysical phenomena. It provided detailed spectral information for these high-energy sources, enabling deeper investigations into the mechanisms responsible for γ -ray production at these extreme energies.

2.4.4 The Fourth *Fermi*-LAT and AGN Catalogs

1. **The Fourth *Fermi*-Large Area Telescope Source Catalog (4FGL).** The Fourth *Fermi*-Large Area Telescope Source Catalog (4FGL) (Abdollahi et al., 2020) introduced the fourth generation of γ -ray source catalogs from the *Fermi*-LAT. Based on the first eight years of data from the *Fermi* satellite, it covered an energy range from 50 MeV to 1 TeV, the broadest to date. Compared to the 3FGL catalog, 4FGL doubled the exposure and included several analytical improvements such as an updated model for diffuse galactic γ -ray emission and bi-monthly and annual γ -ray light curves. The 4FGL catalog detected 5064 sources with a confidence level above 4σ (see Fig. 2.7), providing location and spectral characteristics for each. Over 3130 of these sources were confirmed or associated with blazars, and 239 were identified as pulsars.

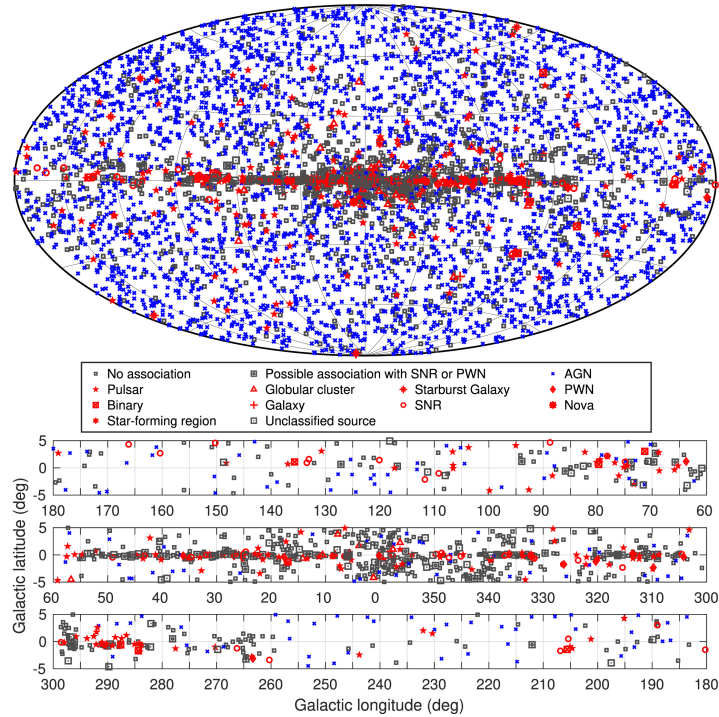


Figure 2.7: Survey locations of the 4FGL sources (upper part) and their positions on the galactic plane (lower three parts categorized by source type). All AGNs are marked in blue, well-defined associations of other classes in red, and unassociated sources or those with unknown properties in black (Abdollahi et al., 2020).

2. **The Fourth *Fermi*-LAT Catalog Data Release 2 (4FGL-DR2).** The 4FGL-DR2 (Ballet et al., 2020) summarized the LAT results from August 4, 2008, to August 2, 2018. This edition was based on a decade of survey data in the 50 MeV-1 TeV energy range. Post-maximum likelihood calculation, the threshold for including new sources was a Test Statistic (TS) ≥ 25 , corresponding to a significance slightly higher than 4σ . The 4FGL-DR2 catalog contained 5787 sources. Even if the TS was ≥ 25 , original 4FGL sources were retained in 4FGL-DR2. Of the 5064

sources detected in 4FGL, 120 fell below the detection threshold, 53 were newly associated, 4 associations were retracted, and 732 new sources were discovered, with 2 associated and 341 having tentative associations at other wavelengths.

3. **The Fourth *Fermi*-LAT Source Catalog Data Release 3 (4FGL-DR3).** The 4FGL-DR3 (Abdollahi et al., 2022) compiled LAT results from August 4, 2008, to August 2, 2020. This version, based on 12 years of survey data in the 50 MeV-1 TeV range, adapted the analysis for more sources with curved spectra. It introduced a more robust spectral parameterization for pulsars and extended spectral points up to 1 TeV. Like 4FGL-DR2, the inclusion threshold for new sources post-maximum likelihood was a $TS \geq 25$. The catalog listed 6658 sources. All previous 4FGL sources (DR1 and DR2, with some exceptions discussed in the paper) were retained in 4FGL-DR3, even if $TS < 25$. Of the 5064 sources detected in 4FGL, 16 were removed, 112 fell below the detection threshold, 74 were newly associated, 10 associations were improved, and 7 were retracted. A total of 1610 new sources were found, with 8 associated and 699 having tentative associations at other wavelengths.
4. **The Fourth *Fermi*-LAT Source Catalog Data Release 4 (4FGL-DR4).** The 4FGL-DR4 (Ballet et al., 2023) summarized the LAT results from August 4, 2008, to August 2, 2022. This edition, based on 14 years of survey data in the 50 MeV-1 TeV range, followed the same maximum likelihood inclusion criteria as 4FGL-DR2 and 4FGL-DR3, with a $TS \geq 25$ threshold for new sources. The catalog included 7194 sources, retaining all previous 4FGL sources (except for some discussed in this paper), even if $TS < 25$.
5. **The Four *Fermi*-LAT AGN Catalog (4LAC).** The 4LAC released a clean catalog of AGNs detected from August 4, 2008, to August 2, 2016, based on the 4FGL catalog (4LAC). It included 2863 high-latitude galactic objects (see Fig. 2.8), an 85% increase over the 3LAC catalog from the previous four years. AGNs constituted at least 79% of the 4FGL class, with the same energy range (50 MeV to 1 TeV). Additionally, 344 γ -ray AGNs were discovered in low-latitude galactic regions. The majority of 4LAC-AGNs were blazars (98%), with the remainder being other types of AGNs. The blazar population comprised 24% FSRQs, 38% BL Lac objects, and 38% unidentified blazar candidates (BCUs). Overall, FSRQs exhibited softer γ -ray spectra, and stronger variability, confirming previous findings. All AGNs detected by ground-based atmospheric Cerenkov telescopes were also included in 4LAC.
6. **The Four LAT AGN Catalog Release 2 (4LAC-DR2).** The 4LAC-DR2, based on a decade of data, detected 723 additional sources, including 285 new 4FGL-DR2 AGNs. Apart from two radio galaxies, these active galactic nuclei are all blazars. The overall sample of new sources comprises 39 FSRQs, 59 BL Lacs, 185 candidate blazars of unknown type (BCUs), and two radio galaxies, NGC 3078 and NGC 4261. The 4LAC-DR2 includes 262 new AGNs at high latitudes ($|b| > 10^\circ$), while the low-latitude sample contains an additional 23 (all BCUs except one BL Lac). A total of 67 sources lack a classification based on their SED. The new 4LAC-DR2 FSRQs have a slightly higher median photon index than the DR1 sample (2.63 vs. 2.45), indicating a softer spectrum. The median redshift is very similar to that of 4LAC-DR1 (1.19 vs. 1.12). PKS 2318-087 (4FGL J2320.8-0823), with a redshift of ($z = 3.164$), has the highest redshift among the new 4LAC-DR2 FSRQs. Four DR1 FSRQs have even higher redshifts, up to 4.31.
7. **The Four *Fermi*-LAT AGN Catalog Release 3 (4LAC-DR3).** The 4LAC-DR3 (Ajello et al., 2022) is derived from the 4FGL-DR3, based on 12 years of γ -ray data above 50 MeV. It features

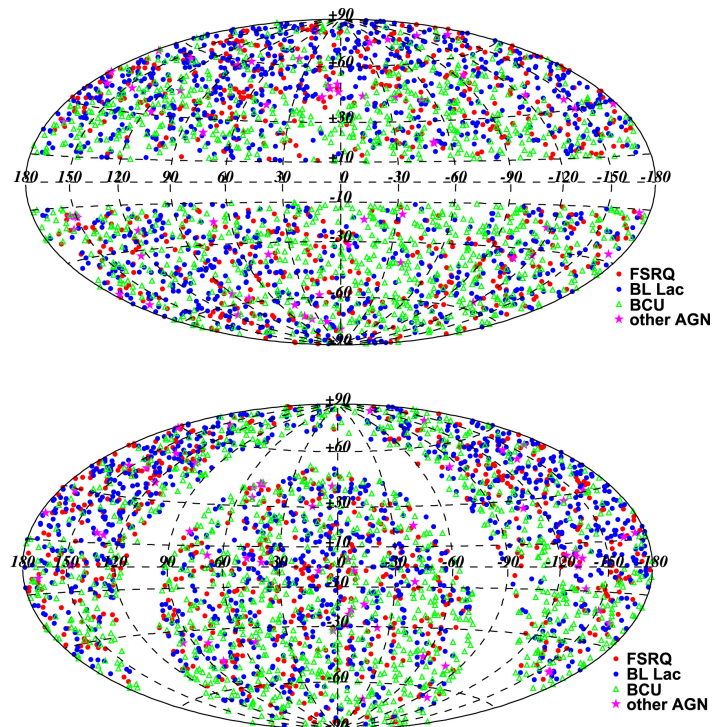


Figure 2.8: Locations of 4LAC sources projected on galactic coordinates (upper part) and J2000 equatorial coordinates (lower part) (Ajello et al., 2020).

updated spectral parameters, SEDs, annual light curves, and associations for all sources. The 4LAC-DR3 includes 1607 new sources relative to the initial 4FGL catalog. Ajello et al. (2022) discusses the properties of 283 new 4FGL-DR2 and 308 new 4FGL-DR3 AGNs, which are predominantly blazars, except for four radio galaxies.

Chapter 3

Methodology

The primary methodologies employed in this thesis are machine learning (ML) and its distinguished branch, deep learning (DL), both of which fall under the umbrella of Artificial Intelligence (AI) and Data Mining. The relationship between AI, ML, and DL is shown in Fig. 3.1.

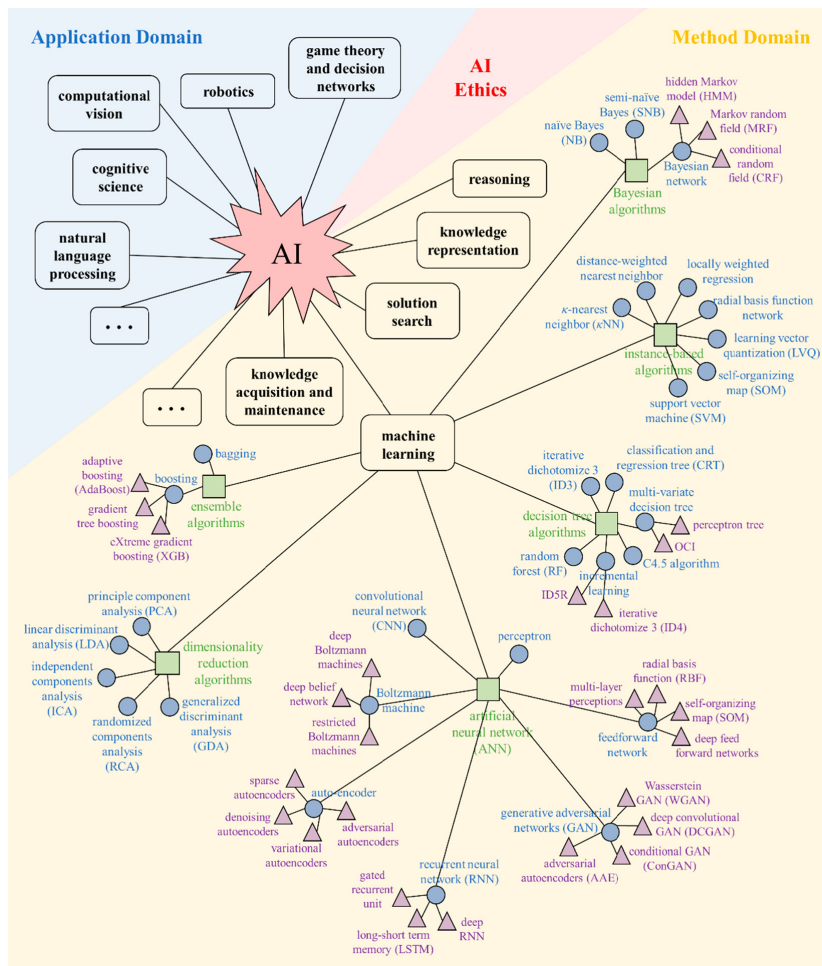


Figure 3.1: The inclusion relationship between AI, ML, and DL (Yang et al., 2019).

3.1 Briefly overview of AI and Data Mining

AI encompasses a range of disciplines, including cognitive science, operations research, statistics, and computer science, and even extends into economics and philosophy. Different researchers define AI from various perspectives. A popular definition, as proposed by Russell & Norvig (2010, 2020), suggests that AI should be defined based on its objectives, forming eight definitions across four groups along two dimensions (see Fig. 3.2): human performance or ideal rationality, and thinking/reasoning or behavior. They argue that the contrast between thinking/reasoning systems and acting systems is central to AI, equating intelligence with rationality. Specifically, AI is seen as the field dedicated to building intelligent agents (see Fig. 3.3), functions that take perceptual tuples from an external environment as inputs and generate behavior (actions) based on these perceptions. This definition posits goal-directed behavior as the essence of intelligence. Goal-directed agents can also be described in economic terms as “rational agents”, possessing a “utility function” that encapsulates all AI goals. Such agents aim to create and execute any plan that maximizes the expected value of this utility function upon completion (Bringsjord & Govindarajulu, 2022).

<p>Thinking Humanly</p> <p>“The exciting new effort to make computers think . . . <i>machines with minds</i>, in the full and literal sense.” (Haugeland, 1985)</p> <p>“[The automation of] activities that we associate with human thinking, activities such as decision-making, problem solving, learning . . .” (Bellman, 1978)</p>	<p>Thinking Rationally</p> <p>“The study of mental faculties through the use of computational models.” (Chamiak and McDermott, 1985)</p> <p>“The study of the computations that make it possible to perceive, reason, and act.” (Winston, 1992)</p>
<p>Acting Humanly</p> <p>“The art of creating machines that perform functions that require intelligence when performed by people.” (Kurzweil, 1990)</p> <p>“The study of how to make computers do things at which, at the moment, people are better.” (Rich and Knight, 1991)</p>	<p>Acting Rationally</p> <p>“Computational Intelligence is the study of the design of intelligent agents.” (Poole <i>et al.</i>, 1998)</p> <p>“AI . . . is concerned with intelligent behavior in artifacts.” (Nilsson, 1998)</p>

Figure 3.2: Four sets of definitions of AI (Russell & Norvig, 2010, 2020).

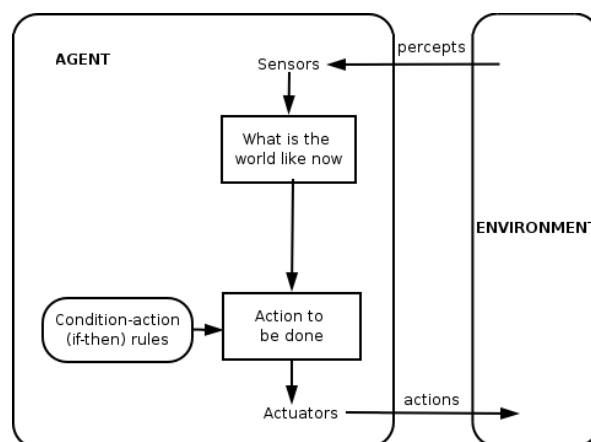


Figure 3.3: Simple reflex agent diagram (Russell & Norvig, 2010, 2020).

Data mining, an integral part of the data science field, involves discovering patterns, correlations, and insights from large data sets using statistical methods, ML, and database systems (Han et al., 2012). It is an interdisciplinary subfield of computer science and statistics, with the overarching goal of extracting information (using intelligent methods) from data sets and transforming it into a comprehensible structure for further use (Hastie et al., 2009; Sammut & Webb, 2017). Data mining is the analysis step of the knowledge discovery from data (KDD) process (Fayyad et al., 1996), involving database and data management aspects, data preprocessing, model and inference considerations, interestingness measures, complexity considerations, post-processing of discovered structures, visualization, and online updating.

The interrelationship between AI and data mining is multifaceted. In the contemporary technological landscape, the synergy between AI and data mining is crucial, representing a fusion of methodologies and objectives that drive the advancement of intelligent data analysis. The intersection of data mining with AI is often seen as an application within the AI field, frequently utilizing AI methods, particularly ML algorithms, for data analysis and interpretation. Conversely, AI systems, especially those based on ML, rely on data mining to identify patterns and regularities in data, aiding in the training and refinement of AI models (Bishop, 2007). This symbiotic relationship is further solidified through the shared use of techniques such as classification, clustering, and regression analysis, as well as ML and the more recent DL, which have become key technologies linking these two domains (LeCun et al., 2015). In practical applications, the convergence of AI and data mining is increasingly evident. For instance, in fields like recommendation systems, financial market analysis, and health informatics, data mining is used to extract insights, while AI is employed to build systems capable of learning from these insights and making intelligent decisions (Jordan & Mitchell, 2015). This collaborative dynamic not only enhances the efficiency of data analysis but also fosters the development of systems capable of intelligently navigating and interpreting complex data environments. In summary, while AI and data mining have distinct objectives and methodologies, their integration in practical applications demonstrates their complementarity, jointly advancing the evolution of large-scale data knowledge extraction and intelligent decision-making. This thesis focuses on the intersection of AI and data mining, specifically on some ML/DL algorithms and data processing procedures, and applies them to the classification of AGN/blazars and the screening of rare sources.

3.2 Historical Context

- **The Inception of AI.** The concept of AI dates back to ancient history, with myths and stories of artificial beings endowed with intelligence or consciousness by master craftsmen. However, as a scientific discipline, AI began in the mid-20th century. The term “Artificial Intelligence” was first coined by John McCarthy in 1956 at the Dartmouth Conference, marking the birth of AI as an independent field (McCarthy, 2004).
- **The Emergence of Data Mining.** Data mining, conceptually older, gained prominence in the 1990s. The term is believed to have been coined in the 1960s, but it was the advent of large databases and the need to extract meaningful information from these that propelled data mining into a field of its own (Fayyad et al., 1996).
- **Evolution of AI.**
 1. *Early AI (1950s - 1970s).* The initial phase of AI was characterized by enthusiasm and significant achievements. Early AI research focused on problem-solving and symbolic

methods. The development of algorithms for games like chess and checkers, and the creation of ELIZA, a natural language processing computer program, are notable milestones (Newell & Simon, 1976).

2. *AI Winter (1970s - mid-1980s)*. The AI winter refers to periods where funding and interest in AI research declined. The limitations of early AI became apparent, leading to skepticism and reduced funding (Crevier, 1993).
3. *Resurgence and Growth (mid-1980s - 2000s)*. The revival of AI was marked by the adoption of ML techniques, the rise of the internet, and increasing computational power. This period saw the development of neural networks, DL, and reinforcement learning (Goodfellow et al., 2016).
4. *Modern AI (2000s - Present)*. The current era of AI is characterized by big data, advanced algorithms, and significant improvements in computational power. AI applications are now widespread, ranging from voice recognition systems to autonomous vehicles (Jordan & Mitchell, 2015).

- **Evolution of Data Mining.**

1. *Early Developments (1960s - 1980s)*. The origins of data mining are in statistical analysis and early database technology. The focus was primarily on data collection and storage (Tukey, 1977).
2. *Growth and Recognition (1990s - 2000s)*. The 1990s saw a surge in interest in data mining, driven by the growth of large databases and the development of new algorithms for data analysis. This period marked the establishment of data mining as a distinct field (Burgess, 1998).
3. *Big Data and Advanced Analytics (2000s - Present)*. The explosion of big data has transformed data mining. The focus has shifted to handling large-scale, diverse datasets and real-time data processing. Techniques like ML, pattern recognition, and network analysis have become central to data mining (Manyika, 2011).

3.3 Segmentation of AI

- **Weak AI vs. Strong AI.** AI systems are generally categorized into weak (or narrow) AI and strong (or general) AI. Weak AI, designed for specific tasks, is prevalent in applications like Siri or Google's search algorithms (Poole et al., 1998). Strong AI, which aims to replicate general human cognitive abilities, remains a largely theoretical concept but is a subject of intense research and ethical debate (Searle, 1980).
- **Advances in ML and DL.** The resurgence of AI in the 21st century is largely attributed to advances in ML and DL. ML, a subset of AI, involves algorithms learning from and making predictions on data (Bishop, 2007). DL, a further subset of ML, employs neural networks with multiple layers to analyze vast datasets, significantly improving tasks like image and speech recognition (LeCun et al., 2015).

Applications of AI and data mining are vast and varied. In healthcare, they are used for disease prediction and diagnosis. In finance, they are applied in fraud detection and risk management. Other sectors like retail, manufacturing, and transportation also benefit from AI and data mining for

customer insights, predictive maintenance, and optimizing logistics (Davenport, 2006). There is no doubt that AI is a universal tool. These technologies are also applicable in the field of astronomy, as will be illustrated with specific examples in the later sections of this thesis.

3.4 ML

ML, a crucial subfield of artificial intelligence (AI), has fundamentally transformed our approach to problem-solving and data analysis across various domains. At its core, ML involves the development of algorithms that enable computers to learn from data and make predictions or decisions (Jordan & Mitchell, 2015). This transformative capability has reshaped how we tackle challenges in numerous fields.

- **Definition and Scope of ML.** ML is defined as the scientific study of algorithms and statistical models that enable computer systems to perform specific tasks without explicit instructions, relying instead on identifying patterns and making inferences (Bishop, 2007). It is a branch of AI that endows systems with the ability to learn and improve automatically through experience (Alpaydin, 2014). The scope of ML ranges from straightforward tasks like data classification to more complex challenges such as image recognition, natural language processing, and predictive analytics.
- **Historical Context and Evolution of ML.** The concept of machines learning from data dates back to the early days of computing. The term “machine learning” was formally introduced in the 1950s by Arthur Samuel (Samuel, 1959). Since then, the field has undergone significant evolution, marked by several key developments:
 1. ML, as a scientific endeavor, originated from the pursuit of AI. In the early days of AI as a discipline, some researchers were intrigued by the idea of machines learning from data. They experimented with various symbolic methods and what were then known as “Neural Networks”. These primarily included perceptrons and other models, which were later recognized as reinventions of generalized linear statistical models (Sar, 1994). Probabilistic reasoning was also adopted, especially in automated medical diagnostics (Russell & Norvig, 2010, 2020).
 2. However, the growing emphasis on logic and knowledge-based methods led to a rift between AI and ML. Probabilistic systems were plagued by theoretical and practical issues in data acquisition and representation (Russell & Norvig, 2010, 2020). By the 1980s, expert systems had come to dominate AI, and statistical methods fell out of favor (Langley, 2011). Work on symbolic/knowledge-based learning continued within AI, leading to inductive logic programming, but more statistical research now fell outside the realm of AI itself, in areas like pattern recognition and information retrieval (Russell & Norvig, 2010, 2020). Neural network research was also abandoned by AI and computer science around the same time. This line of research was continued outside the AI/CS field by researchers from other disciplines under the banner of “Connectionism”. Their major success came from the mid-1980s rediscovery of backpropagation.
 3. ML underwent a reorganization and emerged as its field, flourishing in the 1990s. The focus shifted from achieving AI to solving practical, solvable problems. It moved its emphasis from the symbolic methods inherited from AI to methods and models borrowed from statistics, fuzzy logic, and probability theory (Russell & Norvig, 2002, 2010, 2020).

3.4.1 Supporting Disciplines of ML

- **ML and Optimization Theory.** Optimization theory plays a crucial role in ML, especially in the training of algorithms. The learning process in ML is fundamentally an optimization problem, aiming to minimize a loss function that measures the discrepancy between predicted and actual outcomes. This optimization is vital for algorithms such as gradient descent used in training neural networks (Goodfellow et al., 2016). The interaction between ML and optimization is not limited to loss minimization but extends to more complex aspects like constraint satisfaction and resource allocation, aiming to find the best possible (or sufficiently good) solutions for a given problem under a set of specific constraints (Boyd & Vandenberghe, 2004a). These are foundational in areas like DL and reinforcement learning (Goodfellow et al., 2016).
- **ML and Statistics.** The relationship between ML and statistics is deeply rooted and symbiotic. ML extensively draws from statistical theory, especially in areas of inference, probabilistic models, and concepts of generalization (Hastie et al., 2009). Statistical methods provide the foundation for many ML algorithms, particularly in supervised learning, where the goal is to infer a function from labeled training data (Vapnik, 1998). ML heavily relies on statistical theory to build algorithms capable of learning and predicting from data (Hastie et al., 2009). This relationship is particularly evident in areas like Bayesian learning, where statistical methods are used to update beliefs based on new data (Gelman et al., 2004). The principles of statistical learning guide the development of algorithms that can make predictions and reveal patterns in data, thereby enhancing the ability of trained data to generalize to unseen data (James et al., 2013).

However, ML and statistics, while overlapping in data analysis, differ in focus and approach. ML emphasizes prediction and algorithmic complexity, prioritizing performance and adaptability to data (Breiman, 2001a). Statistics, in contrast, is grounded in probabilistic modeling and inference, with an emphasis on understanding data generation and making population inferences (Efron & Hastie, 2016).

From the above, two ML modeling paradigms can be distinguished: Data models and Algorithmic models. Data models, often statistical, focus on understanding the data generation process and are used for inference and hypothesis testing (McCullagh & Nelder, 1989). Algorithmic models, common in ML, prioritize prediction accuracy and are designed for performance, handling complex data structures without necessarily providing interpretability (Hastie et al., 2009).

- **ML and Computational Complexity** Computational complexity is a key aspect of ML, affecting the feasibility and scalability of ML algorithms. The complexity of an algorithm determines how its computational time and resource requirements grow with the size of the input data (Cormen et al., 2001). In ML, this is particularly important for algorithms that need to process large datasets or complex models, such as DL networks (Goodfellow et al., 2016). The trade-off between model accuracy and computational complexity is a key consideration in designing efficient ML algorithms.

The synergy among these disciplines drives the advancement of ML. Optimization theory provides the tools for building and refining models; statistics offer the principles for learning from and making predictions based on data; and computational complexity ensures the practicality and efficiency of these processes. Together, they form the pillars of an effective and efficient ML system.

3.4.2 Classification of ML algorithms

ML algorithms can be mainly divided into three paradigms (Russell & Norvig, 2010, 2020). supervised learning, unsupervised learning, and reinforcement learning (see Fig. 3.4).

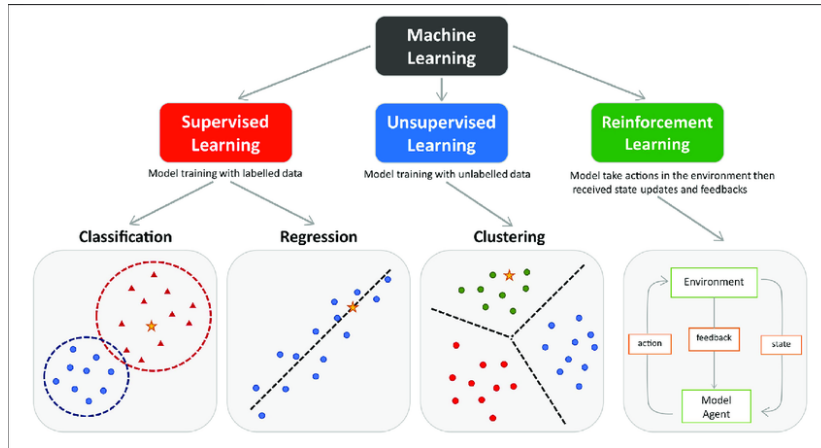


Figure 3.4: Classification of ML algorithms (Peng et al., 2021).

- Supervised machine learning.** Supervised Machine learning (SML), a predominant paradigm in ML, involves training algorithms on a labeled dataset, where the desired output is known. In this approach, the algorithm learns a mapping from inputs to outputs, guided by a known set of input-output pairs. The primary goal is to generalize from the training data to unseen situations in a predictive manner. This form of learning is widely used in applications such as image recognition, speech recognition, and medical diagnosis, where historical data with known outcomes are available. Supervised learning algorithms include linear regression for continuous output prediction and logistic regression, support vector machines, and neural networks for classification tasks. SML is the most commonly used paradigm in ML. Most of the algorithms used in this thesis belong to SML.
- Unsupervised machine learning.** Unsupervised machine learning (UML) deals with finding patterns or intrinsic structures in unlabeled data. Unlike supervised learning, unsupervised learning algorithms do not aim to predict an output but to understand the data's underlying structure. This approach is essential in exploratory data analysis, anomaly detection, and clustering tasks. Common unsupervised learning methods include k-means clustering, hierarchical clustering, and principal component analysis (PCA), which are used to group data based on similarities or to reduce the dimensionality of data for more efficient processing.
- Reinforcement machine learning.** Reinforcement machine learning (RML) is a distinct type of ML where an agent learns to make decisions by performing actions in an environment to achieve a goal. The learning process is driven by the feedback received in the form of rewards or penalties. Unlike supervised learning, reinforcement learning does not require labeled input/output pairs but instead focuses on finding a balance between exploration (of uncharted territory) and exploitation (of current knowledge). Applications of reinforcement learning include robotics, game playing, and autonomous vehicles, where the agent must learn to make a sequence of decisions that maximize some notion of cumulative reward.

Some perspectives suggest defining semi-supervised learning as an intermediary between SML and UML (Russell & Norvig, 2010, 2020), which combines both labeled and unlabeled data for training. It is particularly useful when labeled data is scarce or costly to obtain. Semi-supervised learning aims to improve model performance by using abundant unlabeled data alongside the limited labeled data. This approach is effective in areas like image and speech recognition, where fully labeled datasets are often hard to come by.

The subsequent subsection will detail common algorithms within SML and UML, with all algorithms utilized in this thesis falling under these two categories. DL, as a branch of ML based on Artificial Neural Networks (ANN) and focused on representation learning, possesses unique and exceptional qualities. These will be specifically addressed in Sec. 3.4.6.

3.4.3 ML flowchart

In the realm of ML, the processing flow is a meticulously structured sequence that transforms raw data into actionable insights, Fig. 3.5 illustrates the SML process. The UML flowchart is similar but differs in its handling of unlabeled data. This flow commences with the acquisition of data, a critical phase where diverse data sources are amalgamated and prepared for analysis. Following this, data preprocessing takes center stage, involving cleansing to rectify inconsistencies and missing values, normalization to standardize the range of data features, and feature extraction to identify the most relevant attributes. The heart of the ML process lies in the selection and application of algorithms, where choices are made between SML, UML, and RML based on the nature of the problem and data. SML, for instance, requires labeled datasets to train models, whereas UML algorithms discern patterns without preassigned labels. Post-algorithm selection, the model undergoes training, where it learns from the data, followed by a validation phase to fine-tune parameters and prevent overfitting. The final step is model evaluation, using metrics such as accuracy, precision, recall, and F1-score, to assess its performance. This comprehensive process, from data acquisition to model evaluation, encapsulates the essence of ML, driving the extraction of meaningful patterns and predictions from complex datasets. It is unrealistic to completely introduce the technology of every link in the flowchart. This thesis only introduces some of the most commonly used ones or even the parts that are only used in this thesis.

3.4.4 Basic framework of ML

The task of ML is to learn a model that can make a good prediction of its corresponding output for any given input. Taking SML as an example, the algorithm is designed to create a model that maps input data to output data. Based on the type of output - whether it's a category (discrete) or a numerical value (continuous) - this algorithm is categorized into either classification or regression. In this thesis, the focus will be on the classification algorithms. Before introducing SML algorithms, it is necessary to lay out some concepts (Mitchell, 1997; Boyd & Vandenberghe, 2004b; Mukherjee et al., 2006; Hastie et al., 2009; Nocedal & Wright, 2006; Abu-Mostafa et al., 2012; James et al., 2013; Alpaydin, 2014; Goodfellow et al., 2016; Mohri et al., 2018).

- **Input, Feature, and Output space.** In ML, the input space, often denoted as \mathcal{X} , is the set of all possible values that the input variables (or features) can take. For a dataset with n features, the input space can be mathematically represented as a subset of n -dimensional space, typically \mathbb{R}^n for continuous features or a discrete set for categorical features.

Feature space (\mathcal{F}) is a transformation or subset of the input space (\mathcal{X}), where each dimension corresponds to a feature used in the model. For a dataset with n features, the feature space can

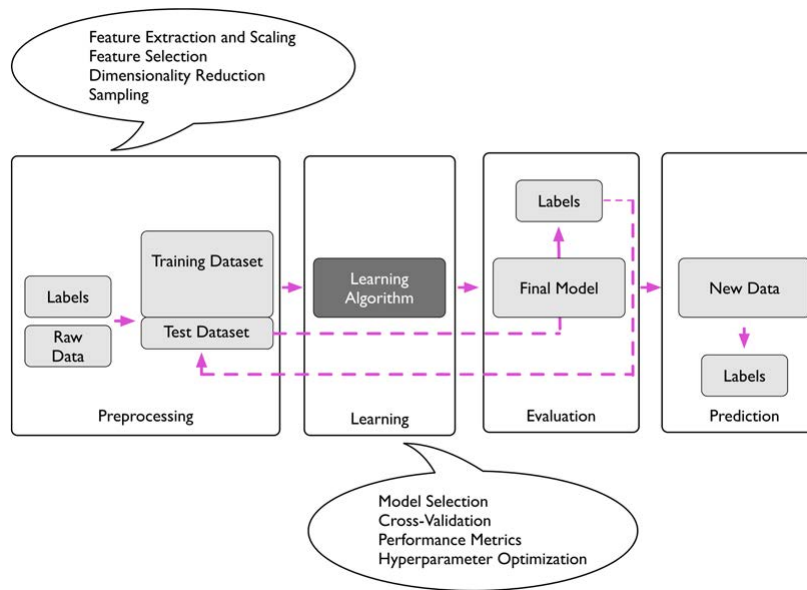


Figure 3.5: SML flowchart.

be represented as \mathbb{R}^n or a transformed version of it. Essentially, models of SML are defined within the feature space.

The output space in ML, denoted as \mathcal{Y} , refers to all possible outputs or responses that an ML model can predict. This space is crucial for defining the nature of the learning task, such as classification or regression, and influences the choice of algorithms and evaluation metrics. The nature of the output space determines whether the task is a classification (discrete output space) or regression (continuous output space).

In the process of SML, input-output pairs are considered as values of random variables defined on input/feature space and output space. Conventionally, the values of input variables are denoted by x , and those of output variables by y . In classification problems, Y often represents categories, also known as labels. This thesis deals with binary classification tasks, where y takes values 0 or 1. An input-output pair (x, y) is also referred to as a sample or sample point. X and Y together constitute a sample space. For instance, if classifying sources in the 4FGL catalog into two categories based on TeV emission, the input space X is the set of physical attributes, with x being specific attribute values, $x = (x_1, x_2, \dots, x_i, \dots, x_n)$, representing a feature/feature vector. For example, the redshift feature z can be represented as $z = (z_1, z_2, \dots, z_i, \dots, z_n)$, where z_i denotes the redshift of the i -th source.

Datasets are usually divided into two groups: a training set to train the model with known data, and a test set, acting as unclassified data for the model to make predictions. Sometimes, a small part of the training set is separated as a validation set for more refined model learning. For a sample (x, y) in the sample space (X, Y) , it is assumed that there exists an unknown true mapping $f : X \rightarrow Y$. Depending on whether the model in SML is a probabilistic model or a non-probabilistic model, this mapping f is a conditional probability distribution $P(Y|X)$ or a decision function $Y = f(X)$. For the former, it is assumed in SML that X and Y follow a joint probability distribution $P(X, Y)$. However, the form of $P(X, Y)$ is often unknown, and the training and test sets are considered to be independently and identically distributed according to

$P(X, Y)$. Formally, the task of SML can be divided into learning and prediction processes. In the learning process, on the training set $T = \{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$, a function \hat{f} (or model) is obtained through learning (or training) to approximate f ; in the prediction process, for a given input x_{n+1} in the test set, the output y_{n+1} is given by the mapping \hat{f} . The learning process involves continuous attempts to select the best model, minimizing the difference between the output y_i of the training set and the model output $\hat{f}(x_i)$.

- **Hypothesis and Version space.** The determination of the hypothesis space signifies the establishment of the scope of learning. The hypothesis space \mathcal{F} is typically a family of parameterized functions: $\mathcal{F} = \{\hat{f}(x, \theta) | \theta \in \mathbb{R}^m\}$, where $\hat{f}(x, \theta)$ represents the model, θ denotes a set of learnable parameters, m is the number of parameters, and \mathbb{R}^m is the parameter space.

For example, a decision tree algorithm has a hypothesis space comprising all the possible decision trees that can be constructed from the features, while a neural network's hypothesis space includes all the possible weight configurations of the network's architecture. The learning process involves searching this space to find the hypothesis that best approximates the true underlying function, typically by minimizing a loss function. This search is guided by training data and is subject to the constraints of the chosen model's structure. The determination of the hypothesis space signifies the establishment of the scope of learning.

Version space, denoted as \mathcal{V} , is the portion of the hypothesis space that agrees with all the training data (see Fig. 3.6). It represents all the hypotheses that have not been ruled out by the training examples. In other words, every hypothesis in the version space correctly classifies all the training examples.

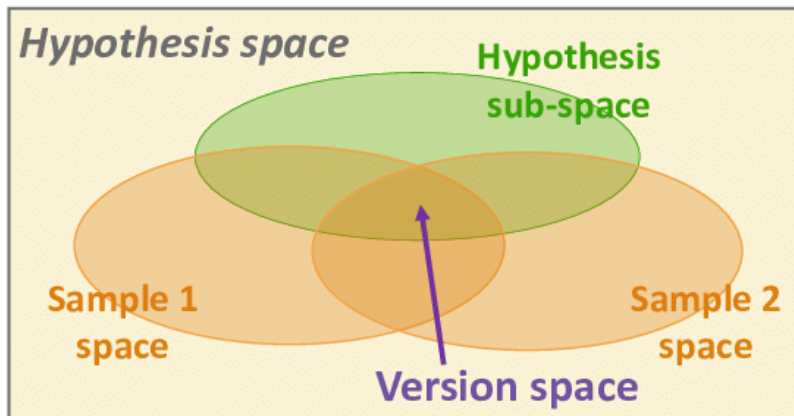



Figure 3.6: Hypothesis space and version space(Hsieh & Wang, 2018).

SML, UML, and RML all encompass three fundamental elements: model, strategy, and optimization algorithm.

1. *Model.* In SML, the model represents the mapping f to be learned, typically in the form of a conditional probability distribution or a decision function. The hypothesis space includes all possible conditional probability distributions or decision functions, generally infinite in number. The goal of learning is to select the best hypothesis f from the version space based on its performance on the training set. Subsequent chapters will introduce common algorithms for SML and UML.

- **Generative and Discriminative Models.** In SML, models can be broadly categorized into generative models and discriminative models. The models learned are referred to as generative or discriminative models, respectively (see Fig. 3.7). Generative and discriminative models each have their advantages and disadvantages and are suitable for different learning problems under various conditions.
 - (a) **Generative Models.** Generative approaches learn the joint probability distribution $P(Y|X)$ and then derive the conditional probability distribution $P(Y|X)$ as the predictive model. The generative model is defined as: $P(Y|X) = \frac{P(Y|X)}{P(X)}$. Such models are termed generative because they model the generation process of the output Y given an input X . Typical generative models include Gaussian Mixture Models.
 - Generative models can restore the joint probability distribution $P(X, Y)$, which discriminative models cannot.
 - They converge faster in learning, meaning the model can converge to the true model more quickly as the sample size increases.
 - Generative models can be used when there are hidden variables, whereas discriminative models cannot.
 - For generative models, $\sum P(x, y) = 1$, but $\sum P(y|x) < 1$, making it difficult to create a clear boundary between different classes, leading to overlapping regions.
 - (c) **Discriminative Models.** Discriminative models, on the other hand, learn the decision function $f(X)$ or the conditional probability distribution $P(Y|X)$ directly as the predictive model. Discriminative models focus on predicting the output Y for a given input X . Typical discriminative methods include Perceptron, Logistic Regression, Support Vector Machines, and Neural Networks. These classification algorithms will be discussed later in this thesis.
 - (d) **Characteristics of Discriminative Models.**
 - Discriminative methods learn the conditional probability $P(Y|X)$ or decision function $f(X)$ directly, often resulting in higher accuracy for prediction.
 - They allow for various degrees of abstraction and feature definition, simplifying the learning problem.
 - For discriminative models, $\sum P(y|x) = 1$, enabling them to create a clear distinction between classes.
2. *Strategy.* The objective of statistical learning is to learn the optimal model from the hypothesis space. Once the hypothesis space \mathcal{F} and learning criteria are determined, finding the optimal model becomes an optimization problem. The training process in machine learning is essentially the process of solving this optimization problem, involving criteria such as expected risk minimization, empirical risk minimization, and structural risk minimization. Loss functions measure the quality of a single prediction, while risk functions measure the model's average prediction quality.
- **Loss Functions and Risk Functions.** In SML, the model $f(x, \theta)$ provides an output y for a given input x . An ideal model $f(x, \theta)$ should consistently match the output of the true mapping function $y = f(x)$ for all possible values of (x, y) . However, the predicted output $f(x, \theta)$ may or may not match the true value y , necessitating a loss function or cost function

Discriminative and Generative Models 

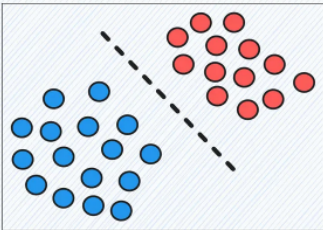
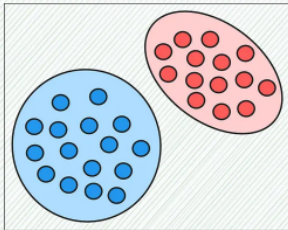
Discriminative Models	Generative Models
	
Learns the decision boundary between classes	Learns the input distribution
Maximizes the conditional probability: $P(Y X)$	Maximizes the joint probability: $P(X, Y)$
Directly estimates $P(Y X)$	Estimates $P(X Y)$ to find $P(Y X)$ using Bayes' rule
Cannot generate new data	Can generate new data
Specifically meant for classification tasks	Typically, they are NOT used to solve classification tasks
Discriminative models don't possess generative properties	Generative models possess discriminative properties

Figure 3.7: Generative vs. Discriminative Models.
[https://www.blog.dailydoseofds.com/p/an-intuitive-guide-to-generative.](https://www.blog.dailydoseofds.com/p/an-intuitive-guide-to-generative)

Source:

to quantify the discrepancy between the model's prediction and the actual label. The loss function is a non-negative real-valued function, denoted as $L(y, f(x, \theta)) = |f(x, \theta) - y| < \varepsilon$, where ε is a small positive number. Common loss functions include:

(a) **0-1 Loss Function.**

$$L(y, f(x, \theta)) = I(y \neq f(x, \theta)), \quad (3.1)$$

where $I(\cdot)$ is the indicator function. Although the 0-1 loss objectively evaluates the model, it is not continuous and has a derivative of zero, making it difficult to optimize. Continuous and differentiable loss functions are often used as substitutes.

(b) **Quadratic Loss Function.**

$$L(y, f(x, \theta)) = (f(x, \theta) - y)^2, \quad (3.2)$$

commonly used in tasks where the label y is a real number (e.g., regression tasks).

(c) **Logarithmic Loss Function.**

$$L(y, f(x, \theta)) = -\log f(x, \theta), \quad (3.3)$$

often interpreted as a log-likelihood function.

(d) **Cross-Entropy Loss Function.** Assuming label y belongs to $\{1, 2, \dots, C\}$, and the model output $f(x, \theta)$ is a conditional probability distribution over C classes. The loss is defined as:

$$L(y, f(x, \theta)) = -\sum_{c=1}^C y_c \log f_c(x, \theta), \quad (3.4)$$

where y is a label vector, y_c indicates the probability of class c , and $f_c(x, \theta)$ is the conditional probability of class c . This loss function is typically used in multi-class tasks.

(e) **Hinge Loss Function.**

$$L(y, f(x, \theta)) = \max(0, 1 - f(x, \theta)), \quad (3.5)$$

commonly used in binary classification problems.

- **Empirical Risk Minimization and Structural Risk Minimization.** The smaller the loss function value, the better the model. If the model's input and output follow a joint probability distribution $P(X, Y)$, the expected value of the loss function can be expressed as the risk function or expected loss:

$$R_{\text{exp}}(\theta) = \mathbb{E}_P[L(y, f(x, \theta))] = \int_{X \times Y} L(y, f(x, \theta)) P(x, y) dx dy. \quad (3.6)$$

This is the theoretical average cost of the model $f(x, \theta)$ for the joint distribution $P(X, Y)$, known as the risk function or expected loss. The goal of ML algorithms is to minimize the expected generalization error represented by this quantity, choosing the model with the minimum expected risk. It is important to note that this expectation is taken over the true underlying distribution $P(X, Y)$. If we knew the true distribution $P(X, Y)$, minimizing the risk would become an optimization problem solvable by optimization algorithms. However, in machine learning problems, we typically do not know the true distribution and only have access to a subset of the data in the training set. This makes supervised learning an ill-formed problem.

- (a) **Empirical risk minimization.** The simplest method to transform a machine learning problem into an optimization problem is by minimizing the expected loss on the training set. This implies using the empirical distribution $\hat{P}(X, Y)$ from the training set to approximate the true distribution $P(X, Y)$. The model's average loss over the training set becomes the empirical risk (or empirical loss), denoted as:

$$R_{\text{emp}}(\theta) = \frac{1}{N} \sum_{n=1}^N L(y, f(x, \theta)). \quad (3.7)$$

This training process based on minimizing the average training error is known as empirical risk minimization (ERM). In this case, the empirical risk becomes the target function for optimization, and supervised learning is transformed into an optimization problem of minimizing the empirical risk function.

- (b) **Structural risk minimization.** Empirical risk minimization is widely adopted, especially in the form of maximum likelihood estimation, and can ensure effective learning when the sample size is sufficiently large. However, in most cases, the training set is not infinite and often represents a small subset of the real data or contains noisy data, which may not accurately reflect the true distribution of all data. Compared to the limited data, models with too many parameters or overly complex structures tend to closely or precisely match a specific dataset, making them unable to fit unknown data well. Empirical risk minimization can easily lead to a model with a low error rate on the training set but a high error rate on unknown data. This phenomenon is known as overfitting, which will be discussed further.

Structural risk minimization (SRM) is a strategy proposed to prevent overfitting. Overfitting often results from limited data and noise, as well as strong model capabilities. To address overfitting, regularization is introduced on top of empirical risk minimization to limit the model's capabilities and prevent it from minimizing the empirical risk excessively. Structural risk minimization is equivalent to regularization. Structural risk = empirical risk + regularization term. In a determined hypothesis space, loss function, and training set, structural risk is defined as follows:

$$R_{\text{SRM}}(\theta) = R_{\text{emp}} + \lambda J(\theta) = \frac{1}{N} \sum_{n=1}^N L(y, f(x, \theta)) + \lambda J(\theta). \quad (3.8)$$

Here, $J(\theta)$ represents the model's complexity, a function defined on the hypothesis space \mathcal{F} , commonly using L1 or L2 norms. $J(\theta)$ can be understood as a penalty term for model complexity. $\lambda > 0$ controls the regularization strength, balancing empirical risk and model complexity. The SRM criterion considers the model with the minimum empirical risk as the optimal model. The optimization model is found by minimizing the structural loss function:

$$\theta^* = \arg \min_{\theta} R_{\text{SRM}}(\theta^*). \quad (3.9)$$

3. *Optimization Algorithm.* Based on the training set and learning strategy, selecting the optimal model from the hypothesis space ultimately involves considering which computational method to use for solving the optimal model. At this point, the SML problem is transformed into an optimization problem (Boyd & Vandenberghe, 2004b; Nesterov, 2004; Luenberger & Ye,

2015). Optimization problems typically do not have analytical solutions and require numerical methods for resolution. Ensuring the discovery of a global optimal solution and making the solving process highly efficient becomes a crucial issue. The following sections introduce four common types of optimization algorithms: Fermat's theorem, first-order derivative (gradient) methods, second-order derivative methods, and divide-and-conquer methods.

- (a) **Fermat's Theorem.** For a differentiable function, a unified approach to finding its extremum is to locate points where the derivative is zero, as stated in Fermat's Theorem. Points where the derivative is zero are called stationary points. It is important to note that a zero derivative is a necessary, but not sufficient, condition for an extremum; these are potential extremum points. For univariate functions, further judgment requires examining the second derivative, while for multivariate/vector functions, the Hessian matrix should be considered. At points where the derivative is zero, the function may not attain an extremum, known as saddle points. Besides saddle points, optimization algorithms may encounter another issue: local extremum, where a stationary point is an extremum but not the global extremum. Imposing certain constraints on optimization problems can effectively avoid these issues. A typical example is convex optimization, which requires the feasible domain of optimization variables to be a convex set, and the objective function to be a convex function. Formally, a convex set C is defined as convex if, for any two points $x, y \in C$, the line segment connecting x and y also lies within C . Mathematically, for all $\theta \in [0, 1]$, it holds that $\theta x + (1 - \theta)y \in C$. A convex function $f : \mathbb{R}^n \rightarrow \mathbb{R}$ is convex if its domain is a convex set and for any two points x, y in its domain and any $\theta \in [0, 1]$, the following inequality holds:

$$f(\theta x + (1 - \theta)y) \leq \theta f(x) + (1 - \theta)f(y). \quad (3.10)$$

The convex optimization problem can be formally expressed as:

$$\min_{x \in C} f(x), \quad (3.11)$$

where $f(x)$ is a convex function, and C is a convex set.

- (b) **Lagrange Multiplier Method.** Fermat's Theorem provides the necessary condition for an extremum of functions without constraints. For practical problems, there are often equality or inequality constraints. For problems with equality constraints, the classical solution is the Lagrange Multiplier Method. This method introduces additional variables, known as Lagrange multipliers, for each constraint. The key idea is to transform a constrained optimization problem into an unconstrained problem in a higher-dimensional space. The Lagrange function is defined as:

$$\mathcal{L}(x, \lambda) = f(x) - \lambda \cdot g(x), \quad (3.12)$$

where $f(x)$ is the objective function to be optimized, $g(x) = 0$ represents the constraint, and λ is the Lagrange multiplier. To find the optimal points, one needs to solve the following system of equations derived from the Lagrange function:

$$\nabla \mathcal{L}(x, \lambda) = 0. \quad (3.13)$$

This involves taking the gradient of the Lagrange function concerning both the original variables and the Lagrange multipliers and setting them to zero. The solutions to this system give the candidate points for the optima under the given constraints.

(c) **Karush-Kuhn-Tucker Conditions.** The Karush-Kuhn-Tucker (KKT) conditions are a set of necessary conditions for a solution in nonlinear programming to be optimal, especially when the problem involves inequality constraints. These conditions generalize the method of Lagrange multipliers to handle inequality constraints. The KKT conditions include both primal and dual feasibility, stationarity, and complementary slackness conditions. Consider an optimization problem of the form: Minimize $f(x)$, subject to $g_i(x) \leq 0$ for $i = 1, \dots, m$, and $h_j(x) = 0$ for $j = 1, \dots, p$. The KKT conditions are:

- **Primal Feasibility.** The solution must satisfy the problem's constraints.
- **Dual Feasibility.** The Lagrange multipliers associated with the inequality constraints must be non-negative.
- **Stationarity.** The gradient of the Lagrangian concerning the variables x must be zero.
- **Complementary Slackness.** For each inequality constraint, the product of the Lagrange multiplier and the constraint function must be zero.

The Lagrangian for this problem is given by:

$$\mathcal{L}(x, \lambda, \mu) = f(x) + \sum_{i=1}^m \lambda_i g_i(x) + \sum_{j=1}^p \mu_j h_j(x), \quad (3.14)$$

where λ_i and μ_j are the Lagrange multipliers. The KKT conditions can then be written as:

$$\nabla_x \mathcal{L}(x, \lambda, \mu) = 0 \quad (3.15)$$

$$g_i(x) \leq 0, \quad \lambda_i \geq 0, \quad \lambda_i g_i(x) = 0 \quad \text{for all } i \quad (3.16)$$

$$h_j(x) = 0 \quad \text{for all } j. \quad (3.17)$$

Numerical Optimization Algorithms. While theoretical derivations and root-finding formulas for equation systems are applicable in cases like linear functions and maximum likelihood estimation under normal distribution, they fall short when dealing with transcendental functions like exponential and logarithmic functions. In such scenarios, where the gradient equals zero, numerical optimization algorithms that utilize derivative information, such as first and second-order derivatives, become essential.

1. **Iterative Methods.** In practical implementations, iterative methods are commonly employed. These methods start from an initial point x_0 and repeatedly move from x_{old} to x_{new} according to a specific rule, forming a sequence that converges to a point where the gradient is zero:

$$\lim_{i \rightarrow \infty} \nabla f(x_{\text{old}}) = 0. \quad (3.18)$$

These rules typically utilize first-order derivative information, i.e., the gradient, or second-order derivative information, i.e., the Hessian matrix. The core of these iterative methods is to obtain an iterative formula that determines the next point from the previous one:

$$x_{\text{new}} = h(x_{\text{old}}). \quad (3.19)$$

2. **First-Order Derivative Methods: Gradient Descent.** Gradient Descent is a fundamental optimization algorithm in first-order derivatives, particularly effective for differentiable objective

functions. Its variants and extensions enhance its efficiency and applicability. The basic update rule for gradient descent is:

$$\theta_{\text{new}} = \theta_{\text{old}} - \eta \cdot \nabla_{\theta} J(\theta_{\text{old}}), \quad (3.20)$$

where θ represents the model parameters, η is the learning rate, and $\nabla_{\theta} J(\theta)$ is the gradient of the objective function.

3. **Variants of Gradient Descent.** Several common variants of Gradient Descent are introduced below (Ruder, 2016).

- **Stochastic Gradient Descent.** Stochastic Gradient Descent (SGD) estimates the gradient of the cost function using a single data point at each iteration, particularly useful for large datasets.

$$\theta_{\text{new}} = \theta_{\text{old}} - \eta \cdot \nabla_{\theta} J(\theta_{\text{old}}, x^{(i)}, y^{(i)}). \quad (3.21)$$

- **Mini-batch Gradient Descent.** A compromise between batch gradient descent and SGD, computing the gradient on a small subset of the training data.

$$\theta_{\text{new}} = \theta_{\text{old}} - \eta \cdot \nabla_{\theta} J(\theta_{\text{old}}, X_{\text{mini-batch}}, Y_{\text{mini-batch}}). \quad (3.22)$$

- **Momentum Gradient Descent.** Incorporates momentum to accelerate gradients and reduce oscillations.

$$v_{\text{new}} = \gamma v_{\text{old}} + \eta \nabla_{\theta} J(\theta), \theta_{\text{new}} = \theta_{\text{old}} - v_{\text{new}}. \quad (3.23)$$

- **Nesterov Accelerated Gradient.** Nesterov Accelerated Gradient (NAG) is a variation that calculates the gradient at a position ahead in the direction of momentum.

$$v_{\text{new}} = \gamma v_{\text{old}} + \eta \nabla_{\theta} J(\theta - \gamma v_{\text{old}}) \theta_{\text{new}} = \theta_{\text{old}} - v_{\text{new}}. \quad (3.24)$$

4. **second derivative method.**

- (a) **Newton's method.** Newton's Method, also known as the Newton-Raphson method, is a powerful technique in optimization, particularly for finding the roots of a real-valued function or the critical points of a real-valued function (i.e., points where the derivative is zero). The method is based on the idea of using a second-order Taylor series expansion to approximate a function near a point, and then iteratively improving the approximation. Given a function $f : \mathbb{R} \rightarrow \mathbb{R}$, Newton's Method updates the estimate of the root or critical point x as follows:

$$x_{\text{new}} = x_{\text{old}} - \frac{f'(x_{\text{old}})}{f''(x_{\text{old}})}, \quad (3.25)$$

where $f'(x)$ and $f''(x)$ are the first and second derivatives of f at x , respectively. Newton's Method can converge very quickly, especially if the starting point is close to the optimal point. However, the method requires the computation of second derivatives, which can be computationally expensive. It also assumes the function is twice differentiable and the Hessian is invertible.

In the context of optimization, particularly for finding the minima or maxima of a function, Newton's Method is applied to the derivative of the function. The update rule becomes:

$$x_{\text{new}} = x_{\text{old}} - \frac{f'(x_{\text{old}})}{f''(x_{\text{old}})}, \quad (3.26)$$

where $f'(x)$ is the gradient, and $f''(x)$ is the Hessian matrix of the objective function.

- (b) **Quasi-Newton methods.** Quasi-Newton methods are a group of optimization algorithms that seek to approximate Newton's Method for optimization, especially in scenarios where the computation of the exact Hessian matrix is computationally expensive or impractical. Quasi-Newton methods approximate the Hessian matrix of second derivatives, which is used in Newton's Method, with an easier-to-compute update at each iteration. This approach significantly reduces the computational burden while maintaining a superlinear convergence rate. The general update rule in Quasi-Newton methods is:

$$x_{\text{new}} = x_{\text{old}} - H^{-1} \nabla f(x_{\text{old}}), \quad (3.27)$$

where x_{old} and x_{new} are the current and updated parameter values, respectively. H^{-1} is the approximation of the inverse Hessian matrix, $\nabla f(x_{\text{old}})$ is the gradient of the function at x_{old} .

Quasi-Newton methods are used in various optimization problems in machine learning, particularly where the objective function is smooth but complex, and the computation of the exact Hessian is not feasible.

- (c) **Trust-Region Newton Method in Optimization.** The trust-region Newton Method is an optimization algorithm that combines the concepts of trust-region methods and Newton's method. It is particularly effective for solving nonlinear optimization problems. Unlike traditional methods that focus solely on the gradient or Hessian information, the Trust Region Newton Method considers both, but within a "trust region" around the current point. This region defines the area where the model is trusted to be a good approximation of the objective function.

The method iteratively solves a subproblem within the trust region to find the step direction and length. The subproblem typically involves minimizing a quadratic approximation of the objective function subject to a constraint that keeps the solution within the trust region:

$$\min_{\Delta x} \left\{ \nabla f(x)^T \Delta x + \frac{1}{2} \Delta x^T B \Delta x \right\} \quad \text{subject to} \quad \|\Delta x\| \leq \delta, \quad (3.28)$$

where Δx is the step to be taken, B is an approximation of the Hessian matrix, δ is the radius of the trust region.

The size of the trust region δ is adjusted based on how well the quadratic model approximates the objective function in the current region. More robust to the choice of initial point and to the ill-conditioning of the problem. However, it often converges in fewer iterations than standard Newton's method, especially for difficult problems.

5. **Divide and Conquer.** Divide-and-conquer is an algorithmic design philosophy that breaks down a large problem into smaller sub-problems to be solved individually. The solutions to these sub-problems are then combined to construct the solution for the entire problem. In optimization methods, this approach is specifically implemented by adjusting only a subset of components of the feature vector x in each iteration, while keeping the other components fixed. This translation captures the essence of the divide-and-conquer strategy in the context of optimization methods, emphasizing the process of solving smaller parts of a problem and then integrating these solutions.

- (a) **Coordinate Descent method.** The Coordinate Descent method is an optimization algorithm that solves optimization problems by successively optimizing along coordinate

directions to find the minimum of a function. It works by iteratively optimizing the objective function along one coordinate direction at a time while keeping other coordinates fixed. This method is particularly effective for problems where optimizing along a single coordinate is significantly simpler than optimizing along all directions simultaneously. Given a function $f : \mathbb{R}^n \rightarrow \mathbb{R}$, the Coordinate Descent algorithm updates one coordinate at a time:

$$x_i^{(k+1)} = \arg \min_{x_i} f(x_1^{(k+1)}, \dots, x_{i-1}^{(k+1)}, x_i, x_{i+1}^{(k)}, \dots, x_n^{(k)}), \quad (3.29)$$

for $i = 1, \dots, n$ and k denotes the iteration number. The Coordinate Descent method iterates over coordinates in a fixed order and chooses a random coordinate to optimize at each iteration.

Coordinate Descent is widely used in large-scale optimization problems. It is conceptually easy to implement and efficient when the optimization problem decomposes nicely along individual coordinates.

- (b) **Sequential Minimal Optimization.** Sequential Minimal Optimization (SMO) is an algorithm for solving the quadratic programming (QP) problem that arises during the training of Support Vector Machines. It is known for its efficiency in large-scale SVM training. SMO breaks the large QP problem into a series of smallest possible QP problems. These small QP problems are solved analytically, which significantly reduces the computational intensity compared to standard QP-solving methods. SMO works by choosing two Lagrange multipliers at each step and finding the optimal values for these multipliers while fixing all others. The algorithm iteratively updates these multipliers until convergence: Choose α_i and α_j , and solve the two-dimensional QP problem.

The SMO maximizes the SVM objective function concerning the chosen multipliers. The algorithm maintains the constraints of the SVM optimization problem at each step. SMO can be faster than traditional QP solvers for SVM training, especially when dealing with large datasets. Besides, it does not require an external QP optimization package.

- (c) **Stage-wise Optimization Algorithms.** Stage-wise optimization algorithms are a class of methods in machine learning where the model is built in a step-by-step fashion. AdaBoost (Adaptive Boosting) is a prominent example of this approach, particularly used in the context of ensemble learning. In stage-wise optimization, the model is constructed by adding a new component at each stage, to reduce the overall model error. The process is iterative, with each step building upon the previous ones.

AdaBoost is one of the most popular stage-wise optimization algorithms. It combines multiple weak learners to form a strong learner. In each stage, AdaBoost changes the weights of misclassified data points so that subsequent learners focus more on difficult cases. Given: Training data $(x_i, y_i), i = 1, \dots, N$. Initialize: Weights $w_i = \frac{1}{N}$, for all i . For $t = 1$ to T : 1. Train weak learner using weights w_i . 2. Calculate error and learner weight α_t . 3. Update weights w_i .

AdaBoost is less prone to overfitting compared to many other algorithms. It can be used with any learning algorithm that accepts weights on the training set. AdaBoost is widely used for binary classification problems and has been extended to multi-class and regression tasks.

- **Model Selection.**

- **Overfitting.** When the hypothesis space contains models of varying complexities (e.g., different numbers of parameters), the issue of model selection arises. If one solely aims to enhance the predictive ability on the training set, it often leads to the phenomenon known as "overfitting." Overfitting occurs when a machine learning model excessively adapts to the training set, resulting in poor performance on new test or validation sets. This is typically due to either excessive model complexity or insufficient training data. Overfitting is characterized by excellent performance on training data but poor performance on test data, where the model may generate overly complex decision boundaries that fail to generalize to new data. This situation equates to low bias but high variance, necessitating a trade-off between bias and variance, which is a bias-variance trade-off problem, as shown in Fig. 3.8.

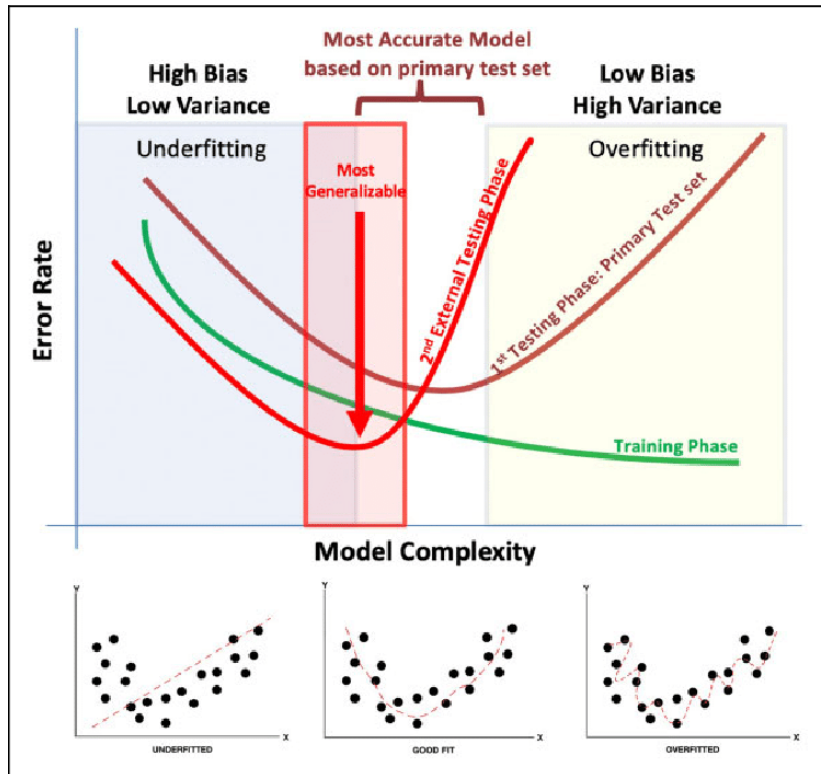


Figure 3.8: Bias-variance trade-off (Rashidi et al., 2019).

- **Regularization.** In addition to simplifying the model and expanding the dataset, regularization is a classic method to combat overfitting. Regularization implements the strategy of structural risk minimization by adding a regularizer or penalty term to the empirical risk. The regularizer is typically a monotonically increasing function of model complexity. Regularization generally takes the following form:

$$\min_{f \in \mathcal{F}} \frac{1}{N} \sum_{i=1}^N L(y_i, f(x_i)) + \lambda J(f), \quad (3.30)$$

where the first term is the empirical risk, and the second term is the regularizer, with $\lambda \geq 0$ as the penalty strength. The most commonly used regularizers are the L1 norm regularizer and the L2 norm regularizer.

The L1 norm regularizer can be expressed as:

$$L(w) = \frac{1}{N} \sum_{i=1}^N (f(x; w) - y_i)^2 + \lambda \|w\|, \quad (3.31)$$

where $\|w\|$ represents the L1 norm of the parameter vector w .

The L2 norm regularizer can be expressed as:

$$L(w) = \frac{1}{N} \sum_{i=1}^N (f(x; w) - y_i)^2 + \frac{\lambda}{2} \|w\|^2, \quad (3.32)$$

where $\|w\|$ represents the L2 norm of the parameter vector w . Fig. 3.9 shows L1 and L2 regularization.

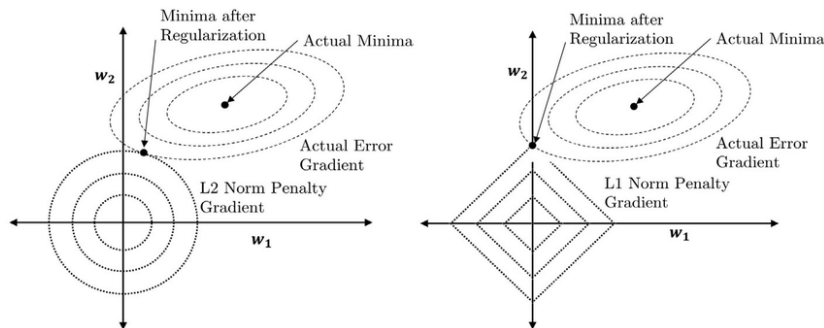


Figure 3.9: L2 norm regularization (left) and L1 norm regularization (right) (Santosh et al., 2022).

The first term's empirical risk is smaller for more complex models (with multiple non-zero parameters), resulting in a larger model complexity in the second term. The role of regularization is to select models with both low empirical risk and low model complexity. The concept of Occam's Razor (Sober, 2015) in machine learning is similar to the idea of regularization: simpler models have better generalization capabilities. If two models have similar performance, the simpler model should be chosen. In machine learning criteria, parameter regularization is often introduced to limit model capabilities and avoid overfitting. One formalization of Occam's Razor is the Minimum Description Length (MDL) principle, which states that for a dataset \mathcal{D} , the best model $f \in \mathcal{F}$ is the one that achieves the best compression of the dataset, i.e., the shortest encoding length. The minimum description length can also be interpreted from a Bayesian learning perspective (MacKay, 2003). The log posterior probability of a model f given a dataset \mathcal{D} is:

$$\max_f \log p(f|\mathcal{D}) = \max_f \log p(\mathcal{D}|f) + \log p(f), \quad (3.33)$$

where $-\log p(f)$ and $-\log p(\mathcal{D}|f)$ can be viewed as the encoding length of the model f and the encoding length of the dataset \mathcal{D} under that model, respectively. In other words, we not only want the model f to encode the dataset \mathcal{D} but also want the model f to be as simple as possible.

- **Cross-Validation.** Cross-validation is a prevalent method for model selection. When sample data is abundant, it can be divided into training, validation, and test sets. The training set is

used for model training, the validation set for model selection, and the test set for final model evaluation. Among models of varying complexities, the one with the smallest prediction error on the validation set is chosen. If the validation set is sufficiently large, it is effective for model selection. However, in practice, data may not always be abundant. To select an appropriate model, cross-validation methods are employed, which involve repeatedly using data; splitting the given data, combining split datasets into training and test sets, and repeatedly training, testing, and selecting models. The following are some of the most commonly used cross-validation methods:

1. **HoldOut Cross-validation.** In this technique, the entire dataset is randomly divided into a training set and a test set. For example, approximately 70% of the dataset is used as the training set and the remaining 30% as the test set. This method is straightforward to implement but is not suitable for imbalanced datasets. For instance, in a binary classification task with labels 0 and 1, if 80% of the data belongs to class 0, dividing the dataset into an 8:2 ratio might result in class 0 data being entirely in the training set and class 1 data in the test set. Moreover, in small datasets, some data reserved for testing might contain important features, and the model might miss these during training.
2. **K-Fold Cross-Validation.** In K-Fold cross-validation, the entire dataset is divided into K equal parts, each called a "Fold." Thus, there are K parts, known as K-Fold. One Fold is used as the test set, and the remaining K-1 Folds as the training set. This process is repeated K times until each Fold has been used as the test set. The final accuracy of the model is calculated by averaging the accuracies of k models. However, this method is also not suitable for imbalanced datasets or time series data. In time series data, the order of samples is important, but in K-Fold cross-validation, samples are selected randomly. Fig. 3.10 illustrates K-Fold Cross-Validation.

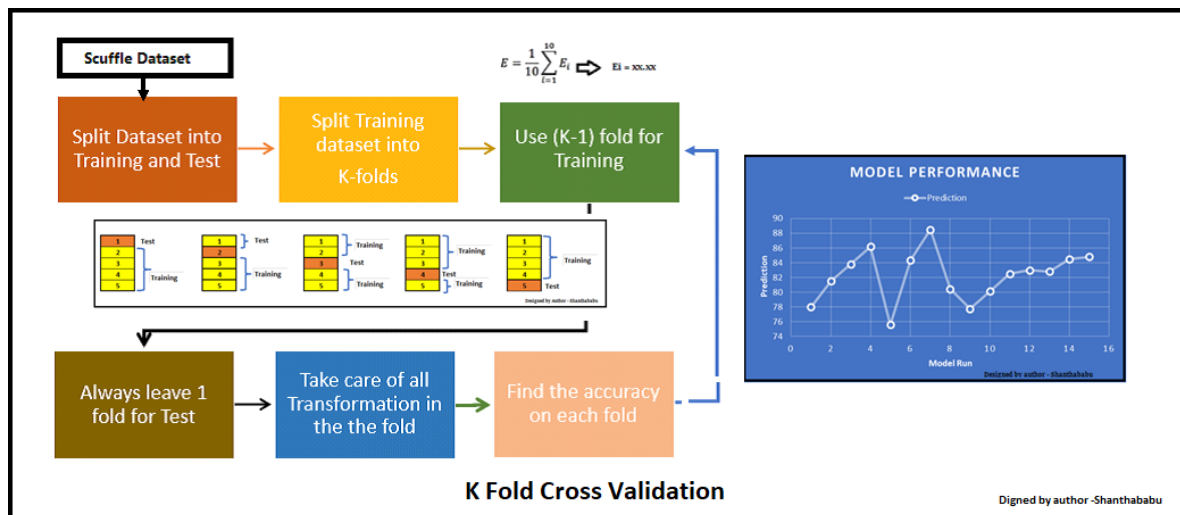


Figure 3.10: K-Fold Cross-Validation <https://www.analyticsvidhya.com/blog/2022/02/k-fold-cross-validation-technique-and-its-essentials/>.

3. **Stratified K-Fold Cross-Validation.** Stratified K-Fold is an enhanced version of K-Fold cross-validation, primarily used for imbalanced datasets. Like K-fold, the entire dataset is divided into K Folds of equal size. However, in this technique, the proportion of in-

stances of the target variable in each Fold matches the proportion in the entire dataset. This method is effective for imbalanced data. Each Fold in stratified cross-validation will have representatives of all classes, in the same proportion as the entire dataset. However, it is also not suitable for time series data. The thesis primarily deals with binary classification tasks with imbalanced classes, and the cross-validation method used is Stratified K-Fold Cross-Validation.

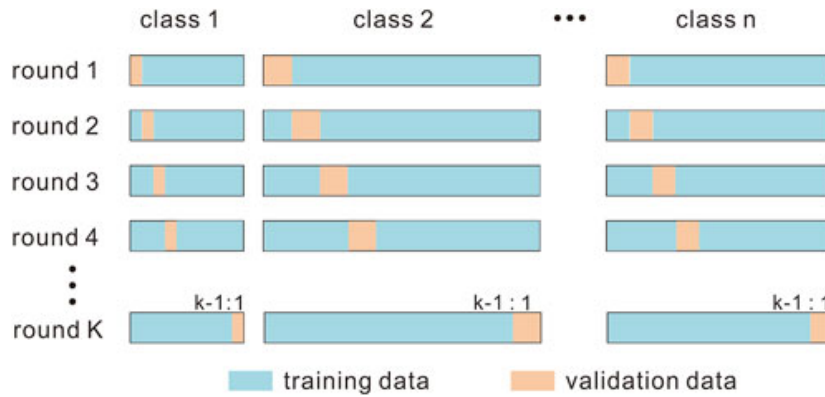


Figure 3.11: Stratified K-Fold Cross-Validation (Duan, 2023).

4. **Leave P Out Cross-Validation.** Leave P Out cross-validation is an exhaustive technique where p samples are used as the validation set and the remaining n samples as the training set. For example, if we have 100 samples in the dataset and use $p=10$, then in each iteration, 10 samples will be used as the validation set and the remaining 90 as the training set. This process is repeated until the entire dataset is divided into p -sample validation sets and $n-p$ training samples. However, this technique is computationally intensive and not suitable for imbalanced datasets. Like K-Fold cross-validation, if we only have samples of one class in the training set, our model will not be able to generalize to the validation set.
5. **Leave One Out Cross-Validation.** Leave One Out cross-validation is another exhaustive technique where one sample point is used as the validation set and the remaining $n-1$ samples as the training set. This process is repeated until every sample in the dataset has been used as a validation point. This is similar to Leave P Out cross-validation with $P=1$. However, it requires less computation time compared to Leave P Out, thus sharing similar advantages and disadvantages.

Metrics of Classification. SML learns a classification model or decision function from data, known as a classifier, with the output being categories. For multi-category classification problems, this thesis primarily discusses binary classification problems. There are several common evaluation metrics for binary classification. The class of interest is usually designated as the positive class, denoted as 1, and the other as the negative class denoted as 0. Based on the predicted and actual categories, there are four combinations: TP - positive class predicted as positive; FN - positive class predicted as negative; FP - negative class predicted as positive; TN - negative class predicted as negative. These four items can be intuitively represented as a confusion matrix (see Fig. 3.12).

These four items lead to many commonly used metrics: Accuracy (ACC) = $(TP + TN) / (FP + FN + TP + TN)$, representing the proportion of correctly predicted samples out of the total samples;

		Actual Values	
		Positive (1)	Negative (0)
Predicted Values	Positive (1)	TP	FP
	Negative (0)	FN	TN

Figure 3.12: Confusion Matrix (Duan, 2023).

True Positive Rate (TPR) = $TP / (FN + TP)$, indicating the model's ability to correctly identify positive classes. A higher TPR means the model is more effective at identifying positive samples; False Positive Rate (FPR) = $FP / (FP + TN)$, representing the proportion of negative samples incorrectly marked as positive out of all actual negative samples. A lower FPR indicates the model performs better in avoiding wrongly identifying negative samples as positive; Precision (PRE) = $TP / (TP + FP)$, representing the proportion of actual positive samples among those predicted as positive by the model; Recall/Sensitivity (REC/SEN) = $TPR = TP / (FN + TP)$, indicating the proportion of actual positive samples correctly identified as positive by the model; F1 score = $2 \times (PRE \times REC) / (PRE + REC)$; Receiver Operating Characteristic Curve (ROC): In the ROC curve, TPR is usually the ordinate, and FPR is the abscissa, used to assess the model's performance at different discrimination thresholds. Area Under the Curve (AUC): An ideal classifier should maximize TPR while minimizing FPR. The area under the ROC curve, AUC, provides a measure of the overall performance of a model. AUC values range between 0 and 1. AUC = 1 indicates a perfect classifier, able to completely distinguish between positive and negative classes; $0.5 < AUC < 1$ indicates the model can differentiate between positive and negative classes, with performance improving as AUC approaches 1; AUC = 0.5 indicates no discriminative ability, equivalent to random guessing; AUC < 0.5 is rare and indicates performance worse than random guessing. ROC and AUC are particularly useful when dealing with imbalanced datasets, as they are unaffected by class distribution. By comparing the AUCs of different models, it is possible to determine which model performs better overall.

In this thesis, we focus on two traditional SML techniques: logistic regression and support vector machine. Additionally, we explore a DL approach based on artificial neural networks, specifically the transfer learning method. For UML, we employ the Gaussian mixture model. The subsequent algorithmic discussions are primarily centered around these specified algorithms and their related methodologies. For comprehensive details on other algorithms, readers are referred to the following key texts: MacKay (2003); Webb (2010); Abu-Mostafa et al. (2012); Raschka (2015a); Mohri et al. (2018); An (2019); Rashidi et al. (2019).

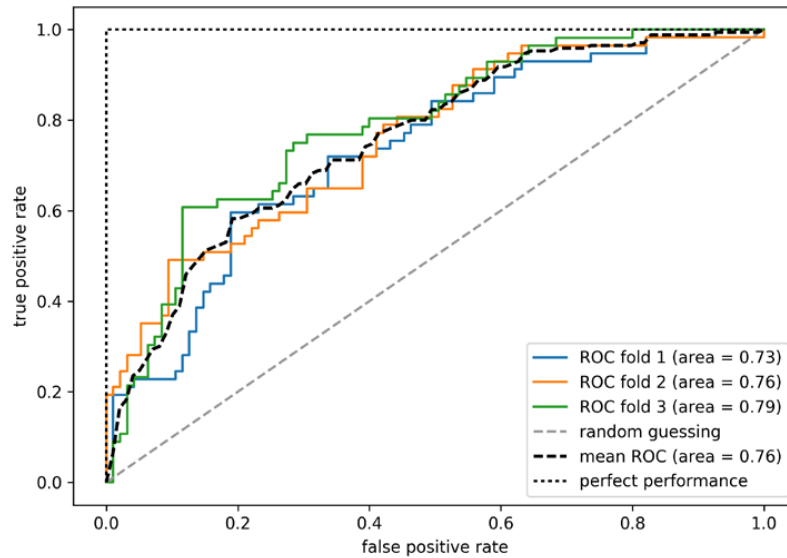


Figure 3.13: ROC Curves (Raschka, 2015a).

3.4.5 SML algorithms

Perceptron and Adeline

The perceptron, introduced by McCulloch & Pitts (1943), is a linear classification model for binary classification, aiming to identify a separating hyperplane that divides the dataset in the feature space into positive and negative categories, classifying it as a discriminative model. The perceptron can be presented in two forms: the primal form and the dual form.

- **Formal Definition**

The perceptron model defines the so-called activation function, which is a linear equation used to map input feature vectors to a binary output (+1 or -1, representing two classes):

$$f(\mathbf{x}) = \text{sign}(\mathbf{w} \cdot \mathbf{x} + \mathbf{b})$$

where \mathbf{w} is the weight vector, \mathbf{x} is the input feature vector, \mathbf{b} is the bias term, and sign is a sign function that maps positive values to +1 and negative values to -1.

The activation function in a perceptron is designed to emulate the response pattern of neurons in the human brain. This function activates when the input signal surpasses a specific threshold, akin to the activation of a neuron. When an instance is input into the perceptron, if its output exceeds a predetermined threshold, it is classified into the +1 category; otherwise, it falls into the -1 category.

- **Geometric Interpretation**

Geometrically, the perceptron model can be viewed as a hyperplane dividing two classes in the feature space. The vector \mathbf{w} determines the orientation of the hyperplane, while \mathbf{b} determines its distance from the origin.

- **Linear Separability of the Dataset**

The perceptron algorithm applies to linearly separable datasets, where a hyperplane can completely and correctly separate the two classes.

- **Loss Function**

The perceptron uses the total distance of misclassified points to the hyperplane as its loss function:

$$L(\mathbf{w}, \mathbf{b}) = - \sum_{x_i \in M} y_i(w_i \cdot x_i + b_i), \quad (3.34)$$

where M is the set of misclassified points.

- **Original Form of the Algorithm**

The perceptron learning strategy aims to minimize the loss function, typically implemented through stochastic gradient descent. The algorithm adjusts \mathbf{w} and \mathbf{b} to move the hyperplane towards misclassified points.

- **Dual Form**

The dual form of the perceptron is based on expressing \mathbf{w} and \mathbf{b} as linear combinations of the training data points. It maintains a vector α , where α_i is associated with the training instance x_i . The update rule involves updating α instead of directly updating \mathbf{w} and \mathbf{b} .

- **Algorithm Procedure**

1. **Initialization.** Set initial values of \mathbf{w} and \mathbf{b} , usually to 0.
2. **Select Misclassified Point.** Find a misclassified data point (x_i, y_i) .
3. **Update Weights and Bias.**

$$\begin{aligned} - w_i &\leftarrow w_i + \eta y_i x_i \\ - b_i &\leftarrow b_i + \eta y_i \end{aligned}$$

where η is the learning rate.

4. **Iteration.** Repeat steps 2 and 3 until no misclassified points are found.

- **Handling Non-Completely Separable Datasets**

The perceptron algorithm, as initially conceived, assumes that the dataset is linearly separable. However, in real-world scenarios, datasets often are not completely separable. In such cases, the perceptron algorithm faces the following limitations:

- **Convergence Issue.** The perceptron algorithm may fail to converge if the dataset is not linearly separable. This is because the algorithm continually adjusts weights and biases without reaching a state where all points are correctly classified.
- **Oscillation.** In non-separable cases, the algorithm may oscillate indefinitely, continuously updating the model parameters without finding an optimal hyperplane.

Adaline (Laboratories et al., 1960), an enhancement of the perceptron model, refines the fundamental concept of the loss function and its optimization. This advancement lays the groundwork for comprehending more complex algorithms, including logistic regression, support vector machines, and artificial neural networks. Adaline distinguishes itself from the perceptron primarily through its activation function, which is a continuous linear function rather than the perceptron's unit step function. Specifically, Adaline employs an identity function as its activation mechanism, expressed as

$$f(\mathbf{w} \cdot \mathbf{x}) = \mathbf{w} \cdot \mathbf{x}, \quad (3.35)$$

, where \mathbf{w} represents the weight vector and \mathbf{x} denotes the input vector.

Fig. 3.14 is the algorithm diagram of the perceptron and Adaline.

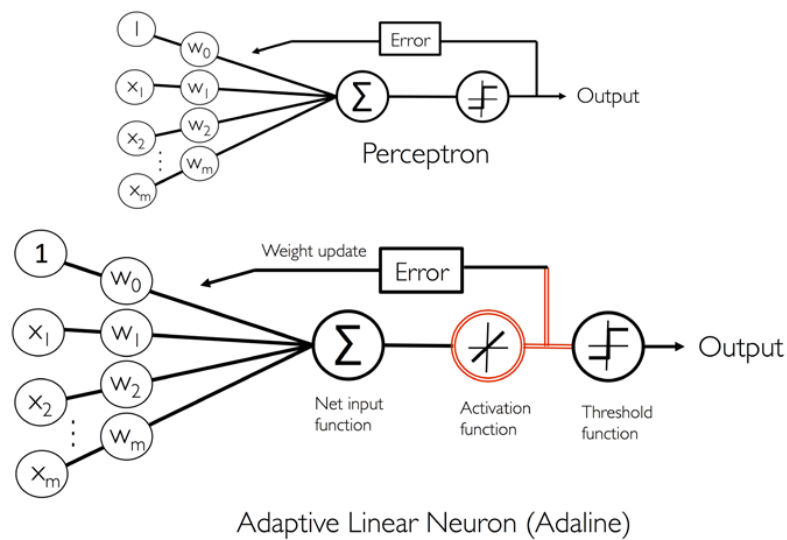


Figure 3.14: The algorithm diagram of the perceptron and Adaline (Raschka, 2015a).

Logistic Regression

Logistic Regression is a fundamental statistical method used for binary classification. It models the probability of a binary outcome based on one or more predictor variables. This method is grounded in the principles of the Exponential Family and Generalized Linear Models (GLMs), offering a robust framework for understanding and predicting binary events.

- Odds and Sigmoid in Logistic Regression

The "odds" in Logistic Regression refer to the ratio of the probability of an event occurring to the probability of it not occurring. This concept is crucial in understanding how Logistic Regression predicts binary outcomes. The odds are modeled using the logit function, which is the natural logarithm of the odds:

$$\text{logit } p = \text{Odds} = \frac{p}{1-p} \quad (3.36)$$

$$\log(\text{Odds}) = \log\left(\frac{p}{1-p}\right) \quad (3.37)$$

The primary objective of Logistic Regression is to estimate the probability that a given instance falls into a particular category. This is achieved through the logistic function, which serves as the inverse of the logit function. Characterized by its S-shaped curve, as depicted in Fig. 3.15, this function is commonly referred to as the sigmoid function due to its distinctive shape.

The sigmoid function (also see Fig. 3.15), often represented as $\sigma(z)$, is defined as:

$$\sigma(z) = \frac{1}{1 + e^{-z}}, \quad (3.38)$$

where $z = \mathbf{w} \cdot \mathbf{x}$, e is the base of the natural logarithm.

In contrast to Adaline, which employs a linear activation function, logistic regression utilizes the sigmoid function as its activation function, as illustrated in Fig. 3.16.

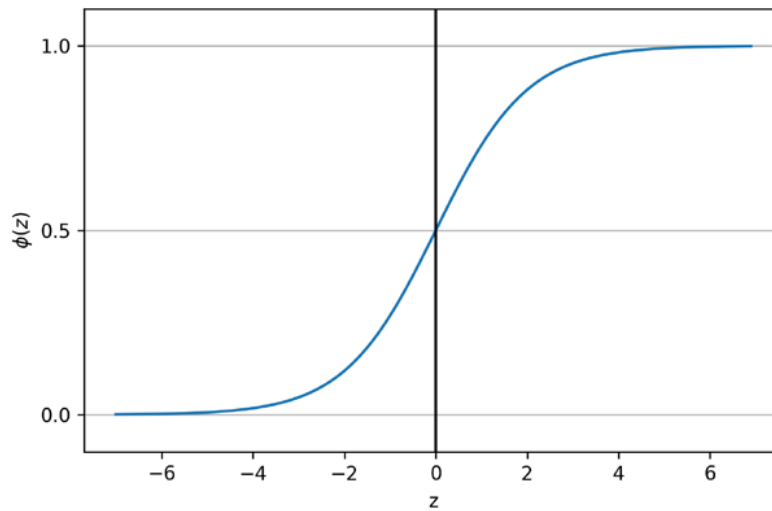


Figure 3.15: Sigmoid function (Raschka, 2015a).

- **Logistic Regression and the Exponential Family**

Logistic Regression can be seen as a special case of GLMs where the response variable's distribution belongs to the Exponential Family, specifically the Bernoulli distribution. The Exponential Family is characterized by the following form:

$$p(y; \theta) = b(y) \exp(\eta(\theta)^T T(y) - a(\theta)) \quad (3.39)$$

In the context of Logistic Regression, the Bernoulli distribution models the probability of the binary outcome, and the link function (logit function) connects the linear predictors to the mean of the distribution:

$$\log\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 X_1 + \dots + \beta_n X_n \quad (3.40)$$

- **Loss Function and Convex Optimization**

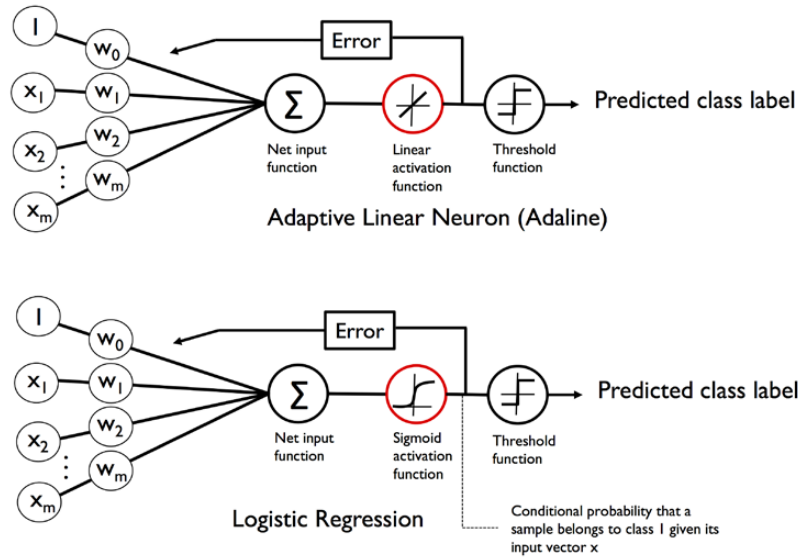


Figure 3.16: Comparison of activation functions in Adaline and Logistic Regression.

The loss function in Logistic Regression is derived from the likelihood function of the Bernoulli distribution. The negative log-likelihood, used for minimization, is a convex function, ensuring that any local minimum is also a global minimum. This property simplifies the optimization process. The log-loss for Logistic Regression is:

$$L(\beta) = - \sum_{i=1}^n [y_i \log(\pi_i) + (1 - y_i) \log(1 - \pi_i)] \quad (3.41)$$

- **Regularization in Logistic Regression**

Regularization techniques are often applied in Logistic Regression to prevent overfitting, especially when dealing with high-dimensional data. Regularization adds a penalty term to the loss function, which helps to control the model complexity. The most common forms of regularization are L1 (Lasso) and L2 (Ridge) regularization.

Support Vector Machine

The Support Vector Machine (SVM) represents another advancement beyond the perceptron and Adaline models. It is fundamentally aimed at maximizing the margin of classification, distinguishing it from its predecessors. The core of SVM lies in solving a convex quadratic programming problem, which equivalently minimizes the regularized hinge loss function. SVM encompasses a range of models, from simple to complex (Boser et al., 1992; Burges, 1998; Vapnik, 2000). In scenarios where data is linearly separable, a linear SVM is employed. For approximately linearly separable data, SVM adapts through the maximization of a soft margin. Non-linear SVMs address non-linear classification challenges by employing kernel tricks combined with soft margin maximization, thus offering a versatile solution for various data separability conditions.

- **linear SVM machine in linearly separable case**

In the context of a linear SVM machine in a linearly separable case, data is considered linearly separable if there exists a hyperplane that can separate the data points of one class from another without any overlap. Mathematically, for a binary classification with labels $y_i \in \{-1, 1\}$, a dataset is linearly separable if there exist parameters \mathbf{w} (weight vector) and b (bias) such that:

$$y_i(\mathbf{w} \cdot \mathbf{x}_i + b) > 0, \quad \text{for all } i \quad (3.42)$$

- **Geometric Margin and Functional Margin**

The functional margin of a hyperplane concerning a training example (\mathbf{x}_i, y_i) is defined as:

$$\hat{\gamma}_i = y_i(\mathbf{w} \cdot \mathbf{x}_i + b) \quad (3.43)$$

The geometric margin is the Euclidean distance from the hyperplane to the closest data point and is normalized by the norm of the weight vector:

$$\gamma_i = \frac{\hat{\gamma}_i}{\|\mathbf{w}\|} \quad (3.44)$$

- **Maximum Margin Hyperplane**

The maximum margin hyperplane is the one that maximizes the minimum geometric margin. This is equivalent to minimizing $\|\mathbf{w}\|$ while ensuring that the functional margin for each training example is greater than or equal to 1.

- **Convex Quadratic Programming**

The optimization problem in SVM is a convex quadratic programming problem, formulated as:

$$\min_{\mathbf{w}, b} \frac{1}{2} \|\mathbf{w}\|^2 \quad (3.45)$$

subject to:

$$y_i(\mathbf{w} \cdot \mathbf{x}_i + b) \geq 1, \quad \text{for all } i \quad (3.46)$$

- **Support Vectors and Margin Boundaries**

Support vectors are the data points that lie closest to the decision surface (or hyperplane). They are critical elements as they define the margin boundaries.

- **Dual Problem** The dual problem in SVM involves transforming the primary optimization problem into a dual form using Lagrange multipliers. The dual problem is given by:

$$\max_{\alpha} \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i,j=1}^n y_i y_j \alpha_i \alpha_j (\mathbf{x}_i \cdot \mathbf{x}_j) \quad (3.47)$$

subject to:

$$\sum_{i=1}^n \alpha_i y_i = 0 \quad \text{and} \quad \alpha_i \geq 0, \quad \text{for all } i \quad (3.48)$$

1. Support Vector Machine for approximately linearly separable data

In the context of SVM for data that is not completely separable, the concept of soft margins and slack variables is introduced to handle misclassifications. The optimization goal of SVM in such cases is to find a balance between maximizing the margin and minimizing the misclassification errors. This is achieved through convex quadratic programming and the introduction of a regularization term.

- **Soft Margin SVM**

When dealing with data that is not completely separable, SVM uses a soft margin approach. This approach introduces slack variables ξ_i to allow some misclassifications. The optimization problem is formulated as:

Objective Function:

$$\min_{w,b,\xi} \frac{1}{2} \|w\|^2 + C \sum_{i=1}^n \xi_i \quad (3.49)$$

Constraints:

$$y_i(w \cdot x_i + b) \geq 1 - \xi_i \quad \text{and} \quad \xi_i \geq 0 \quad \text{for all } i \quad (3.50)$$

Here, w is the weight vector, b is the bias, ξ_i are the slack variables representing the degree of misclassification, and C is the regularization parameter that controls the trade-off between maximizing the margin and minimizing the misclassification.

- **Dual Formulation**

The dual formulation of the soft margin SVM is derived using Lagrange multipliers. The dual problem is often preferred due to its computational efficiency and the ability to apply the kernel trick for non-linear classification.

– **Dual Objective Function:**

$$\max_{\alpha} \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i,j=1}^n y_i y_j \alpha_i \alpha_j (x_i \cdot x_j) \quad (3.51)$$

– **Constraints:**

$$\sum_{i=1}^n \alpha_i y_i = 0 \quad \text{and} \quad 0 \leq \alpha_i \leq C \quad \text{for all } i \quad (3.52)$$

In this dual problem, α_i are the Lagrange multipliers, and the constraints ensure that the solution lies within a specified range, governed by the regularization parameter C .

- **Hinge Loss Function**

The hinge loss function is a key component in the formulation of SVMs. It is defined as:

$$L(y, f(x)) = \max(0, 1 - y \cdot f(x)) \quad (3.53)$$

This loss function penalizes predictions that are on the wrong side of the margin, and its minimization leads to the optimal separating hyperplane in the SVM framework.

- **Regularization**

Regularization in SVMs is crucial for preventing overfitting, especially in cases where the data is not completely separable. It is achieved through the regularization parameter C , which controls the trade-off between maximizing the margin and minimizing the misclassification errors.

2. Non-linear Classification with SVM

SVMs are powerful tools for both linear and non-linear classification tasks. For non-linear classification, SVMs employ kernel functions to map input data into a higher-dimensional space where linear separation is feasible. This approach is known as the kernel trick.

- **Kernel Functions and the Kernel Trick** The kernel trick is a method used in SVMs to solve non-linear problems by transforming them into linear ones in a higher-dimensional space. The transformation is achieved through a kernel function $K(x, x')$, which computes the inner product of two vectors x and x' in the transformed feature space. Common kernel functions include polynomial kernels, Gaussian (Radial Basis Function - RBF) kernels, and others.
- **Positive Definite Kernels and Hilbert Spaces** Positive-definite kernels are functions that satisfy certain mathematical conditions, ensuring that the corresponding feature space is a Hilbert space. This property is crucial for the theoretical foundation of kernel methods in SVMs.
- **Non-linear SVM Classification Decision Function**

Given a non-linear SVM, the classification decision function is defined using a positive definite kernel function. The decision function for a new input x is given by:

$$f(x) = \text{sign} \left(\sum_{i=1}^N \alpha_i y_i K(x_i, x) + b \right), \quad (3.54)$$

where:

- α_i are the Lagrange multipliers obtained from the solution of the dual problem.
- y_i are the labels of the training data.
- x_i are the support vectors.
- $K(x_i, x)$ is the kernel function, a positive definite function used to compute the inner product in the transformed feature space.
- b is the bias term.
- N is the number of support vectors.

The kernel function $K(x_i, x)$ maps the input data into a higher-dimensional space where linear separation is possible. Common choices for K include polynomial kernels, Gaussian (RBF) kernels, and others, depending on the specific requirements of the dataset and problem. Fig. 3.17 shows the kernel trick.

- **Convex Quadratic Programming in Non-linear SVMs** When the kernel function is positive definite, the optimization problem of SVM is a convex quadratic programming problem, and the solution exists.

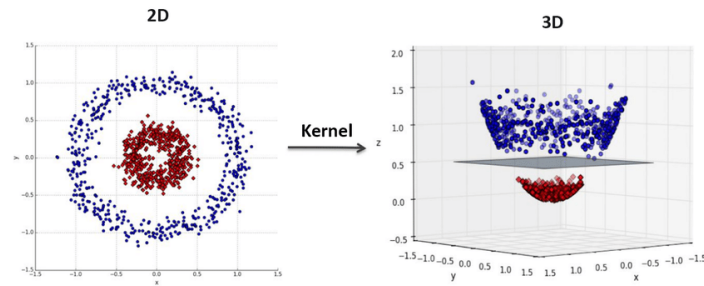


Figure 3.17: Kernel trick diagram (Hachimi et al., 2020).

3.4.6 UML algorithms

Unsupervised learning is a type of machine learning that deals with unlabeled data. Unlike supervised learning, where models are trained on data paired with known outputs, unsupervised learning algorithms infer patterns directly from the input data without reference to known, or labeled, outcomes.

The objectives of unsupervised learning can vary, including but not limited to:

- **Clustering:** One of the primary tasks in unsupervised learning is clustering, where the algorithm groups data points into distinct clusters based on their similarities. Common clustering algorithms include K-means, hierarchical clustering, the Gaussian Mixture Model based on the EM algorithm, and DBSCAN.
- **Dimensionality Reduction:** Unsupervised learning is often used to reduce the number of variables in a dataset while preserving as much information as possible. Techniques like PCA are widely used for dimensionality reduction, which will be discussed in detail later.
- **Association:** Discovering rules that describe large portions of the data, such as frequent itemsets or correlations.
- **Anomaly Detection:** Identifying rare items, events, or observations that raise suspicions by differing significantly from the majority of the data.

logistic regression vs. SVM In practical classification tasks, the results obtained from logistic regression and SVM often exhibit similarities. Logistic regression focuses on maximizing the conditional likelihood of the training set, continuously updating the posterior probabilities. This approach makes it more adept at handling outliers compared to SVM, which primarily concentrates on the points near the decision boundary. Another advantage of logistic regression is its inherent simplicity and ease of model updating, making it a convenient choice for many applications. In this section, we focus exclusively on two algorithms pertinent to this thesis: K-means and Gaussian Mixture Models (GMM). For a broader exploration of unsupervised learning algorithms, readers are directed to additional resources (Hinton & Sejnowski, 1999; Kotsiantis & Pintelas, 2004; Duda et al., 2012).

K-means algorithm

The K-means algorithm is a widely used method in unsupervised learning, particularly for clustering tasks. It aims to partition a set of observations into K clusters, where each observation belongs to the cluster with the nearest mean. The algorithm can be formalized as follows:

1. **Initialization:** Select K points as the initial centroids. These points can be chosen randomly or based on a specific strategy.
2. **Assignment Step:** Assign each observation to the cluster whose centroid is nearest. The distance is typically calculated using the Euclidean distance. Formally, this step involves partitioning the observations based on the given centroids. If we denote C_i as the set of observations in the i -th cluster, the assignment can be represented as:

$$C_i = \{x_j : \|x_j - \mu_i\|^2 \leq \|x_j - \mu_k\|^2 \text{ for all } k \neq i\}$$

where x_j is an observation, μ_i is the centroid of the i -th cluster, and $\|\cdot\|$ denotes the Euclidean norm.

3. **Update Step:** Recalculate the centroids of the clusters. The new centroid of each cluster is calculated as the mean of all observations in that cluster. Mathematically, the update of the centroid μ_i of the i -th cluster is given by:

$$\mu_i = \frac{1}{|C_i|} \sum_{x_j \in C_i} x_j$$

where $|C_i|$ is the number of observations in the i -th cluster.

4. **Repeat:** Steps 2 and 3 are repeated until the centroids no longer change significantly, indicating that the clusters have stabilized.

The objective of the K-means algorithm is to minimize the within-cluster sum of squares (WCSS), which is the sum of squared distances between each observation and its corresponding centroid. The WCSS can be expressed as:

$$\text{WCSS} = \sum_{i=1}^K \sum_{x_j \in C_i} \|x_j - \mu_i\|^2$$

. Fig. 3.18

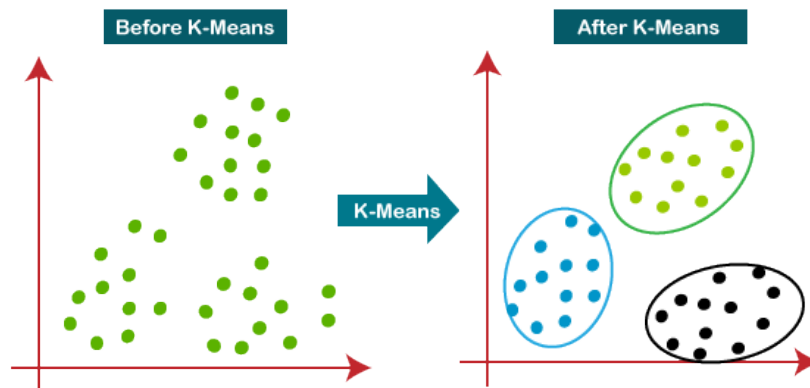


Figure 3.18: Schematic diagram of K-means algorithm (Raschka, 2015a).

Expectation-Maximization Algorithm

The Expectation-Maximization (EM) algorithm (Dempster et al., 1977) is an iterative method used for finding maximum likelihood estimates of parameters in statistical models, where the model depends on unobserved latent variables. The algorithm alternates between performing an expectation (E) step and a maximization (M) step.

- **Formal Definition**

Consider a statistical model with observed data X , unknown parameters θ , and latent variables Z . The likelihood of the observed data is given by $L(\theta; X) = p(X|\theta)$.

The EM algorithm seeks to maximize the log-likelihood $\log L(\theta; X)$, which is difficult due to the latent variables Z . The algorithm works iteratively as follows:

- **E Step**

Calculate the expected value of the log-likelihood function, concerning the conditional distribution of Z given X under the current estimate of the parameters $\theta^{(t)}$:

$$Q(\theta|\theta^{(t)}) = \mathbb{E}_{Z|X, \theta^{(t)}}[\log L(\theta; X, Z)]$$

- **M Step**

Find the parameter that maximizes this quantity:

$$\theta^{(t+1)} = \arg \max_{\theta} Q(\theta|\theta^{(t)})$$

- **Iterative Process**

1. **Initialize** $\theta^{(0)}$, the initial parameter estimates.
2. **Repeat until convergence:**

- **E Step:** Compute $Q(\theta|\theta^{(t)})$.
- **M Step:** Update $\theta^{(t+1)} = \arg \max_{\theta} Q(\theta|\theta^{(t)})$.

- **Jensen's Inequality in EM**

Jensen's inequality is crucial in the EM algorithm, especially in the E step. It provides a lower bound to the log-likelihood. The inequality states that for any random variable Y and a convex function f ,

$$f(\mathbb{E}[Y]) \leq \mathbb{E}[f(Y)]$$

In EM, this inequality allows deriving the Q function in the E step as a lower bound for the log-likelihood, which is then maximized in the M step. This approach guarantees that the log-likelihood increases with each iteration, ensuring convergence to a local maximum.

Gaussian Mixture Model

The Gaussian Mixture Model (GMM) is a probabilistic model that assumes all the data points are generated from a mixture of a finite number of Gaussian distributions with unknown parameters. GMMs are commonly used as a parametric model of the probability distribution of continuous measurements or features in a biometric system, such as the iris, fingerprint, or face recognition systems.

- **Formal Definition**

A Gaussian Mixture Model is a weighted sum of M component Gaussian densities as given by the equation:

$$p(x|\lambda) = \sum_{i=1}^M w_i \cdot g(x|\mu_i, \Sigma_i) \quad (3.55)$$

where:

- x represents the data points.
- λ represents the parameters of the mixture model.
- w_i are the mixture weights.
- $g(x|\mu_i, \Sigma_i)$ are the component Gaussian densities.
- Each component density is a multivariate Gaussian function of the form:

$$g(x|\mu_i, \Sigma_i) = \frac{1}{(2\pi)^{d/2} |\Sigma_i|^{1/2}} \exp\left(-\frac{1}{2}(x - \mu_i)^T \Sigma_i^{-1} (x - \mu_i)\right) \quad (3.56)$$

- μ_i and Σ_i are the mean and covariance of the i^{th} component.
- d is the dimensionality of the data.

- **Parameters Estimation**

The parameters of a GMM are typically estimated using the EM algorithm. The EM algorithm iteratively adjusts the parameters to maximize the likelihood of the data given in the model.

- **Expectation Step (E-step)**

In the E-step, the algorithm computes the responsibilities, which reflect the probability that each data point belongs to each cluster:

$$\gamma(z_{ik}) = \frac{w_k \cdot g(x_i|\mu_k, \Sigma_k)}{\sum_{j=1}^M w_j \cdot g(x_i|\mu_j, \Sigma_j)} \quad (3.57)$$

where $\gamma(z_{ik})$ is the responsibility of cluster k for data point i .

- **Maximization Step (M-step)**

In the M-step, the algorithm updates the parameters using the current responsibilities:

- Update the weights:

$$w_k^{new} = \frac{1}{N} \sum_{i=1}^N \gamma(z_{ik}) \quad (3.58)$$

- Update the means:

$$\mu_k^{new} = \frac{\sum_{i=1}^N \gamma(z_{ik}) x_i}{\sum_{i=1}^N \gamma(z_{ik})} \quad (3.59)$$

- Update the covariances:

$$\Sigma_k^{new} = \frac{\sum_{i=1}^N \gamma(z_{ik}) (x_i - \mu_k^{new})(x_i - \mu_k^{new})^T}{\sum_{i=1}^N \gamma(z_{ik})} \quad (3.60)$$

DL algorithms

DL, a transformative approach within ML, fundamentally relies on the concept of distributed representation. This paradigm posits that observations emerge from the complex interplay of various factors. DL extends this concept by proposing that these interactions can be decomposed into multiple layers, each representing different levels of data abstraction (Bengio et al., 2013a).

- **Hierarchical Abstraction in DL** DL employs a hierarchical abstraction process, where higher-level, more complex concepts are derived from lower-level, simpler ones. This hierarchical structure is typically constructed in a layer-wise manner, often employing greedy algorithms designed to select features that most effectively contribute to the learning process (Bengio et al., 2013a). A significant portion of DL algorithms falls under the category of unsupervised machine learning (UML), enabling their application to unlabeled datasets, which are more abundant and accessible compared to labeled data (Bengio et al., 2013a).
- **Representation Learning**
 - **Traditional ML Feature Learning**

Traditional ML feature learning includes techniques like Principal Component Analysis (PCA) (Pearson, 1901) and Linear Discriminant Analysis (LDA) (FISHER, 1936), as well as Manifold Learning, which emerged in 2000 to discover inherent structures in high-dimensional data.
 - **Deep Neural Network-Based Feature Learning**

Deep Neural Network-Based Feature Learning gained prominence with the demonstration of Rumelhart et al. (1986) that the backpropagation algorithm could learn useful intrinsic representations in a network's hidden layers. Hinton et al. (2006) introduced methods for greedy hierarchical pre-training and deep neural network fine-tuning, addressing neural networks' challenges with model overfitting and gradient diffusion (Hinton et al., 2006).
- **Artificial Neural Networks in DL**

The conceptual foundation of Artificial Neural Networks (ANNs), a cornerstone of DL, is rooted in the seminal work of Hubel & Wiesel (1959) on the neuronal structure of the human

brain. Minsky & Papert (1969) proved that the linear model of the perceptron cannot represent complex functions and cannot solve the exclusive OR (XOR) problem. The Universal Approximation Theorem (Cybenko, 1989; Chen et al., 1995; Lu et al., 2017) posits that a feedforward neural network (FNN) with a single hidden layer can approximate continuous functions.

- **The Depth in DL**

Deep neural networks (DNNs) are typically an extension of FNNs, although they have also been adapted into recurrent neural networks (RNNs) for specific applications like language modeling (Mikolov et al., 2010). The term “deep” in DL refers to the number of hidden layers in the network, denoting the depth of the credit assignment path (CAP).

- **Applications of DL**

In SML tasks, DL techniques eliminate the need for manual feature engineering by transforming data into compact, intermediate representations akin to principal components. These algorithms are also applicable to UML tasks, providing a significant advantage due to the abundance of unlabeled data. Various DL frameworks have been successfully applied in fields such as computer vision, speech recognition, natural language processing, audio recognition, and bioinformatics, yielding impressive results.

Artificial Neural Networks

The inception of artificial neural networks (ANNs) is deeply rooted in the study of the human central nervous system. These computational models are comprised of elementary units known as artificial neurons, which are interconnected in a manner akin to the neural networks observed in biological systems.

- **Definition and Structure**

There is currently no unified formal definition of ANN. However, it generally has three components: nodes or neurons, organized in layers, capable of handling tasks like pattern recognition, classification, and prediction. Each neuron is interconnected with others via links analogous to the axon-synaptic-dendritic connections found in biological systems. These nodes, upon receiving data through these connections, execute designated operations and tasks on the received information. Each link is characterized by a weight, which signifies the extent of influence one neuron exerts over another (Russell & Norvig, 2010, 2020). This weighting mechanism facilitates the selective transmission of signals between neurons, thereby enabling the network to process and learn from the input data effectively. Fig. 3.19 is the schematic diagram of human brain neurons, artificial neurons, and ANN. The following is a brief introduction to ANN:

- **Input Layer:**

- * Receives raw input data.
 - * Each neuron represents one feature of the input data.
 - * Formally represented as a vector $X = [x_1, x_2, \dots, x_n]$.

- **Hidden Layers:**

- * Situated between the input and output layers.
- * Apply transformations to inputs: $H = f(W_h \cdot X + B_h)$.

– **Output Layer:**

- * Produces the network's output.
- * Output computed as $Y = g(W_o \cdot H + B_o)$.

• **Neurons**

ANNs are composed of simulated neurons. Each neuron is connected to other nodes through links like biological axon-synaptic-dendritic connections. All nodes connected through a link receive some data and use it to perform specific operations and tasks on the data. Each link has a weight that determines the strength of the influence of one node on another node (Russell & Norvig, 2010, 2020), allowing the weight to select the signal between neurons. Neurons can be summarized as the following:

- Fundamental unit of ANNs.
- Computes a weighted sum and bias: $z = \sum_i (w_i \cdot x_i) + b$.
- Activation function introduces non-linearity.

• **Activation Functions**

Activation functions in ANNs introduce non-linearity, enabling the network to model complex data patterns. An activation function transforms the linear combination of inputs into a non-linear output:

$$f(z) = \text{Non-linear function}(z), \quad (3.61)$$

where z is the weighted sum of inputs and biases.

– **Common Types**

- * **Sigmoid Function:** Maps inputs to a range between 0 and 1.

$$\sigma(z) = \frac{1}{1 + e^{-z}}, \quad (3.62)$$

- * **Hyperbolic Tangent (tanh):** Outputs values between -1 and 1.

$$\tanh(z) = \frac{e^z - e^{-z}}{e^z + e^{-z}}, \quad (3.63)$$

- * **Rectified Linear Unit (ReLU):** Outputs the input if positive, else zero.

$$\text{ReLU}(z) = \max(0, z), \quad (3.64)$$

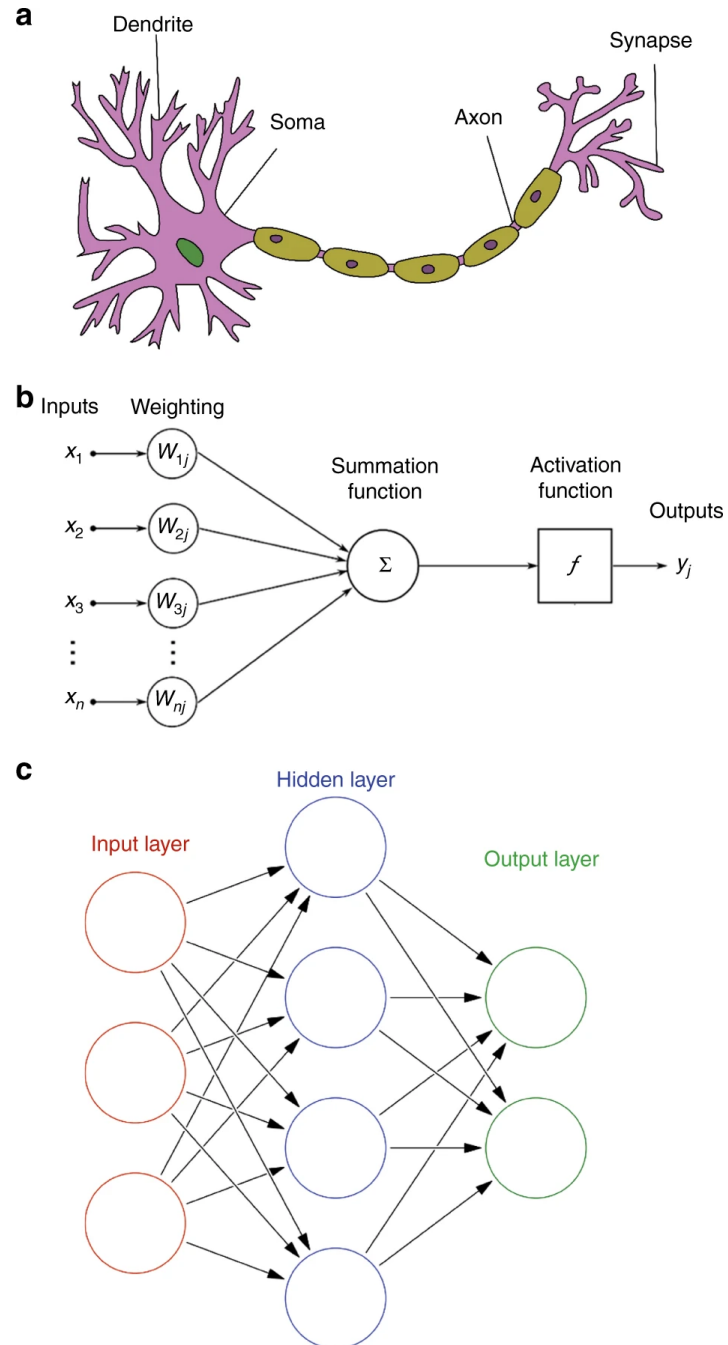


Figure 3.19: human brain neurons, artificial neurons, and ANN (Bre et al., 2017).

- * **Leaky ReLU**: Allows a small gradient when inactive.

$$\text{Leaky ReLU}(z) = \max(\alpha z, z), \quad (3.65)$$

where α is a small constant.

- * **Softmax Function**: Used in multi-class classification to turn logits into probabilities.

$$\text{Softmax}(z_i) = \frac{e^{z_i}}{\sum_j e^{z_j}}$$

– Backpropagation

Backpropagation calculates the gradient of the loss function concerning the network weights. Backpropagation computes the gradient of the loss function for each weight by the chain rule, propagating the error backward.

– Process

1. **Forward Pass**: Compute output and loss.
2. **Backward Pass**: Compute the gradient of the loss function.

$$\frac{\partial \mathcal{L}}{\partial w} = \frac{\partial \mathcal{L}}{\partial y} \cdot \frac{\partial y}{\partial w}$$

where \mathcal{L} is the loss function, and y is the network output.

3. **Update Weights**: Adjust weights to minimize loss.

$$w = w - \eta \cdot \frac{\partial \mathcal{L}}{\partial w}$$

where η is the learning rate.

– Regularization

In addition to the common L1 (Lasso) and L2 (Ridge) regularization, Dropout is also a regularization technique that randomly deactivates a subset of neurons during training. This process is mathematically represented as:

$$h' = h \odot \text{Dropout}(\rho), \quad (3.66)$$

where h is the input to the dropout layer, and h' is the output after applying dropout (Wan et al., 2013).

– Formal Representation

A simple three-layer ANN:

- * Input: $X = [x_1, x_2, \dots, x_n]$
- * Hidden: $H = f(W_h \cdot X + B_h)$
- * Output: $Y = g(W_o \cdot H + B_o)$

Disadvantages of DL

- **Gradient Vanishing and Gradient Exploding**

DL is not omnipotent and will still face some difficulties. In the context of DL, particularly in training deep neural networks (DNNs), two significant challenges are the phenomena of *gradient vanishing* (Hochreiter, 1991) and *gradient exploding* (Pascanu et al., 2012a). These issues are pivotal in understanding the limitations and behavior of DNNs during the training process.

- **Gradient Vanishing**

Gradient vanishing refers to the problem where the gradients of the loss function concerning the network's parameters become increasingly small as the error is backpropagated through the layers. This effect is more pronounced in deeper networks. Mathematically, it can be expressed as:

$$\lim_{l \rightarrow 0} \frac{\partial \mathcal{L}}{\partial w^{(l)}} \approx 0, \quad (3.67)$$

where \mathcal{L} is the loss function, $w^{(l)}$ represents the weights of the l -th layer, and l denotes the layer index, counting backward from the output layer.

- **Causes**

The primary cause of gradient vanishing is the use of certain activation functions, like the sigmoid or hyperbolic tangent, which saturate at their asymptotes, leading to derivatives close to zero. In deep networks, as the gradient is backpropagated, the multiplication of these small derivatives leads to an exponentially smaller gradient for layers closer to the input layer Hochreiter & Schmidhuber (1997); Glorot & Bengio (2010).

- **Gradient Exploding**

Conversely, gradient exploding occurs when the gradients of the loss function become excessively large, potentially leading to divergent behavior in the learning process. This is represented as:

$$\lim_{l \rightarrow 0} \left| \frac{\partial \mathcal{L}}{\partial w^{(l)}} \right| \rightarrow \infty. \quad (3.68)$$

- **Causes**

Gradient exploding is often observed in networks with deep architectures and is typically due to the multiplication of gradients larger than one through the layers during backpropagation. This is particularly problematic with activation functions that do not have an upper bound on their derivatives Pascanu et al. (2012b).

- **Mitigation Strategies for Gradient Vanishing**

- * **Use of ReLU Activation Function:** The Rectified Linear Unit (ReLU) and its variants (e.g., Leaky ReLU) help mitigate the vanishing gradient problem due to their non-saturating nature Nair & Hinton (2010).
- * **Weight Initialization Techniques:** Proper initialization methods, such as Xavier or He initialization, can prevent gradients from diminishing too quickly He et al. (2015).

- * **Batch Normalization:** Normalizing the inputs of each layer ensures that they have a mean of zero and a variance of one, which helps maintain stable gradients Ioffe & Szegedy (2015).

– Mitigation Strategies for Gradient Exploding

- * **Gradient Clipping:** This technique involves scaling down gradients when they exceed a certain threshold, thus preventing them from becoming too large Pascanu et al. (2012b).
- * **Careful Network Design:** Avoiding excessively deep architectures or using skip connections (as in ResNet) can reduce the risk of exploding gradients.

• Interpretability and Opacity in DL

DL has achieved remarkable success across various domains, from image recognition to natural language processing. However, the interpretability of these models, often referred to as their "black box" nature, remains a significant challenge in the field.

– The Black Box Nature of Deep Learning

DL models, particularly DNNs, are often considered opaque due to their complex structure and the high dimensionality of the data they process. This opacity arises from the intricate interplay of numerous parameters within these models, making it challenging to understand the specific reasons behind their decisions or predictions Castelvechi (2016).

– Challenges in Interpretability

- * **Complexity of Models:** DNNs can consist of millions of parameters and numerous layers, making it difficult to trace how input data is transformed into outputs.
- * **Non-linearity:** The non-linear activation functions used in DNNs add to the complexity, as they create non-linear mappings between inputs and outputs.
- * **High Dimensionality:** DL models often work with high-dimensional data, which is inherently difficult to visualize or interpret in human-understandable terms.

Transfer learning

Transfer learning (TL) is an ML method where a model developed for a task is reused as the starting point for a model on a second task. It is a popular approach in deep learning where pre-trained models are used as the starting point for computer vision and natural language processing tasks. The formidable capabilities of ANNs in feature learning and nonlinear fitting render them the optimal choice for implementing Transfer Learning. These networks excel in discerning intricate patterns and relationships within data, a trait that is crucial for the successful application of Transfer Learning methodologies.

- **Historical Background** The concept of transfer learning dates back to the 1990s, but it gained significant traction in the 2010s with the advent of deep learning. The success of transfer learning in deep learning is largely attributed to the introduction of large-scale datasets and the development of high-capacity models that can learn rich representations from these datasets West et al. (2007); Karimpanal & Bouffanais (2019); Bozinovski (2020).
- **Formal Definition**

TL involves leveraging knowledge from a source domain \mathcal{D}_S with its task \mathcal{T}_S to enhance learning in a different target domain \mathcal{D}_T and its task \mathcal{T}_T . In TL, a domain \mathcal{D} comprises a feature space \mathcal{X} and a marginal probability distribution $P(X)$, while a task \mathcal{T} includes a label space \mathcal{Y} and a predictive function f . The function f maps inputs $x \in \mathcal{X}$ to labels $y \in \mathcal{Y}$, learned from input-output pairs x_i, y_i . TL is particularly effective in imbalanced clustering tasks, using knowledge from \mathcal{D}_S to improve performance in \mathcal{D}_T , especially when \mathcal{D}_T has limited samples. This approach helps in modeling minority classes and addressing class imbalance challenges. The process of TL is illustrated in Fig. 3.20.

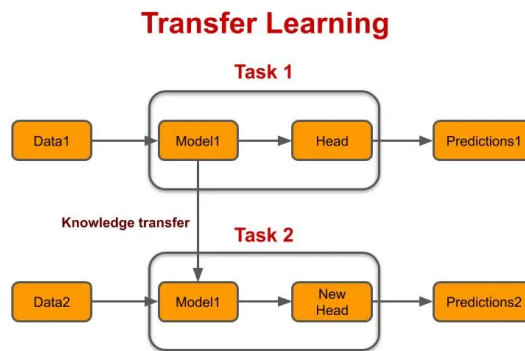


Figure 3.20: TL process diagram <https://medium.com/modern-nlp/transfer-learning-in-nlp-f5035cc3f62f>.

- **TL in Practice** In practice, transfer learning can be implemented in several ways, such as:
 - **Feature Extraction:** Using the representations learned by a previous network to extract meaningful features from new samples.
 - **Fine-Tuning:** Adjusting the weights of a pre-trained network by continuing the back-propagation process, allowing the model to learn task-specific features from the new dataset.

3.4.7 Data Preprocessing

Data preprocessing is a fundamental step in the ML workflow (Russell & Norvig, 2010, 2020; Raschka, 2015a), pivotal for enhancing the quality and utility of data for effective model training. This section explores the intricate processes and methodologies involved in data preprocessing, underscoring its critical role in ML.

- *Data Cleaning.* Dealing with missing values is a critical aspect of data preprocessing in ML. The presence of missing data can significantly impact the performance of models, making it essential to address them effectively. This section delves into various strategies for handling missing values.
 1. **Deletion Methods.** The simplest approach is to ignore or remove the rows or columns with missing values. While straightforward, this method, known as listwise or pairwise deletion, can lead to significant data loss and potential bias, especially if the missingness is not random.
 2. **Imputation Techniques.** Imputation involves substituting missing values with estimated ones. The basic form of imputation replaces missing values with central tendency measures such as mean, median, or mode. More sophisticated techniques include:
 3. **KNN Imputation.** This method estimates missing values based on the k-nearest neighbors, considering the similarity between instances.
 4. **Regression Imputation.** Missing values are predicted using regression models, where other variables in the dataset serve as predictors.
 5. **Multiple Imputation.** It involves creating multiple imputations for missing values to reflect the uncertainty around the true value. This approach is particularly useful when the data is missing at random (MAR).
 6. **Algorithmic Approaches.** Certain machine learning algorithms can handle missing values inherently. For instance, decision trees and random forests can split nodes using available data without imputing missing values
 7. **Data Augmentation.** In some cases, missing values can be addressed through data augmentation, where synthetic data is generated to fill in the gaps. This approach is often used in DL contexts.
- *Feature Discretization.* Feature discretization converts a range of continuous values into a smaller number of intervals or bins. This process can be particularly useful when dealing with non-linear relationships in the data, as it allows inherently linear models (like linear regression or LR) to capture non-linear effects. Here are many common methods of discretization (Dougherty et al., 1995; Kotsiantis & Kanellopoulos, 2005; Huang et al., 2014).
 1. **Equal-width binning.** Divide the range of values into intervals of equal size. The main disadvantage is its sensitivity to outliers, as they can skew the distribution of the bins.
 2. **Equal-frequency binning.** Each bin contains approximately the same number of data points. This method is less sensitive to outliers but can lead to bins with very different width ranges.
 3. **K-means binning.** Uses the K-means clustering algorithm to partition data into k clusters. Each cluster forms a bin, and the boundaries are determined by the cluster centroids.
 4. **Decision tree-based binning.** Utilizes decision trees to determine optimal binning. The splits made by the decision tree define the bin boundaries, which can capture complex structures in the data.

- *Dealing with categorical.* In the realm of ML, the treatment of nominal and ordinal features is a crucial aspect of data preprocessing, as it significantly influences the performance of the learning algorithms. This subsection delves into the methodologies employed for effectively processing these categorical data types.
 1. **Nominal Features.** Nominal features, representing discrete and unranked data categories, necessitate specific preprocessing techniques to transform them into a format suitable for machine learning algorithms. The most prevalent method is one-hot encoding, where each category is transformed into a new binary feature. This approach effectively deals with the non-ordinal nature of nominal data, ensuring that machine learning models do not erroneously attribute an order or priority to these categories. However, one-hot encoding can lead to a high-dimensional feature space, especially with features with many categories, which might necessitate dimensionality reduction techniques.
 2. **Ordinal Features.** Ordinal features, in contrast, are categorical data types with a natural order or ranking among the categories. The primary challenge in processing ordinal features is to encode the inherent order in the data appropriately. Common techniques include label encoding, where each category is assigned a unique integer based on its order. This method preserves the ordinal relationship between categories, making it suitable for algorithms that can interpret numerical values and their order. Another approach is to use binary encoding, which reduces the feature space compared to one-hot encoding while maintaining the order of information.

3.4.8 Splitting Dataset into Training, Validation, and Testing Sets

Splitting a dataset into training, validation, and testing sets is a crucial step in the development of machine learning models. This process helps in assessing the model's performance and ensuring that it generalizes well to new, unseen data.

- *Standard Splitting Methods*
 1. **Random Split.** The dataset is randomly divided into training, validation, and testing sets. A common split ratio is 70% for training, 15% for validation, and 15% for testing.
 2. **Stratified Split.** This method is used to ensure that each split of the dataset contains approximately the same percentage of samples of each target class as the original dataset. It is particularly useful when dealing with imbalanced datasets.
 3. **Time-based Split.** In time-series data, the dataset is split based on time, ensuring that the training set contains earlier data while the validation and testing sets contain later data.
- *Handling Imbalanced Datasets.* Imbalanced datasets, where some classes are underrepresented, pose a significant challenge in model training and evaluation. Specialized techniques are used to ensure a fair representation of all classes in each dataset split:
 1. **Stratified Sampling.** Ensures that each class is adequately represented in each split of the dataset. This method maintains the same class proportion in training, validation, and testing sets as in the original dataset.

2. **Oversampling and Undersampling.** Involves adjusting the class distribution of the dataset. Oversampling the minority class or undersampling the majority class can be applied before the split to balance the dataset.
 3. **Synthetic Data Generation.** Techniques like the Synthetic Minority Over-sampling Technique (SMOTE) can be used to generate synthetic samples for minority classes before the dataset split.
- *Feature scaling.* Feature scaling is a crucial step in the preprocessing of data for machine learning algorithms. It involves adjusting the scale of the features in the dataset to a common range or distribution. This is important because features on different scales can disproportionately influence the model's learning process. Two common methods of feature scaling are normalization and standardization.
 - **Normalization.** Normalization, often referred to as Min-Max scaling, rescales the feature values to a range of [0, 1]. The formula for normalization is given:

$$x_{\text{normalized}} = \frac{x - \min(x)}{\max(x) - \min(x)} \quad (3.69)$$

where x is an individual data point, $\min(x)$ is the minimum value of the feature across all data points, and $\max(x)$ is the maximum value of the feature. This scaling method is useful when the algorithm assumes the data to be on a bounded interval.

- **Standardization.** Standardization, on the other hand, rescales data to have a mean (μ) of 0 and a standard deviation (σ) of 1. The formula for standardization is:

$$x_{\text{standardized}} = \frac{x - \mu}{\sigma} \quad (3.70)$$

, where x represents an individual data point, μ is the mean of the feature, and σ is the standard deviation of the feature. Standardization is less affected by outliers and is often used when the algorithm assumes data to be normally distributed.

Both normalization and standardization are essential techniques in data preprocessing, ensuring that each feature contributes equally to the model's performance and improving the convergence of the algorithm.

3.4.9 Data Reduction

Data compression, a pivotal technique in the realm of machine learning and data science, plays a crucial role in managing the ever-growing volume of data in modern computational tasks. It involves reducing the size of a dataset while retaining its essential information, thereby enhancing storage efficiency and computational speed. In the context of ML, data compression is not merely about reducing storage requirements. It is also instrumental in addressing the “curse of dimensionality”, a term coined by Bellman et al. (1957); BELLMAN (1961) in the context of dynamic programming, refers to various phenomena that arise when analyzing and organizing data in high-dimensional spaces, often encountered in fields like ML and data analysis. As the dimensionality of a dataset increases, the volume of the space increases exponentially, making the available data sparse. This sparsity is problematic for any method that requires statistical significance, as more data is needed to support the model reliably.

In high-dimensional spaces, the distance between data points becomes less meaningful, as highlighted by Beyer et al. (1999). This undermines the effectiveness of many ML algorithms, especially those relying on distance calculations, like KNN. The increased dimensions also lead to overfitting, where a model learns noise in the training data instead of the intended outputs, reducing its predictive power on new, unseen data.

Dimensionality reduction techniques, such as PCA and LDA, are often employed to mitigate the curse of dimensionality. These methods reduce the number of random variables under consideration, either by selecting only the most relevant variables (feature selection) or by transforming them into a lower-dimensional space (feature extraction), as discussed by Jolliffe (2002) and Hastie et al. (2009).

- **Feature selection** The primary goal of feature selection is to improve model performance, whether measured by accuracy, precision, speed, or robustness. It also helps in understanding the underlying processes that generated the data, as noted by (Guyon & Elisseeff, 2003). Feature selection methods fall into three broad categories: filter methods, wrapper methods, and embedded methods (see Fig. 3.21).

Feature selection, an essential process in ML, involves selecting a subset of relevant features (variables, predictors) for use in model construction. This process simplifies models to make them easier to interpret, reduces the computational cost of modeling, and, most importantly, can enhance the generalization capabilities of the model by eliminating irrelevant or redundant data.

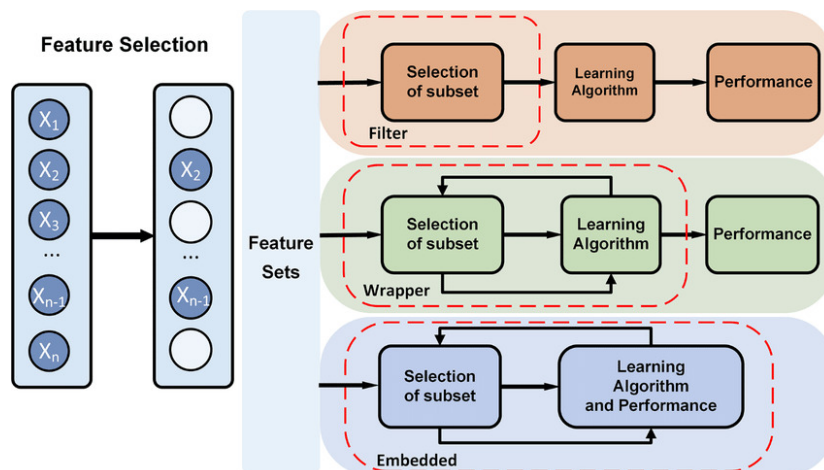


Figure 3.21: Three types of methods for Feature selection: filter, wrapper, and embedded method (Xie et al., 2020).

- **Filter methods** Filter methods in machine learning are a set of techniques used for feature selection, primarily based on statistical measures. They evaluate the importance of features independently of any machine learning algorithm. This independence from the predictive model makes filter methods computationally efficient, and particularly useful for high-dimensional datasets. Here are some of the common filter methods (Koller & Sahami, 1996; Cover & Thomas, 2006; Amorim, 2009).

1. **Variance Threshold.** This method removes all features whose variance does not meet a certain threshold. It operates under the assumption that features with low

variance are less informative. The variance threshold is defined as: $\text{Var}(X) = \sigma^2 = \frac{1}{N} \sum_{i=1}^N (x_i - \mu)^2$, where σ^2 is the variance, N is the number of observations, x_i is each individual observation, and μ is the mean of the feature.

2. **Pearson Correlation Coefficient.** This method evaluates the linear correlation between two variables. The Pearson correlation coefficient is calculated as:

$$r = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}}, \quad (3.71)$$

where x_i and y_i are the individual sample points indexed with i , \bar{x} and \bar{y} are the sample means. A high absolute value of r indicates a strong relationship between the features.

3. **Mutual Information (Information Gain).** Mutual information measures the amount of information one can obtain from one random variable given another. The mutual information between two variables X and Y is calculated as:

$$I(X; Y) = \sum_{y \in Y} \sum_{x \in X} p(x, y) \log \left(\frac{p(x, y)}{p(x)p(y)} \right), \quad (3.72)$$

where $p(x, y)$ is the joint probability distribution function of X and Y , and $p(x)$ and $p(y)$ are the marginal probability distribution functions of X and Y respectively.

4. **Chi-Squared Test.** This statistical test is used to determine if there is a significant association between two categorical variables. The formula for the chi-squared statistic is:

$$\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i}, \quad (3.73)$$

where O_i is the observed frequency, and E_i is the expected frequency under the null hypothesis.

- **Wrapper methods.** Wrapper methods in ML are feature selection techniques that use a predictive model to evaluate the combination of features and determine their effectiveness. Unlike filter methods, wrapper methods consider the interaction between features and are typically more computationally intensive. Here are some of the common wrapper methods: (Pudil et al., 1994; Kohavi & John, 1997; Saeys et al., 2007).

1. **Sequential Forward Selection (SFS).** This method starts with an empty model and successively adds features that improve the model's performance the most until a specified stopping criterion is reached. The general idea can be expressed as:

$$S_{k+1} = S_k \cup \{x^+\}, \quad (3.74)$$

where S_k is the set of selected features at step k , and x^+ is the feature that maximizes the performance criterion when added to S_k .

2. **Sequential Backward Selection (SBS).** This method starts with the full set of features and iteratively removes the least significant feature that causes the least performance loss. The process is described as:

$$S_{k-1} = S_k - \{x^-\}, \quad (3.75)$$

where S_k is the set of selected features at step k , and x^- is the feature that, when removed from S_k , has the least impact on performance.

3. **Recursive Feature Elimination (RFE).** RFE recursively removes the least important features based on the model weights. It is particularly useful with models that assign importance to each feature, such as linear models with coefficients. The process can be summarized as:

$$S_{k-1} = S_k - \{\text{least important feature}\} \quad (3.76)$$

until the desired number of features is reached.

- **Embedded methods.** Embedded methods are a type of feature selection technique in ML that performs feature selection as part of the model training process. These methods are particularly efficient as they combine the qualities of filter and wrapper methods, offering a balance between computational efficiency and model performance. Here are some of the common wrapper methods (Quinlan, 1986; Tibshirani, 1996; Breiman, 2001b).

1. **L1 (LASSO) Regularization.** L1 regularization adds a penalty equal to the absolute value of the magnitude of coefficients. This method can shrink some coefficients to zero, effectively performing feature selection. The optimization problem can be formulated as:

$$\min_{\beta} \left\{ \frac{1}{2n} \|Y - X\beta\|_2^2 + \lambda \|\beta\|_1 \right\}, \quad (3.77)$$

where β represents the coefficients, X is the feature matrix, Y is the target vector, n is the number of samples, and λ is the regularization parameter.

2. **Decision tree.** DT inherently performs feature selection by selecting the most informative features to split on at each node. The selection is based on measures like Gini impurity or information gain.
3. **Random Forest.** RF extends the concept of decision trees. Each tree in the forest is built from a sample drawn with replacement from the training set. When splitting a node during the construction of the tree, the split that is chosen is the best split among a random subset of the features. This randomness helps in achieving a more diversified set of features.

- **Feature extraction.** Feature extraction (Sarangi et al., 2020) is a process in ML where new features are constructed from the original set of features to reduce the dimensionality of the dataset. This process is crucial for improving the efficiency of machine learning algorithms, especially when dealing with high-dimensional data. Feature extraction techniques transform the input data into a set of features that are more meaningful and informative for the learning algorithms.

Three of the most common feature extraction methods are PCA (Pearson, 1901; Jolliffe, 2014), LDA (FISHER, 1936), and Kernel Principal Component Analysis (Kernel PCA, Schölkopf et al. 1997; Schölkopf et al. 1998).

1. **PCA.** PCA is a statistical technique used to emphasize variation and capture strong patterns in a dataset. It transforms the original data into a new set of variables, the principal components, which are orthogonal and uncorrelated. The first principal component accounts for the most variance in the data. The mathematical formulation of PCA involves an eigenvalue decomposition of the data covariance matrix or singular value decomposition of the data matrix. The PCA transformation is defined as:

$$Y = W^T \cdot X, \quad (3.78)$$

where X is the data matrix, W is the matrix of eigenvectors, and Y is the matrix of principal components.

2. **LDA.** LDA is a SML method used to find the linear combinations of features that best separate two or more classes of objects or events. The resulting combination may be used as a linear classifier or for dimensionality reduction. The goal of LDA is to maximize the ratio of between-class variance to the within-class variance in any particular data set, thereby ensuring maximum separability. The LDA transformation can be represented as:

$$W = \arg \max \frac{|W^T S_B W|}{|W^T S_W W|}, \quad (3.79)$$

where S_B is the between-class scatter matrix and S_W is the within-class scatter matrix.

3. **Kernel PCA.** Kernel PCA applies a nonlinear kernel function to the original data to map it into a higher-dimensional feature space where linear separation is possible. In this new space, the standard PCA is then applied to find the principal components. The kernel function, such as the polynomial, radial basis function (RBF), or sigmoid kernel, defines the nature of the high-dimensional space. The transformation in Kernel PCA can be represented as: $\Phi : x_i \mapsto \Phi(x_i)$, $Y = \Phi(X)W$ where x_i are the original data points, Φ is the nonlinear mapping function defined by the kernel, X is the matrix of original data, W is the matrix of eigenvectors in the feature space, and Y is the transformed data in the kernel principal component space.

Fig. 3.22 and Fig. 3.23 visualize the PCA, LDA, and Kernel PCA.

- **Hyperparameters.** Hyperparameters are external configurations to the model that are not learned from data but are set before the training process. They are crucial in controlling the behavior of the learning algorithm and can significantly impact the performance of the model.

Selecting the optimal hyperparameters is a critical step, often achieved through processes like grid search, random search, or Bayesian optimization:

- **Grid Search.** Grid search is a methodical approach to hyperparameter tuning that involves exhaustively searching through a specified subset of hyperparameters. It defines a grid of hyperparameter values and evaluates the model's performance for each combination in the grid. Grid search is simple to understand and implement. Guarantees finding the optimal solution if it lies within the grid. However, it is computationally intensive, especially with a large number of hyperparameters or large value ranges.

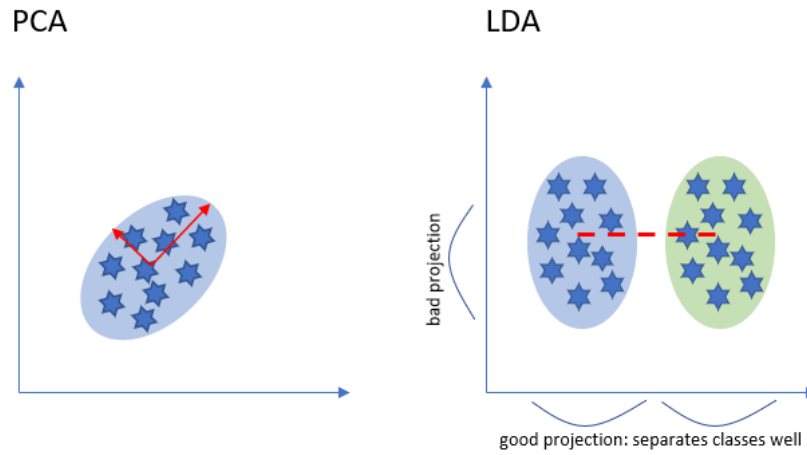


Figure 3.22: PCA versus LDA. Figure courtesy of https://sebastianraschka.com/Articles/2014_python_lda.html

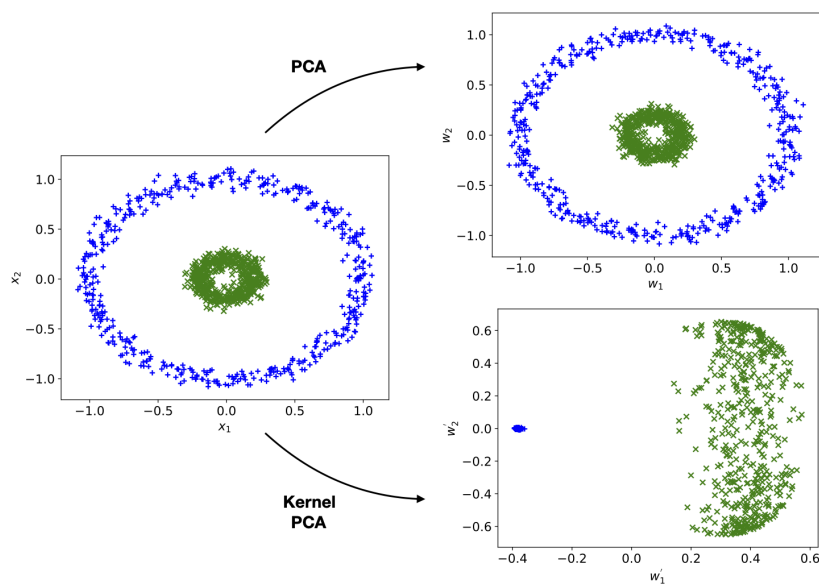


Figure 3.23: PCA versus Kernel PCA. Figure courtesy of https://ml-lectures.org/docs/structuring_data/ml_without_neural_network-2.html

- **Random Search.** Random search selects random combinations of hyperparameters to evaluate. It defines a search space as ranges for hyperparameters and randomly selects combinations for evaluation. Random search is more efficient than grid search, especially when only a few hyperparameters significantly influence the model. However, there is no guarantee of finding the optimal solution, and the search might miss important areas of the hyperparameter space.
- **Bayesian Optimization.** Bayesian optimization uses a probabilistic model to guide the search for the best hyperparameters. It builds a probabilistic model of the function mapping from hyperparameter values to the target evaluated on the validation set. Bayesian optimization is more efficient than random or grid search, particularly for high-dimensional spaces. Nevertheless, bayesian optimization is more complex to implement and understand. Requires careful choice of the prior and the update rule.

3.5 ML/DL applications in astronomy

In the era of astronomical big data, driven by advancements in observational tools, there has been an exponential increase in astronomical data. For instance, the Hubble Space Telescope (HST) Data Warehouse provides access to engineering data from 1990 to 2002, with a current data volume of 2.8 TB (Miebach, 2002). The Sloan Digital Sky Survey (SDSS) had amassed 40 TB of data by the early 21st century (York et al., 2000), while the Large Synoptic Survey Telescope (LSST) is expected to collect approximately 30 TB of data nightly over a decade, potentially discovering 100,000 variable objects each night. The total image data volume of LSST is projected to be around 70 PB, with catalogs estimated to be 10-20 PB (Borne, 2008). Gaia (Gaia Collaboration et al., 2016) is mapping a three-dimensional chart of the Milky Way and will provide precise positional and radial velocity measurements for over a billion stars in the Milky Way and the entire Local Group. The Jet Propulsion Laboratory Adaptive Recognition Tool (JARTool, Burl et al. 1994) is used for identifying volcanoes in over 10,000 Venusian images returned by the Magellan mission. The Euclid space mission is expected to generate data in the tens of PB (An, 2019), and the Square Kilometre Array will produce 700 PB of scientific data annually. In this data-intensive era of astronomy, facing large, complex, and multidimensional datasets, there is a need for automated, efficient, and intelligent methods capable of performing tasks akin to those of astronomers, such as identifying known objects, discovering unknown entities, and recognizing rare phenomena. All these tasks can be attempted using ML/DL technologies (Wang et al., 2018; Rodriguez et al., 2022; Patil, 2023).

The surge in astronomical data necessitates the adoption of cutting-edge advanced data processing methods for data processing and analysis. This era of data-intensive astronomy calls for automated, efficient, and intelligent methods capable of performing tasks akin to those of astronomers, such as recognizing known objects, discovering unknown entities, and identifying rare phenomena, all achievable through ML/DL techniques (Wang et al., 2018; Rodriguez et al., 2022; Patil, 2023). Many astronomical classification tasks, such as the classification of stars and galaxies, galaxy morphology, and AGN/blazar classification, have successfully applied ML/DL technologies. Some detailed examples are provided below.

- **Classification of Stars and Galaxies.** Distinguishing between stars and galaxies is a fundamental task in astronomical data analysis. The most common method for this purpose is morphological classification, which relies on the appearance, shape, and size of objects in images. Stars, being relatively closer to Earth, typically appear as more or less circular point sources,

while galaxies, being farther away, usually present extended structures with varying shapes and luminosity distributions. A widely used technique is the analysis of light profiles. Stars, due to their point-like nature, have light profiles that can be well described by a PSF. On the other hand, galaxies have more complex light distributions that cannot be fully captured by PSF models. This distinction forms the basis for star-galaxy classification in many surveys, including the Sloan Digital Sky Survey (SDSS) (Stoughton et al., 2002). However, other astrophysical objects, such as quasars and supernovae, also appear as point sources. Therefore, dividing photometric catalogs into stars and galaxies, or more generally, stars, galaxies, and other objects, is an important issue. ML algorithms, particularly supervised learning methods, have become increasingly popular in this task. These algorithms can be trained on labeled datasets containing known stars and galaxies, learning to recognize patterns and features that differentiate these two categories. For example, SVMs have been effectively used for star-galaxy classification (Fadely et al., 2012). DL, which has seen significant success in recent years, allows ANNs to automatically learn hierarchical features from image data, making them highly suitable for classifying a diverse range of astronomical objects Kennamer et al. (2018); Kim & Brunner (2016). Devices like APM and DPOSS, which scan photographic plates, have used various ML/DL algorithms. In addition to ANNs (Odewahn et al., 1992, 1993; Bazell & Peng, 1998; Andreon et al., 2000; Philip et al., 2002; Odewahn et al., 2004; Collister et al., 2007), there are DTs (Odewahn et al., 2004; Weir et al., 1995; Ball et al., 2006), and hybrid modeling (Qin et al., 2003), most of which exceed 95% efficiency. This efficiency is usually achieved by analyzing a set of morphological parameters extracted from photometric measurements, often supplemented with additional data such as color or seeing conditions. These studies demonstrate the efficacy of ML/DL algorithms in accurately classifying stars and galaxies in large astronomical datasets.

- **Classification of Galaxy Morphology.** Galaxies exhibit a variety of sizes and morphologies. The Hubble Sequence stands out as the primary method for classifying galaxy morphology, categorizing galaxies into ellipticals, spirals, barred spirals, and irregulars, each with further subdivisions (Fath, 1909; Hubble, 1926; Sandage, 1961; van den Bergh, 1998). This system is linked to key physical characteristics that are crucial for the formation and evolution of galaxies (Roberts & Haynes, 1994; Firmani & Avila-Reese, 2003). Other classification systems, such as the Yerkes system, utilize the concentration index (Morgan, 1958, 1959; 197, 1977), while the de Vaucouleurs index and S´ersic index focus on the light profile of galaxies (de Vaucouleurs, 1948; Freeman, 1970; Sersic, 1968; Graham & Driver, 2005).

The Hubble Sequence has been extended, particularly for data mining, using the T system, which numerically represents Hubble types from E to Irr as values from -5 to 10. Studies classifying galaxies such as E/S0, Sabc, and Sd/Irr based on surface brightness and light profiles using ANNs include works by (Odewahn et al., 1996; Windhorst et al., 1999; Driver et al., 2003). Additionally, Odewahn et al. (2002) applied Fourier decomposition and ANNs to galaxy images for bar detection and T-type assignment. The innovative Galaxy Zoo project Masters et al. (2019) employs crowdsourcing for morphological classification, inviting the public to categorize SDSS images. This project introduced new categories, such as edge-on views and spiral arm winding, and generated numerous scientific insights. While slower than automated algorithms, this method provides a valuable complement to them. Recent advancements in DL (especially CNNs) have played a key role in galaxy shape classification. Studies have shown that AI can automatically classify galaxy shapes by training CNNs with crowdsourced data from Galaxy Zoo (Dieleman et al., 2015). Although matching human accuracy, these networks also inherit human biases.

- **Classification of AGN/Blazars.** Identifying AGNs/quasars from astronomical observations is a key and complex task, well suited for data mining techniques. It is well-known that active galactic nuclei are detected to varying degrees across different wavebands (X-ray, optical, radio), with no single waveband capable of identifying all AGNs. Traditional classification of AGNs often uses the BPT (Baldwin et al., 1981), delineating AGN star-forming regions through a curved demarcation line. Data mining enhances this approach by enabling more refined or multidimensional separation and incorporating passive objects into the analysis. This advancement facilitates the calculation of the likelihood of an object harboring an AGN, without the necessity of detecting all or any specific emission lines.

Several research groups have employed ANNs (Carballo et al., 2004; Claeskens et al., 2006; Carballo et al., 2008) and DTs (Suchkov et al., 2005; Ball et al., 2006; ?) to select quasar candidates from survey data. Other methodologies include PCA (Yip et al., 2004), SVM and Learning Vector Quantization (Zhang & Zhao, 2003), k-d trees (Gao et al., 2008), and clustering techniques like principal surfaces and negative entropy clustering (D'Abrusco et al., 2009), along with kernel density estimation (Richards et al., 2009). Many of these studies integrate multiwavelength data, particularly from X-ray, optical, and radio sources. Similarly, various types of AGN candidates can be effectively selected and classified (Sanchez Saez, 2022). The availability of multiwavelength data enhances the capability of data mining algorithms to construct models of requisite complexity, thereby leveraging the full spectrum of information to extract more comprehensive AGN samples.

Certain blazars, due to the lack of optical spectroscopic information, cannot be definitively categorized as either BL Lac or FSRQ. These sources are referred to as blazar candidates of an unknown type (BCU) in the Fermi-LAT catalog. Numerous outstanding works on BCU classification have employed a variety of ML/DL techniques, such as LR, SVM, DT, RF, XGBoost, CatBoost, ANN, GMM, K-means, etc., with accuracies exceeding 90% (Chiaro et al., 2016; ?; Kang et al., 2019; Fan et al., 2022; Xiao et al., 2022b; Agarwal, 2023; Tolamatti et al., 2023). The subsequent four chapters will showcase how this thesis applies ML/DL techniques for the classification of AGN/blazars and the identification of rare sources.

3.6 Class-imbalance problem

ML classification methodologies assume a balanced distribution of classes (Weiss, 2004, 2005). However, this is not always the case in real-world scenarios. It is not uncommon to encounter classification problems where one class is significantly underrepresented compared to others, a situation known as the class-imbalance problem (He & Garcia, 2009; Ortigosa-Hernández et al., 2017). In binary classification tasks, the disparity between the two classes can reach magnitudes of 100:1, 1000:1, or even 10000:1 (Kubat et al., 1998; He & Shen, 2007; He & Garcia, 2009). In Chapter 7, we address a sample class imbalance where the positive class, neutrino blazars (NBs), consists of only a few dozen instances, whereas the negative class, non-NBs, comprises thousands. Notably, the class-imbalance issue can be further divided into relative imbalance/rarity and absolute imbalance/rarity (Weiss, 2004, 2005), indicating that the same proportion of positive to negative classes can have different implications depending on the absolute numbers involved.

It has been established that class imbalance adversely affects the training of traditional ML classifiers, including Multilayer Perceptrons and Convolutional Neural Networks (CNNs), leading to overfitting and results dominated by the majority class (Japkowicz & Stephen, 2002; Drummond & Holte,

2005; Mazurowski et al., 2008; Buda et al., 2018). Some studies have explored solutions to the class-imbalance problem for ANNs, such as developing novel loss functions for ANN training (Wang et al., 2016) and designing a new CNN approach that involves two-phase training of the network, initially on a class-balanced dataset followed by fine-tuning of the output layer (Havaei et al., 2017).

Traditional solutions to the class-imbalance problem can be summarized into five categories (Weiss, 2004, 2005):

1. Collecting more data is always beneficial but often impractical.
2. Changing the evaluation metrics:
 - Precision/recall (also known as specificity/sensitivity) calculated through the confusion matrix.
 - F1-score (a controllable weighted average of precision and recall).
 - ROC curves, etc.
3. Data-level methods, also known as resampling methods, focus on modifying the training dataset to enable standard learning algorithms to train effectively. These methods can be further classified into:
 - Methods that remove samples from the majority class, i.e., undersampling, such as NearMiss (Zhang & Mani, 2003), ENN (Wilson, 1972), Tomek link (Tomek, 1976).
 - Methods that generate new samples for the minority class, i.e., oversampling, such as SMOTE (Chawla et al., 2002), ADASYN (He et al., 2008), Borderline-SMOTE (Han et al., 2005).
 - Data Augmentation, which enhances the quantity and quality of original data without substantially increasing the data, approaching the value generated by more data, thereby improving the model's learning effect. This includes sample transformation-based data augmentation and deep learning-based methods such as Variational Auto-Encoding network or VAE (Cinelli et al., 2021) and Generative Adversarial Network or GAN (Goodfellow et al., 2014).
4. Algorithm-level methods focus on modifying existing standard machine learning algorithms to correct their preference for the majority class. A popular branch within this category is cost-sensitive learning (Freund & Schapire, 1997; Elkan, 2001), which assigns higher misclassification costs to minority class samples and lower costs to majority class samples.
5. Ensemble learning methods focus on combining a data-level or algorithm-level method with ensemble learning to obtain a powerful ensemble classifier. These methods are increasingly popular in practice due to their outstanding performance in class imbalance tasks.

This thesis addresses two class-imbalance problems that fall under the category of absolute imbalance, where the number of positive classes of interest is only in the decades, while the negative classes number in the hundreds or even thousands. The thesis adopts a conservative approach to data-level methods due to uncertainties regarding their impact on the original dataset distribution and to conserve computational resources. For the binary classification problem in Chapter 6, where the imbalance is

less severe, traditional ML algorithms such as Logistic Regression (LR) are employed, coupled with changing evaluation metrics and algorithm-level methods. For the binary classification in Chapter 7, where the degree of imbalance is greater, a more innovative ANN-based Transfer Learning (TL) method is used.

3.7 Sklearn and Pytorch

This thesis uses the machine learning library `Sklearn` developed based on the Python programming language and the DL framework `Pytorch` to complete data preprocessing and classification tasks.

Sklearn (Pedregosa et al., 2011a; Buitinck et al., 2013), also known as `scikit-learn`, is a widely acclaimed open-source machine learning library in Python. Initially developed by David Cournapeau in 2007 as part of the Google Summer of Code project, `Sklearn` is renowned for its simplicity and accessibility. It offers a comprehensive range of supervised and unsupervised learning algorithms, including tools for model fitting, data preprocessing, model selection, and evaluation. Built upon NumPy, SciPy, and matplotlib, `Sklearn` provides efficient tools for data mining and data analysis.

`PyTorch`, in contrast, is an open-source ML library based on the `Torch` library (Ketkar, 2017), extensively used for applications such as computer vision and natural language processing. Developed by Facebook's AI Research lab, `Pytorch` offers two high-level features: Tensor computation with strong GPU acceleration and Deep Neural Networks built on a tape-based autograd system. Known for its flexibility and ease of use in building and training neural networks, `Pytorch` has become a popular choice for researchers and developers in the field of DL.

Both `Sklearn` and `Pytorch` are integral components of the Python ecosystem for ML, each offering unique advantages. While `Sklearn` is often the preferred choice for traditional ML algorithms, `Pytorch` is favored for its dynamic computational graph and efficient memory usage, making it particularly suitable for DL applications. The selection between `Sklearn` and `Pytorch` typically depends on the specific requirements of the project, such as the complexity of the model, the size of the dataset, and the need for GPU acceleration.

Chapter 4

Radio Dualism in in AGN: A Detailed Study

As introduced in Chap. 1, AGN encompasses a variety of types, such as quasars, Seyfert galaxies, BL Lacertae objects, Fanaroff-Riley (FR) classifications, and low-ionization nuclear emission-line regions (LINERs). A key characteristic of AGNs is their division into two categories: radio-loud (RL) AGNs (including broad-line radio galaxies, blazars, and both FR-I and FR-II types) and radio-quiet (RQ) AGNs (such as Seyfert galaxies and LINERs). This categorization is based on numerous studies (Strittmatter et al., 1980; Kellermann et al., 1989; White et al., 2000; Baloković et al., 2012; Cao, 2016; Zhang et al., 2021).

Initially, Strittmatter et al. (1980) observed a bimodal distribution in the radio-to-optical flux density ratio (RL, $R = f_r/f_o$) among optically selected quasars, leading to the suggestion of two distinct quasar populations: RL and RQ. Kellermann et al. (1989) analyzed 114 bright quasars using 5 GHz and 4400 Å flux data, affirming the bimodal distribution of R and categorizing RQs and RLs based on specific R value ranges. Shastri et al. (1993) later introduced a specific $R = 10$ value to differentiate between RLs and RQs, using 5 GHz and optical B band data.

Subsequent research by various scholars (Ivezić et al., 2002; Zamfir et al., 2008; Zhang et al., 2021) has supported the notion of this dichotomy. However, White et al. (2000) compiled a comprehensive sample of 636 quasars and found a significant number of intermediate radio-loudness sources, challenging the bimodal distribution theory. This finding was corroborated by later studies (Lacy et al., 2001; Calderone et al., 2013).

The dichotomy in radio loudness implies different fundamental properties for RLs and RQs, such as variations in the origin of radio emission, black hole masses, accretion rates, and spins (McLure & Jarvis, 2004; Sikora et al., 2007; Laor & Behar, 2008).

In examining the AGN radio dichotomy, it's evident that the threshold value of R varies depending on the sample used. Zhang et al. (2021) utilized a dataset of 2419 sources from the thirteenth quasar and AGN catalog (Véron-Cetty & Véron, 2010), proposing a division at $\log R = 1.26$ to distinguish between RLs and RQs. This division was validated through statistical analysis, maximizing the separation accuracy in a diagram comparing radio luminosity against radio loudness.

This study reexamines the AGN radio dichotomy using the Véron-Cetty & Véron (2010) catalog, aiming to delve deeper into the underlying physical properties that differentiate these categories.

Throughout this paper, we have consistently used the following cosmological parameters: a Hubble constant (H_0) of 73 kilometers per second per megaparsec, a matter density parameter (Ω_m) of 0.3, and a dark energy density parameter (Ω_Λ) of 0.7.

4.1 Samples

In this research, a comprehensive dataset of 2943 astronomical objects, encompassing 2120 quasars, 245 BL Lacs, and 578 other types of AGNs, was assembled. This dataset, detailed in columns (4), (5), and (6) of Tab. 4.1, includes measurements such as radio flux density at 6 cm, apparent V band magnitudes, and $B - V$ color indices, sourced from the work of Véron-Cetty & Véron (2010).

To account for the reddening effect caused by interstellar dust and gas, which can alter the observed spectrum of celestial objects, the study incorporates a correction for Galactic extinction. This correction, based on the model by Schlegel et al. (1998), is denoted as E_{B-V} and is listed in column (7) of Tab. 4.1. The V band magnitude is adjusted using the formula $A_V = 2.742 \times E_{B-V}$, as suggested by Zhang et al. (2021). This adjustment allows for the conversion of the V band magnitude to the B band magnitude, using the known $B - V$ color, and subsequently to the corresponding flux densities f_V and f_B .

Furthermore, to derive the intrinsic intensity of the sources, a K -correction is applied to both the radio and optical fluxes. This correction, represented as $f'_\nu = f_\nu(1+z)^{\alpha_\nu-1}$, adjusts the observed flux density f_ν at frequency ν for the effects of redshift z . The study adopts values of α_ν as 0.76 ± 0.54 for the optical band and 0.25 ± 0.52 for the radio band, following the methodology of Zhang et al. (2021).

The concept of radio loudness is central to this study. It is defined as the ratio of the intrinsic radio flux density to the optical flux density, expressed as:

$$\log R = \log \frac{f'_r}{f'_o}, \quad (4.1)$$

where f'_r and f'_o are the intrinsic radio and optical flux densities, respectively. The study utilizes the radio flux density at 6 cm and the optical flux density in the B band, as recorded in Tab. 4.1.

Additionally, the study calculates the monochromatic luminosity of the sources using the equation:

$$L_\nu = 4\pi d_L^2 \nu f'_\nu, \quad (4.2)$$

where d_L represents the luminosity distance, calculated as $d_L = (1+z) \cdot \frac{c}{H_0} \cdot \int_1^{1+z} \frac{1}{\sqrt{\Omega_M x^3 + 1 - \Omega_M}} dx$. This formula incorporates key cosmological parameters and the redshift of the objects.

4.2 Results

4.2.1 Distribution of Source Numbers vs. Redshift

Our sample spans a redshift range from 0 to 5.11, excluding 86 BL Lacs without available redshifts and two radio galaxies (M31 and NGC3031) with zero redshifts. Fig. 4.1 illustrates the redshift bin count distribution ($bin_width = 0.1$) for all three source classes. Gaussian fits for quasar and BL Lac distributions yield $\mu_Q = 1.24$, $\sigma_Q = 0.07$ and $\mu_A = 0.25$, $\sigma_A = 0.26$, respectively. However, Gaussian fitting is not applicable for the other AGN distribution, as these AGNs predominantly lie within a redshift range of 0 to 0.1, as indicated by the 'green square' in the upper-left corner of Fig. 4.1.

4.2.2 Radio Loudness

Fig. 4.2 displays the radio loudness ($\log R$) distribution of our sample. The $\log R$ spans over six orders of magnitude, with quasars ranging from -0.9 to 5.271 ($\langle \log R_Q \rangle = 2.81 \pm 0.72$), BL Lacs from 1.131

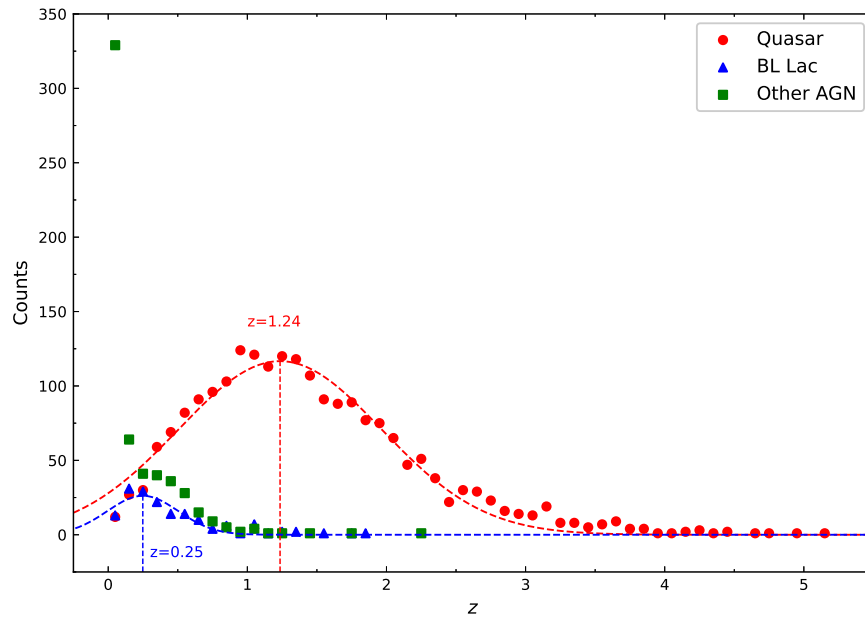


Figure 4.1: Redshift distribution of source numbers. Quasars (red dots), BL Lacs (blue triangles), and other AGNs (green squares) are shown. Dashed curves represent Gaussian fits.

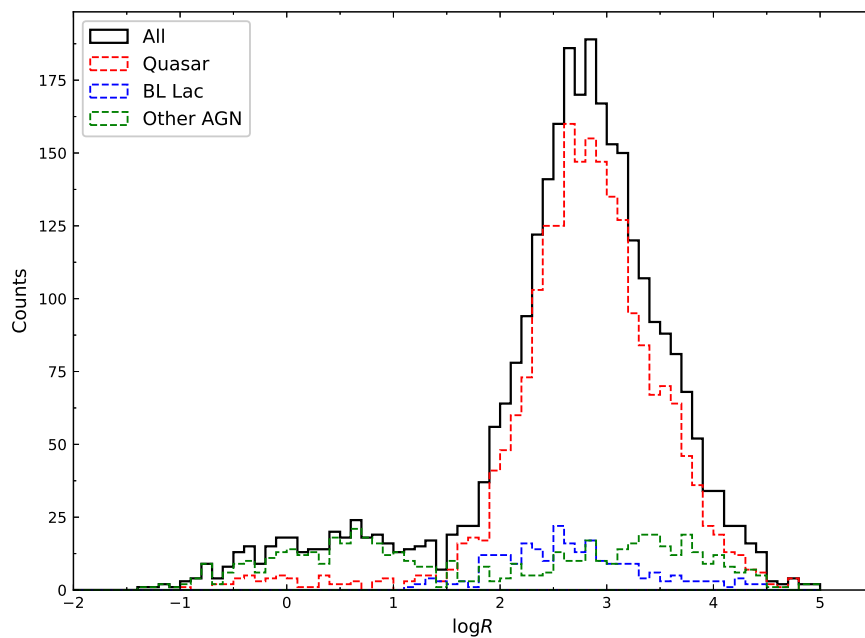


Figure 4.2: Radio loudness distribution. Quasars (red), BL Lacs (blue), and other AGNs (green) are represented.

to 4.473 ($\langle \log R_B = 2.66 \pm 0.68 \rangle$), and other AGNs from -2.329 to 5.712 ($\langle \log R_A = 1.89 \pm 1.62 \rangle$). Other AGNs exhibit the broadest range and lowest $\log R$ values, while quasars have the highest.

A clear dichotomy in the distribution is observed, with the lower $\log R$ bump primarily composed of other AGNs and some quasars, and the higher bump mostly consisting of quasars. Scikit-learn (also known as `sklearn`), an ML library for Python, offers a variety of algorithms for classification, regression, and clustering. In this study, we use the GMM introduced in Chap. 3 to analyze the $\log R$ distribution, as illustrated in Fig. 4.3. GMM is a probabilistic model that posits that all data points are generated from a finite number of Gaussian distributions with unknown parameters. Based on prior research (Shastri et al., 1993; White et al., 2000; Zhang et al., 2021), we postulate the presence of two Gaussian components in the $\log R$ distribution.

The GMM procedure involves determining the proportions of each Gaussian component within the overall distribution, as well as their respective means and standard deviations. This is achieved through an EM algorithm (also see Chap. 3), an iterative process comprising two steps: 'E' and 'M'. In the 'E' step, we use initial or previously iterated parameter values to compute the posterior probability of hidden variables using the Bayesian formula. In the 'M' step, we maximize the likelihood function to estimate the means and standard deviations. These steps are repeated until the convergence of the estimated parameters.

To enhance our model's generalizability, we perform stratified sampling at an 80% scale and repeat the entire GMM process 1,000 times. The outcome reveals a dichotomy represented by two Gaussian components marked as Gaussian[0] and Gaussian[1], with weights of 11.7% for Gaussian[0] and 88.3% for Gaussian[1]. The means and standard deviations of these components are $\mu[0] = 0.42 \pm 0.03$, $\sigma[0] = 0.54 \pm 0.03$ for Gaussian[0], and $\mu[1] = 2.91 \pm 0.01$, $\sigma[1] = 0.39 \pm 0.01$ for Gaussian[1]. Besides, the average log-likelihood of the GMM under the $\log R$ is -1.28. The intersection of these curves at $\log R = 1.37 \pm 0.02$ is used as a threshold to categorize radio sources into two groups: RL for $\log R \geq 1.37$ and RQ for $\log R < 1.37$. Following this new classification, our sample includes 2057 quasars, 238 BL Lacs, and 314 other AGNs identified as RLs, constituting 97.0%, 97.1%, and 54.3% of their respective totals.

4.2.3 Examining the Relationship Between Radio and Optical Luminosity

The observed dual-mode distribution of RL and RQ sources in terms of their $\log R$ values suggests potential disparities in the relationship between radio and optical luminosities across these two radio categories.

In our study, we examined the radio luminosity at 6 cm (denoted as $\log_{6\text{cm}}$) and optical luminosity in the B band (represented as \log_B), employing linear regression analysis. For Radio-Loud sources, the relationship is established as follows:

$$\log L_{6\text{cm}} = (1.09 \pm 0.01) \log L_B - (6.40 \pm 0.66), \quad (4.3)$$

This is accompanied by a Pearson partial correlation coefficient of $r = 0.95$ and a remarkably low chance probability of $p < 10^{-10}$, after accounting for the uncertainty of data and redshift effects. In the computation of $\log B$ and $\log L_{6\text{cm}}$, we employ the error propagation formula $\Delta f = \sqrt{\sum_{i=1}^n \left(\frac{\partial f}{\partial x_i} \Delta x_i \right)^2}$ to transfer the uncertainties associated with the calculation of flux densities to the optical and radio luminosities, where Δf and Δx_i stands for the uncertainty of dependent variables and independent variables. The uncertainties in the optical and radio luminosities are denoted as Err_B and $Err_{6\text{cm}}$, respectively. The reciprocal of the total error, given by $\frac{1}{\sqrt{Err_{6\text{cm}}^2 + Err_B^2}}$ is utilized as the weight for each data point in the calculation of the correlation coefficients r and p .

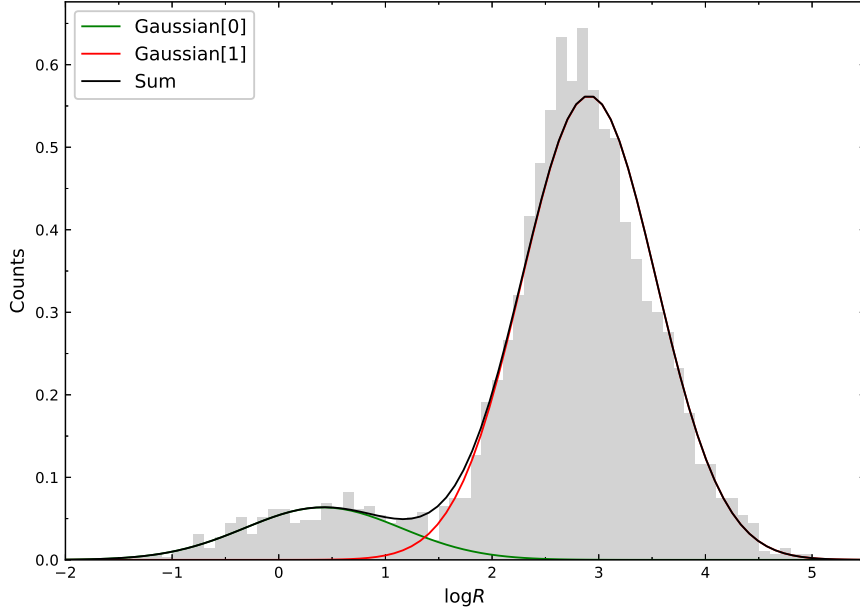


Figure 4.3: Density probability distribution of radio loudness fitted with two Gaussian components.

In contrast, for Radio-Quiet sources, the regression equation is:

$$\log L_{6\text{cm}} = (1.19 \pm 0.03)\log_B - (13.37 \pm 1.47), \quad (4.4)$$

Here, we observe a correlation coefficient $r = 0.72$ and a chance probability of $p < 10^{-10}$, once again after correcting for data error and redshift influences, as elucidated through Pearson partial analysis. These findings are graphically represented in Fig. 4.4.

4.3 Discussion

4.3.1 Comparison of Source Distributions

Our study adopts a more flexible approach in source selection compared to the methodology used by Zhang et al. (2021), resulting in an additional 524 sources. These extra sources comprise 232 quasars, 97 BL Lacs, and 196 other AGNs. We utilized the Kolmogorov–Smirnov (K-S) test to examine potential discrepancies in redshift, $f_{6\text{cm}}$.

The distribution patterns of three distinct classes of astronomical sources are illustrated in Fig 4.1. Quasars are observed to have a notably higher mean redshift compared to BL Lacs and other AGNs, which primarily include Seyfert galaxies and LINERs. The distribution of BL Lacs and other AGNs largely falls within the redshift range of $0 < z < 1$, with BL Lacs peaking at $z = 0.25$ and the majority of other AGNs clustering in the $0 < z < 0.1$ redshift bin. This pattern aligns with observations indicating that BL Lacs, Seyfert galaxies, and LINERs are typically found at lower redshifts, a trend attributed to their relatively lower optical luminosity compared to quasars.

Our sample, however, is not exhaustive, due to two primary limitations. Firstly, RLs are estimated to constitute about 10% of all AGNs (Kellermann et al., 2016; Panessa et al., 2019), suggesting a larger prevalence of RQs. However, our sample predominantly contains RL AGNs. Secondly, the redshift

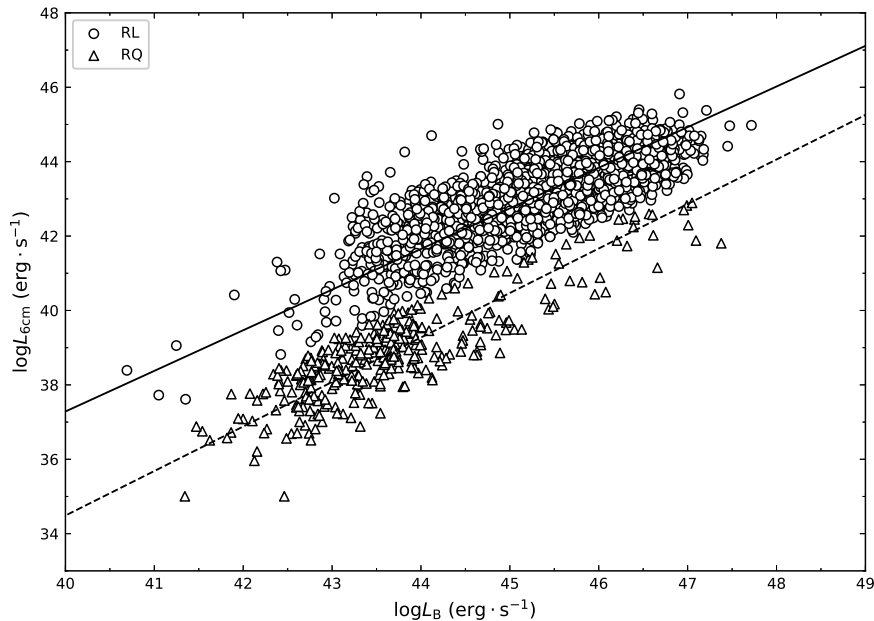


Figure 4.4: Radio luminosity versus B-band luminosity. RLs (circles and solid lines) and RQs (triangles and dashed lines) are shown.

distribution of quasars, as reported in studies by the Sloan Digital Sky Survey (SDSS), indicates a range from 0 to 3.5 with multiple peaks (Pâris et al., 2018). In contrast, our sample exhibits a singular peak at $z = 1.24$ for quasars.

The limitations in our sample stem from our selection criteria, which focused on sources with available data in both the B band (including V band and the color index $B-V$) and at a 6cm wavelength. While sources are abundant with B band data, 6 cm radio observations are less commonly available, particularly for quasars. Quasars are known for stronger radio emissions in comparison to the other two classes (Kellermann et al., 2016; Radcliffe et al., 2021), making them more likely to be detected due to the lower sensitivity requirements of radio telescopes. It is important to consider that such sample incompleteness may impact the results concerning radio dichotomy. The primary effects on our findings could be twofold: (1) The boundary for $\log R$ in our study might be underestimated due to sample incompleteness, as the RQ peak, expected to be more dominant and widespread in the bimodal $\log R$ distribution, is not fully represented (see Fig. 4.2). (2) The correlation between $\log L_{6\text{cm}}$ and $\log L_B$ might differ significantly in a more complete sample. With a higher proportion of RQs, typically found in the lower-left region of Fig. 4.4, the sample would include various AGN types, each potentially exhibiting different radio-optical statistical relationships.

Despite these limitations, our study still encompasses the largest sample for analyzing $\log R$ bimodal distributions to date. Our findings on $\log R$ are in concordance with previous studies, as will be discussed in the following section. This consistency lends credibility to our research on the topic of radio dichotomy.

4.3.2 Revisiting the Radio Dichotomy in Quasars: A Comprehensive Analysis

The phenomenon of radio dichotomy in quasars has been a subject of debate for over four decades, with studies yielding varying conclusions (Strittmatter et al., 1980; Kellermann et al., 1989; White et al., 2000; Baloković et al., 2012). A major factor contributing to these inconsistent findings is the limited sample size in earlier research. This issue was addressed in a study by Zhang et al. (2021), which utilized a larger sample of 2419 radio sources to affirm the existence of radio dichotomy. They proposed a demarcation at $reqr_{resultsRQR} = 1.26$ to distinguish between RLs and RQs quasars. Our analysis, however, suggests a slightly higher boundary of $\log R = \langle 1.37 \pm 0.02 \rangle$, deviating from Zhang's result. This discrepancy can be attributed to the variations in sample sizes. The traditional boundary of $\log R = 1$, as suggested by Shastri et al. (1993) through an analysis of 45 quasars, is notably lower than both our findings and those of Zhang et al. (2021). An increasing trend in $\log R$ values with larger sample sizes is observed. It is important to note that RLs exhibit radio intensities several times stronger than those of RQs.

Multiple hypotheses have been proposed to explain the radio dichotomy (Kellermann et al., 1989, 2016). These include:

1. **Intermittent Radio Activity:** This theory posits that quasars are typically radio-silent, becoming radio-emitters only during periods of unusual activity. However, this contrasts with the observed variability time scales, which are typically months to years, not the 10^5 to 10^6 years required for this hypothesis (Barvainis et al., 2005).
2. **Absorption Mechanisms:** The role of absorption by intervening plasma and synchrotron self-absorption has been considered. Nevertheless, radio observations at lower bands, less affected by opacity, do not significantly alter the RLs to RQs ratio (Strittmatter et al., 1980; Condon et al., 1980; Ennis et al., 1982; Robson et al., 1985).
3. **Geometric Effects:** The possibility of relativistic beaming in a subset of quasars, where the jet is aligned towards the observer, has been explored (Urry & Padovani, 1995; Fan et al., 2004, 2014; Xiao et al., 2015; Pei et al., 2016). While this model explains superluminal motions observed in quasars (Kellermann et al., 2004; Xiao et al., 2019), the differential distribution of 6 cm flux density in a sample of 75 sources contradicts this as a sole explanation for the dichotomy (Kellermann et al., 1989; Scheuer & Readhead, 1979).
4. **Host Galaxy Influence:** Investigations into quasar host galaxies indicate that radio emission at lower luminosities ($10^{21} \lesssim L_{6\text{cm}} \lesssim 10^{23}$ W/Hz) may primarily result from star formation within the host galaxy (Kellermann et al., 2016). However, these conclusions are based on a limited sample size and a narrow redshift range ($0.2 < z < 0.3$).
5. **Stellar or Black Hole Properties:** Factors such as black hole mass, accretion rate, outflow, and spin have been implicated in discussions of radio dichotomy (Wilson & Colbert, 1995; Dunlop et al., 2003; Ghisellini et al., 2004; Sikora et al., 2007; Gopal-Krishna et al., 2008; Best & Heckman, 2012). For instance, Laor & Behar (2008) suggested a correlation between radio and X-ray flux density, indicative of coronal activity in RQs. Cao (2016) proposed that the dichotomy primarily stems from the angular velocity of circumnuclear gas in AGNs.

Despite these diverse interpretations, a consensus remains elusive. Our findings highlight a distinct separation between RL and RQ clusters at a luminosity of $L_{6\text{cm}} \sim 10^{40}$ erg \cdot s $^{-1}$ (Stocke et al., 1992), suggesting a significant disparity in radio and optical power dependencies between the two

populations. Regression analysis reveals strong associations between $\log L_{6\text{cm}}$ and $\log L_B$ for both RLs and RQs, with nearly identical slope values. This slope similarity indicates a potential commonality in the radio-optical correlation across both categories. Considering that optical emission in both RLs and RQs is predominantly driven by thermal radiation from accretion disks, it is plausible that their radio emissions share a common origin. For RLs, this is widely accepted to be synchrotron emission from relativistic electrons in jets. Consequently, we propose that RQs may exhibit a form of 'mini-jet' or uncollimated outflow, similar to the 'aborted jet' concept, which involves the intermittent ejection of material at sub-escape velocities (Ghisellini et al., 2004). Collisions between infalling and outgoing matter could generate high-energy emissions, such as X-rays.

4.3.3 Refined Classification of Radio Loud and Radio Quiet Sources

Previous research, including our study, confirms that RL and RQ in celestial sources can be distinctly classified based on their radio loudness ($\log R$). Nonetheless, relying solely on this criterion ($\log R$) leads to two potential misclassifications: (1) sources with high radio luminosity ($L_r > 10^{43}$ erg/s) but even higher optical luminosity, resulting in $\log R \lesssim 1.37$, are incorrectly categorized as RQ; (2) sources with low radio luminosity ($L_r < 10^{41}$ erg/s) coupled with lower optical luminosity, yielding a $\log R > 1.37$, are misclassified as RL. These classifications are misleading as it's implausible for a source with $L_r > 10^{43}$ erg/s to be RQ and one with $L_r < 10^{41}$ erg/s to be RL.

To address these inaccuracies, we propose an additional criterion: radio luminosity. As depicted in Fig. 4.5, the diagram of radio luminosity at 6cm ($\log L_{6\text{cm}}$) versus radio loudness ($\log R$) delineates RLs in the upper-right and RQs in the lower-left. We aim to establish a boundary in this two-dimensional parameter space, using radio luminosity as a function of radio loudness, to separate RLs from RQs accurately.

For this purpose, we employ SVM from *sklearn*, a machine learning method well-suited for classification and regression challenges. SVM functions by identifying a hyperplane in an N-dimensional space that maximizes the margin between two linearly separable sample sets. If the samples are non-linearly separable, SVM maps them into a higher-dimensional space to locate the optimal hyperplane. Our approach involves dividing our RL and RQ samples into a training set (70%) and a test set (30%), utilizing the training set to determine the optimal hyperplane and the test set to assess its classification accuracy.

In our two-dimensional space, represented by $\log R$ and $\log L_R$, we hypothesized the hyperplane as a line, expressed as $w_1 \log R + w_2 \log L_R + m = 0$. The parameters w_1 , w_2 , and m were deduced through SVM training. We employed the *svm.LinearSVR* algorithm from *sklearn* for this task, iterating different combinations of hyperparameters until convergence to the optimal values was achieved.

Our analysis yielded an optimal dividing line with an impressive classification accuracy of 99.0%:

$$\log L_{6\text{cm}} = -2.7 \log R + 44.3. \quad (4.5)$$

This line effectively segregates RLs (upper right region) from RQs (lower left region). This finding is in contrast with the steep dividing line proposed by Zhang et al. (2021), who suggested a boundary of $\log R = 1.26$ for RL and RQ separation through analytical methods. Our approach, utilizing an SVM methodology, results in a considerably flatter slope (-2.7) for the dividing line.

Therefore, we advocate for a dual-criterion approach, incorporating both radio loudness ($\log R = 1.37$) and the SVM-derived dividing line ($\log L_{6\text{cm}} = -2.7 \log R + 44.3$), to more accurately classify radio sources as either RLs or RQs.

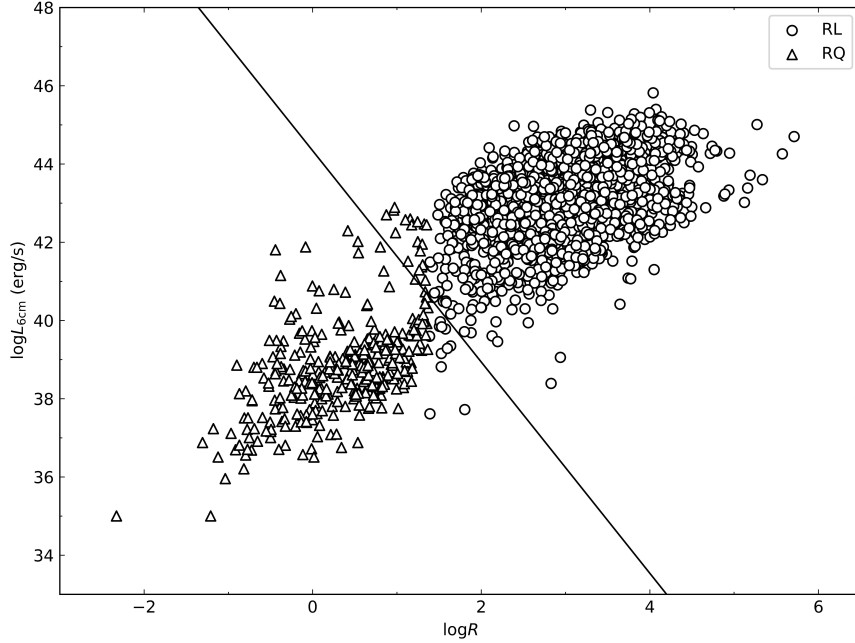


Figure 4.5: Construct a diagram illustrating the relationship between radio luminosity and radio loudness. In this diagram, represent RLs with circles and RQs with empty triangles. Include a black line in the diagram to demarcate the dividing line between these two categories.

4.4 Summary

In this study, we have curated an extensive AGN dataset, comprising 2943 sources, to investigate the bimodal distributions of $\log R$ and the dichotomy in radio emissions.

Utilizing SVM analysis, we scrutinized the $\log R$ distribution, successfully decomposing it into dual Gaussian components using the Gaussian Mixture Model (GMM). We determined the intersection point of these Gaussian curves at $\log R = \langle 1.37 \pm 0.02 \rangle$, proposing this as a demarcation for separating RL and RQ sources. Recognizing the potential for misclassification with a single criterion for RL/RQ differentiation, we introduced radio luminosity as a pivotal factor in this segregation. Consequently, we recommend a dual-criterion, defined as $\log R = 1.37$ and $\log L_{6\text{cm}} = -2.7 \log R + 44.3$, for more accurate classification of radio sources into RLs or RQs.

Furthermore, our analysis reveals moderate correlations between $\log L_{6\text{cm}}$ and $\log L_B$ for both RLs and RQs, with the correlations appearing consistent across both categories. This suggests a common origin for radio emissions in both RLs and RQs, potentially attributable to jets and mini-jets, including aborted jet or outflow mechanisms.

Table 4.1: The radio sources of our sample.*

Name (1)	Class (2)	Redshift (3)	f_{6cm} (4)	V_{mag} (5)	B-V (6)	E(B-V) (7)	Radio Class (8)
SDSS J00037-1108	Q	1.57	0.118	19.93	0.33	0.0364	RL
FIRST J00051-1010	Q	1.3	0.043	19.21	0.42	0.0377	RL
UM 18	Q	1.9	0.257	16.21	0.35	0.0304	RL
BG CFH 17	Q	1.62	0.056	19.31	0.24	0.0361	RL
PKS 0003+15	Q	0.45	0.34	16.4	0.11	0.0491	RL
...
...
...

The full details of this table can be found in the comprehensive listing provided in Xiao et al. (2022c), where all items are thoroughly documented.

Chapter 5

The Jet Apparent Motion and Central Engine Dynamics in *Fermi* Blazars

5.1 Introduction

The unique and striking observational characteristics of blazars, such as their rapid and extensive variability, pronounced and fluctuating polarization, intense and irregular γ -ray emissions, and apparent superluminal motion, have been extensively studied and documented in various scholarly works (Wills et al., 1992; Urry & Padovani, 1995; Fan, 2002; Villata et al., 2006; Fan et al., 2014; Xiao et al., 2015; Gupta et al., 2016; Xiao et al., 2019, 2020; Abdollahi et al., 2020; Fan et al., 2021b). These extreme observational features are largely attributed to the presence of a relativistic jet, as proposed by Blandford & Koenigl (1979), which is oriented towards the observer at a narrow-angle (ϕ), thereby creating a Doppler beaming effect (Ghisellini et al., 1993; Fan et al., 2013; Pei et al., 2016; Xiao et al., 2020). The Doppler factor ($\delta = [\Gamma(1 - \beta\cos\phi)]^{-1}$, where Γ is the bulk Lorentz factor and β is the jet speed in units of the speed of light, c) is a crucial parameter in jet analysis as it quantifies the extent of flux amplification and timescale compression as observed from Earth. However, the determination of δ is challenging as it relies on indirect measurements due to the unobservable nature of both β and ϕ . Various indicators of the beaming effect, such as the core-dominance parameter (Pei et al., 2016, 2020) and the apparent superluminal motion of knots (Zhang & Fan, 2008; Xiao et al., 2019, 2020), have been introduced in the literature to address this challenge.

The study of blazar jets, particularly their knots, has been a focus of Very Long Baseline Interferometry (VLBI) for many years. These knots, particularly those exhibiting superluminal motion, are observable as localized intensity enhancements in radio images of jets. The heightened flux corresponds to a significant brightness temperature, calculated as $T = \frac{2c}{\pi k} \frac{S}{\theta_{\text{eq}}^2 \nu^2}$, where S represents the flux at frequency ν , and θ is the angular diameter of a source (Readhead, 1994). Observations often record brightness temperatures far exceeding the jet equipartition threshold of approximately 5×10^{10} K, where the powers of radiating particles and the magnetic field are comparable. Nonetheless, the brightness temperature is also constrained by the inverse Compton cooling effect, limiting it to a maximum of 1.0×10^{12} K (Kellermann & Pauliny-Toth, 1969). While various authors have reported differing maximum brightness temperatures in the rest frame (Readhead, 1994; Liodakis et al., 2018), early VLBI observations generally align with the theoretical brightness temperature limit set by inverse Compton cooling. However, Kellermann (2003) proposed that this agreement might be coincidental, influenced by Earth's size and the range of observed flux densities.

In addition to brightness temperature studies, the dynamics of knots also present compelling re-

search avenues. The first observation of superluminal motion in blazar 3C 279 (Cohen et al., 1971) corroborated the theoretical proposal of Rees (1966). Subsequent studies, such as those by Cohen et al. (2007), have shown that the distribution of observed apparent velocity and luminosity is in line with relativistic beaming models. This concept has gained wide acceptance in the research community. The apparent velocity of blazar jet knots (β_{app}) is often used to estimate the Doppler factor and viewing angle (Xiao et al., 2019, 2020), as both superluminal motion and Doppler boosting arise from the same geometric orientation of the jet. Notably, Jorstad et al. (2001) and Kellermann et al. (2004) observed that AGNs with γ -ray emissions, as detected by EGRET and later by *Fermi*, tend to exhibit higher apparent velocities. Kellermann et al. (2007) also observed a correlation between luminosity and apparent velocity, a finding supported by subsequent studies (Lister et al., 2009; Piner et al., 2012; Xiao et al., 2019).

The relationship between blazar knots and γ -ray emissions was initially explored two decades ago (Jorstad et al., 2001) and has been further affirmed in recent studies (Jorstad & Marscher, 2016; Jorstad et al., 2017; Weaver et al., 2022) by the Boston University γ -ray Blazar Monitoring Program and the MOJAVE project. These initiatives have significantly contributed to our understanding of radio brightness and polarization variations in AGN jets. The launch of the *Fermi* Large Area Telescope (*Fermi*-LAT) has heralded a new era in blazar studies, with the release of multiple source catalogs, including the fourth *Fermi*-LAT source catalog (4FGL_DR2, Abdollahi et al., 2020), which lists 3511 blazars (4LAC_DR2, Ajello et al., 2020). Xiao et al. (2019, 2020) utilized MOJAVE and *Fermi*-LAT data to compare blazars with and without *Fermi* γ -ray detection, revealing that those detected by *Fermi* exhibit larger apparent velocities, indicative of a stronger beaming effect. Lister et al. (2019) further discovered a robust correlation between apparent jet velocity and synchrotron peak frequency, suggesting that *Fermi*-LAT γ -ray associated AGNs in the MOJAVE sample tend to exhibit more variable position angles than non-*Fermi* AGNs.

Despite extensive research, the direct connection between knot dynamics and the central engine of blazars, as well as the formation and acceleration mechanisms of knots, remain topics of active investigation. This study aims to delve deeper into the relationship between the apparent motion of knots and the properties of the jet and accretion disk. Additionally, we seek to identify potential candidates for future VLBI observations, thereby contributing to the ongoing exploration of these enigmatic celestial phenomena.

5.2 Data Collection Methodology

In this study, we focus on investigating the properties of VLBI-detected blazars, specifically examining the relationship between their apparent motion, jets, and accretion disks. To this end, we have curated a dataset of blazars, each with documented proper motion (μ) or apparent velocity (β_{app}), sourced from a range of established studies (Vermeulen et al., 1994; Jorstad et al., 2005; Britzen et al., 2008; Piner et al., 2007; Piner & Edwards, 2018; Lister et al., 2019) as detailed in Col. (4), (5), and (6) of Tab. 5.1, where Col. (1) 4FGL name; Col. (2) B1959 Name; Col. (3) redshift; Col. (4) feature identification number; Col. (5) angular proper motion measured in microarcseconds per year; Col. (6) reference of proper motion; Col. (7) frequency; Col. (8) radio flux density Measured at the frequency specified in Col. (7); Col. (9) reference of radio flux density; Col. (10) the dominance of the core at a frequency of 5 GHz; Col. (11) the integral photon flux (1-100 GeV); Col. (12) the photon spectral index; Col. (13) classification of blazars, ‘F’ represents FSRQs, ‘B’ stands for BL Lacs, and ‘U’ is used to denote Blazar Candidates of Uncertain type (BCUs); Col. (14) the variability index; Col. (15) mass of SMBH;

Col. (16) luminosity of accretion disk; Col. (17) flux of IC peak; Col. (18) references to Col. (15), (16), and (17). C99: Cao & Jiang (1999), C12: Chai et al. (2012) L18: Lister et al. (2018), L19: Lister et al. (2019), P21: Paliya et al. (2021), P14: Piner & Edwards (2014), P18: Piner & Edwards (2018), Sh12: Shaw et al. (2012), T96: Taylor et al. (1996), X91: Xie et al. (1991), Z12: Zhang et al. (2012b). Our data collection adhered to two primary criteria:

1. For sources with multiple VLBI-detected components, the highest recorded apparent velocity, denoted as $\beta_{\text{app}}^{\text{max}}$, was selected.
2. We prioritized sources from the MOJAVE program (Lister et al., 2019), given its status as an ongoing monitoring initiative utilizing the VLBA, the world's most advanced resolution telescope. This preference led us to exclude duplicative sources from other literature.

We further augmented our dataset by cross-referencing these sources with the 4LAC_DR2 catalog to incorporate spectral and variability data, as indicated in Col. (11) to (14) of Tab. 5.1. The final dataset encompasses 407 *Fermi* blazars with available μ^{max} , including 372 blazars with either documented or calculable $\beta_{\text{app}}^{\text{max}}$ values, derived using the equation:

$$\beta_{\text{app}} = \frac{\mu}{H_0} \int_1^{1+z} \frac{1}{\sqrt{\Omega_M x^3 + 1 - \Omega_M}} dx \quad (5.1)$$

(Vermeulen et al., 1994; Zhang & Fan, 2008; Xiao et al., 2019).

Additionally, we collated radio flux density (S_ν) data from various studies (Col. (7) to (9) of Tab. 5.1) and 5 GHz core dominance parameter ($\log R$, Col. (10) of Tab. 5.1) from Pei et al. (2020). Data pertaining to black hole mass (M_{BH} , Col. (15) of Tab. 5.1) and accretion disk luminosity ($\log L_{\text{Disk}}$, Col. (16) of Tab. 5.1) were also compiled from existing literature. Lastly, the intensity of the inverse Compton peak ($\log F_{\text{IC}}$, Col. (17) of Tab. 5.1) information was sourced from Paliya et al. (2021).

For convenience, we refer to these VLBI-detected *Fermi* blazars as VFBs throughout our study.

5.3 Results

5.3.1 Association Between Maximum Apparent Velocity and Luminosity, Photon Index, and Variability Index in *Fermi* GeV γ -ray Emission

We propose a model where γ -ray photons adhere to a power-law distribution, represented as

$$\frac{dN}{dE} = N_0 E^{-\alpha_{\text{ph}}}, \quad (5.2)$$

where α_{ph} denotes the photon spectral index. The coefficient N_0 is defined as $N_0 = N_{(E_L \sim E_U)} \left(\frac{1}{E_L} - \frac{1}{E_U} \right)$, for $\alpha_{\text{ph}} = 2$, and alternatively, as $N_0 = \frac{N_{(E_L \sim E_U)} (1 - \alpha_{\text{ph}})}{(E_U^{1-\alpha_{\text{ph}}} - E_L^{1-\alpha_{\text{ph}}})}$, for other values of α_{ph} . Here, $N_{(E_L \sim E_U)}$ signifies the integrated photon count in the range of 1 GeV to 100 GeV, expressed in photons $\cdot \text{cm}^{-2} \cdot \text{s}^{-1}$.

The integral flux F , in $\text{GeV} \cdot \text{cm}^{-2} \cdot \text{s}^{-1}$, is given by

$$F = N_{(E_L \sim E_U)} \frac{E_U - E_L}{E_U \times E_L} \ln \frac{E_U}{E_L} \quad (5.3)$$

for $\alpha_{\text{ph}} = 2$, and by

$$F = N_{(E_L \sim E_U)} \frac{1 - \alpha_{\text{ph}}}{2 - \alpha_{\text{ph}}} \frac{(E_U^{2-\alpha_{\text{ph}}} - E_L^{2-\alpha_{\text{ph}}})}{(E_U^{1-\alpha_{\text{ph}}} - E_L^{1-\alpha_{\text{ph}}})} \quad (5.4)$$

Table 5.1: The Selected Ensemble of Blazars from the *Fermi* Catalog

4FGL name (1)	B1950 name (2)	z (3)	ID (4)	μ_{max} $\mu\text{as/yr}$ (5)	Ref (6)	ν GHz (7)	S_{ref} mJy (8)	Ref (9)	$\log R$ (10)	Flux1000 $\text{ph cm}^{-2} \text{s}^{-1}$ (11)	α_{ph} (12)	Class (13)	VJ (14)	$\log(M_{\text{BH}}/M_{\odot})$ (15)	$\log L_{\text{disk}}$ erg s^{-1} (16)	$\log F_{\text{IC}}$ $\text{erg cm}^{-2} \text{s}^{-1}$ (17)	Ref (18)
J0005.9+3824	0003+380	0.229	2	317 ± 25	L19	15	600.2	L18	1.13	$4.25\text{E-}10$	2.67	F	26.77	8.95 ± 0.4	44.52 ± 0.15	-12.04	P21
J0006.3-0620	0003-066	0.347	8a	330.4 ± 9.7	L19	15	2351.9	L18	0.26	$1.40\text{E-}10$	2.13	B	20.41	7.47 ± 0.26	44.32 ± 0.13	-11.65	P21
J0009.1+0628	0006+061	1.563	1	134 ± 11.4	L19	15	185.4	L18	0	$4.77\text{E-}10$	2.10	B	13.05
J0014.1+1910	0011+189	0.477	2	159 ± 16	L19	15	133.1	L18	0	$2.25\text{E-}10$	2.28	B	39.85
J0014.8+6118	0012+610	...	1	13 ± 10	L19	$7.91\text{E-}10$	2.05	U	6.15
...	U

The full details of this table can be found in the comprehensive listing provided in Xiao et al. (2023d), where all items are thoroughly documented.

otherwise. The γ -ray luminosity L_γ is calculated as

$$L_\gamma = 4\pi d_L^2 (1+z)^{(\alpha_{\text{ph}}-2)} F, \quad (5.5)$$

where d_L represents the luminosity distance, defined in (Komatsu et al., 2011), and $(1+z)^{(\alpha_{\text{ph}}-2)}$ accounts for the K -correction.

Our analysis involves calculating γ -ray luminosity for various blazar subclasses and exploring its relationship with apparent velocity. The upper panel of Fig. 5.1 illustrates the correlation between $\log L_\gamma$ and $\log \beta_{\text{app}}^{\text{max}}$, yielding a linear regression

$$\log L_\gamma = (1.68 \pm 0.08) \log \beta_{\text{app}}^{\text{max}} + (44.91 \pm 0.06),$$

with a correlation coefficient $r = 0.27$ and a chance probability $p = 7.1 \times 10^{-8}$, determined through a Pearson analysis.

The middle panel of Fig. 5.1 presents the relationship between α_{ph} and $\log \beta_{\text{app}}^{\text{max}}$, resulting in the linear regression

$$\alpha_{\text{ph}} = (0.15 \pm 0.02) \log \beta_{\text{app}}^{\text{max}} + (2.22 \pm 0.02),$$

with a correlation coefficient $r = 0.33$ and a chance probability $p = 3.1 \times 10^{-11}$.

Furthermore, the variability index (VI), an indicator of variability level as defined in (Nolan et al., 2012b; Acero et al., 2015), is examined. The *Fermi Collaboration* introduced a correction term in the variability index formula (Eq. 5.6) to reconcile the weights in full analysis and yearly intervals. This index is calculated as

$$VI = 2 \sum_i \log \left[\frac{\mathcal{L}_i(F_i)}{\mathcal{L}_i(F_{\text{glob}})} \right] - \max(\chi^2(F_{\text{glob}}) - \chi^2(F_{\text{av}}), 0), \quad (5.6)$$

$$\chi^2(F) = \sum_i \frac{(F_i - F)^2}{\sigma_i^2}, \quad (5.7)$$

where F_{av} is the average flux over one-year intervals, and F_{glob} is the flux from a global fit. A source is deemed variably significant if $VI > 18.48$, corresponding to a 99% confidence level in a χ^2 distribution (Abdollahi et al., 2020). The bottom panel of Figure 5.1 demonstrates the correlation between $\log VI$ and $\log \beta_{\text{app}}^{\text{max}}$, leading to a linear regression

$$\log VI = (0.32 \pm 0.07) \log \beta_{\text{app}}^{\text{max}} + (1.94 \pm 0.07),$$

with a correlation coefficient $r = 0.21$ and a chance probability $p = 2.7 \times 10^{-5}$.

5.3.2 Correlation Analysis: Maximum Apparent Velocity, Black Hole Mass, Accretion Disk Luminosity, and Normalized Accretion Luminosity

Several methodologies exist for estimating the mass of black holes (M_{BH}). A prevalent approach involves applying the virial theorem. This method approximates the mass of the central black hole via the equation:

$$M_{\text{BH}} \approx \frac{rv^2}{G}, \quad (5.8)$$

where v represents the velocity dispersion of matter at a distance r . The velocity dispersion is ascertainable through spectroscopic analyses, denoted as $v = f v_{\text{FWHM}}$, with v_{FWHM} being the full

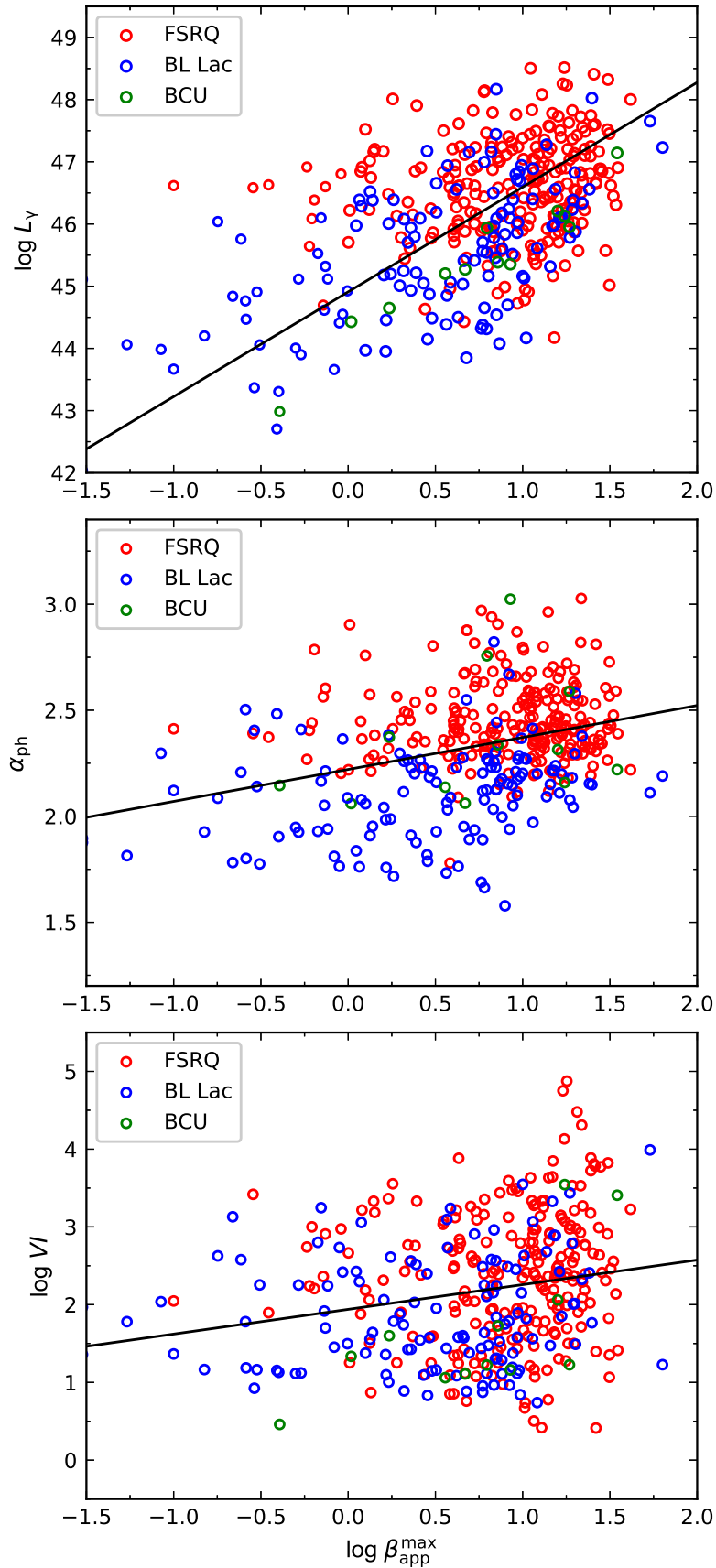


Figure 5.1: the correlations of $\log L_{\gamma}$, α_{ph} , and $\log VI$ with $\log \beta_{\text{app}}^{\text{max}}$ across different subclasses of blazars. Here, FSRQs are represented by red markers, BL Lacs by blue, and BCUs by green.

width at half maximum of the broad emission line and f a factor contingent upon geometry and kinematics, typically near unity (Peterson et al., 1999, 2000; McLure & Dunlop, 2001; Vestergaard, 2002). Reverberation mapping reliably determines the distance r , contributing to the ‘Size-Luminosity Relation’ (Kaspi et al., 2000b). Consequently, the virial M_{BH} is calculable using:

$$\log \left(\frac{M_{\text{BH}}}{M_{\odot}} \right) = a + b \log \left(\frac{\lambda L_{\lambda}}{10^{44} \text{ erg} \cdot \text{s}^{-1}} \right) + 2 \log \left(\frac{\text{FWHM}}{\text{km} \cdot \text{s}^{-1}} \right), \quad (5.9)$$

with λL_{λ} indicating the continuum luminosity at a specified wavelength (e.g., 5100 Å for H β , 3000 Å for Mg II, and 1350 Å for C IV). Coefficients a and b are sourced from McLure & Jarvis (2004) and Vestergaard & Peterson (2006). Alternative methods, such as stellar velocity dispersion and host galaxy bulge luminosity, also serve as estimators for M_{BH} (Gültekin et al., 2009). Paliya et al. (2021) employed these techniques to compute M_{BH} for a selection of *Fermi* blazars, incorporating 269 sources from our sample.

The upper panel of Fig. 5.2 illustrates the linear relationship between $\log (M_{\text{BH}}/M_{\odot})$ and $\log \beta_{\text{app}}^{\text{max}}$, formulated as:

$$\log (M_{\text{BH}}/M_{\odot}) = (0.12 \pm 0.06) \log \beta_{\text{app}}^{\text{max}} + (8.57 \pm 0.06), \quad (5.10)$$

with a correlation coefficient (r) of 0.12 and a probability value (p) of 0.05.

The luminosity of the accretion disk (L_{Disk}) is typically derived from either SED modeling of the blue bump or the broad emission-line luminosity ($L_{\text{BLR}} = 0.1 L_{\text{Disk}}$), where 0.1 represents the BLR covering factor. The computation of L_{BLR} utilizes the equation:

$$L_{\text{BLR}} = \sum_i L_i \cdot \frac{\langle L_{\text{BLR, rel}} \rangle}{\sum_i L_{i, \text{rel}}}, \quad (5.11)$$

where $\langle L_{\text{BLR, rel}} \rangle = 556$, L_i is the observed line luminosity, and $L_{i, \text{rel}}$ is the relative line luminosity. Given that the reference flux for Ly α contributes 100, the relative weights for H α , H β , Mg II, and C IV lines are 77, 22, 34, and 63, respectively, culminating in a total broad emission-line flux of 556 (Celotti et al., 1997; Francis et al., 1991; Paliya et al., 2021; Xiao et al., 2022a). Utilizing these methodologies, Paliya et al. (2021) calculated L_{Disk} or established 3σ upper limits for a sample of *Fermi* blazars, inclusive of 254 sources from our collection.

Displayed in the middle panel of Fig. 5.2 is the positive correlation between $\log L_{\text{Disk}}$ and $\log \beta_{\text{app}}^{\text{max}}$, expressed as:

$$\log L_{\text{Disk}} = (1.59 \pm 0.10) \log \beta_{\text{app}}^{\text{max}} + (44.22 \pm 0.08), \quad (5.12)$$

with a correlation coefficient (r) of 0.32 and a chance probability (p) of 1.1×10^{-7} after removing the redshift effect.

The Eddington luminosity (L_{Edd}) is a function of the Eddington accretion rate (\dot{M}_{Edd}) and is expressed as $L_{\text{Edd}} = \dot{M}_{\text{Edd}} c^2 = 1.26 \times 10^{38} (M/M_{\odot}) \text{ erg} \cdot \text{s}^{-1}$. L_{Edd} was calculated for sources with known black hole mass in our sample. The correlation between $\log (L_{\text{Disk}}/L_{\text{Edd}})$ and $\log \beta_{\text{app}}^{\text{max}}$ is depicted in the lower panel of Fig. 5.2. The linear regression for this correlation is given by:

$$\log (L_{\text{Disk}}/L_{\text{Edd}}) = (0.71 \pm 0.11) \log \beta_{\text{app}}^{\text{max}} - (1.88 \pm 0.11), \quad (5.13)$$

with a correlation coefficient (r) of 0.40 and a probability value (p) of 4.8×10^{10} .

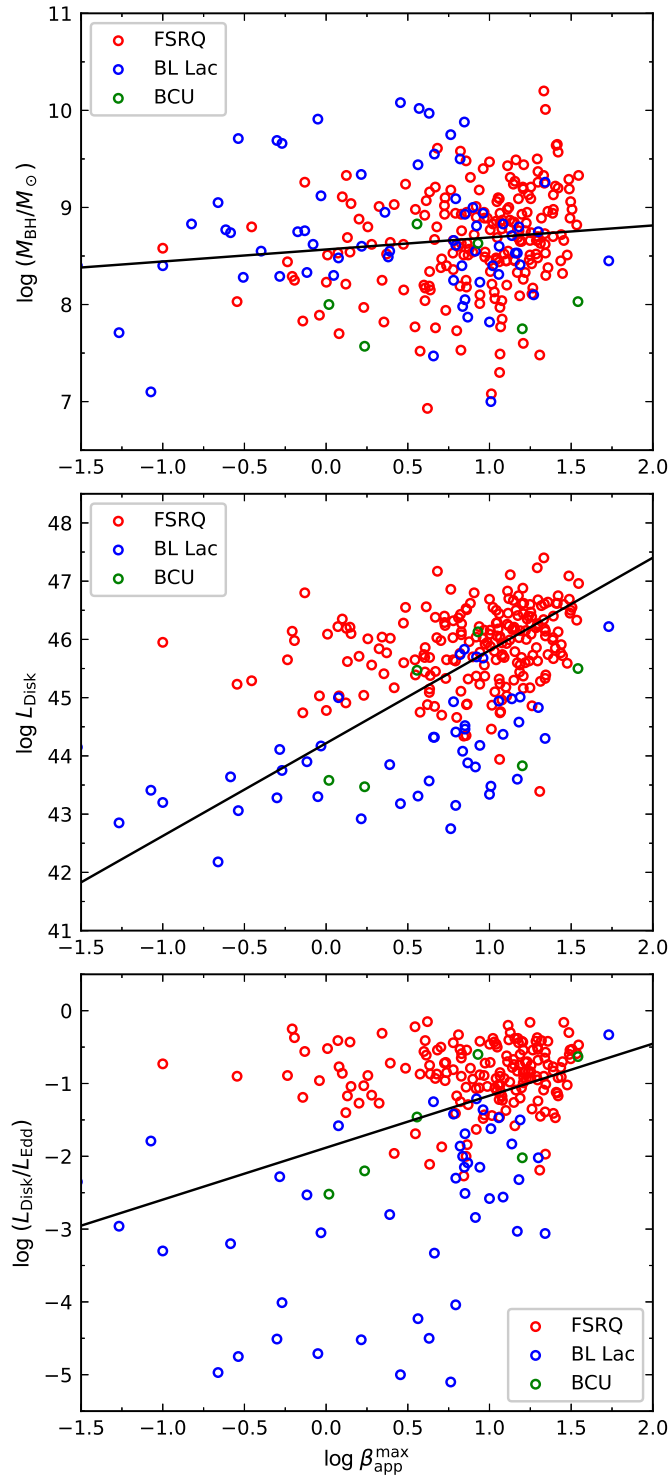


Figure 5.2: the correlations among various logarithmic parameters: the black hole mass ($\log \left(\frac{M_{\text{BH}}}{M_{\odot}} \right)$), the disk luminosity ($\log L_{\text{Disk}}$), and the ratio of disk luminosity to Eddington luminosity ($\log \left(\frac{L_{\text{Disk}}}{L_{\text{Edd}}} \right)$). These parameters are analyzed against the maximum apparent velocity ($\log \beta_{\text{app}}^{\text{max}}$) for VFBS. FSRQs are denoted in red, BL Lacs in blue, and BCUs in green.

5.4 Discussion

5.4.1 Apparent Motion and Central Engine in Blazars: A Correlational Analysis

In our investigation of the correlation between the maximum apparent velocity and GeV γ -ray luminosity for the Variable Frequency Blazars (VFBs) in our sample, a positive correlation has been observed between $\log L_\gamma$ and $\log \beta_{\text{app}}^{\text{max}}$. It is crucial, however, to acknowledge a significant caveat about the Doppler beaming effect, which is known to amplify blazar γ -ray emissions (Xiao et al., 2015; Pei et al., 2016; Fan et al., 2016; Xiao et al., 2019, 2020; Zhang et al., 2020; Paliya et al., 2021).

To account for this effect, we recalculated the γ -ray luminosity (L_γ) by incorporating the Doppler beaming influence as per the following formula:

$$L_\gamma^{\text{in}} = L_\gamma^{\text{ob}} / \delta^{q+1}, \quad (5.14)$$

where δ represents the Doppler factor, computed using the methodology proposed by Zhang et al. (2020). Here, q is defined as $2 + \alpha$ for continuous jet emissions (or $q = 3 + \alpha$ for discrete jet emissions), with α being the spectral index ($\alpha_{\text{ph}} - 1$) where $f_\nu \propto \nu^{-\alpha}$.

A linear regression analysis of the intrinsic γ -ray luminosity $\log L_\gamma^{\text{in}}$ against $\log \beta_{\text{app}}^{\text{max}}$ (illustrated in Fig. 5.3) yields the following results:

- For $q = 2 + \alpha$, with the redshift effect negated, we obtain

$$\log L_\gamma^{\text{in}} = (1.84 \pm 0.10) \log \beta_{\text{app}}^{\text{max}} + (41.72 \pm 0.08),$$

accompanied by a correlation coefficient $r = 0.26$ and a p-value of 5.8×10^{-7} .

- For $q = 3 + \alpha$, again with the redshift effect negated, the regression gives

$$\log L_\gamma^{\text{in}} = (2.02 \pm 0.13) \log \beta_{\text{app}}^{\text{max}} + (40.65 \pm 0.10),$$

with a correlation coefficient $r = 0.23$ and a p-value of 1.0×10^{-5} .

These positive correlations between $\log \beta_{\text{app}}^{\text{max}}$ and both $\log L_\gamma$ and $\log L_\gamma^{\text{in}}$ suggest a potential link between the motion of jet knots and the activity of the blazar's central engine.

In astrophysical jet studies, the total jet power, P_{jet} , is commonly decomposed into radiation power, P_{rad} , and kinetic power, P_{kin} . The former is linked to the jet's nonthermal radiation, while the latter relates to its propagation dynamics. A robust method to estimate these powers is through broadband Spectral Energy Distribution (SED) fitting, particularly with simultaneous data, providing insights into both kinetic and radiation components (Ghisellini et al., 2014; Tan et al., 2020).

Ghisellini & Tavecchio (2010) and Ghisellini et al. (2014) formulated the radiation power as

$$P_{\text{rad}} = 2f \frac{\Gamma^2 L_{\text{jet}}^{\text{bol}}}{\delta^4}, \quad (5.15)$$

where the factor of 2 accounts for bi-directional jets. The parameter f varies with the emission process, taking the value of 16/5 for the Synchrotron Self-Compton (SSC) process, predominantly in BL Lacertae objects, and 4/3 for the External Compton (EC) process, usually in Flat Spectrum Radio Quasars (FSRQs). Additionally, for the EC process, δ^4 is modified to $\delta^4(\delta/\Gamma)^2$.

Two assumptions are pivotal in calculating P_{rad} . Firstly, for blazars, the bulk Lorentz factor Γ is approximated to be equal to the Doppler factor δ , owing to their small viewing angles (Ghisellini &

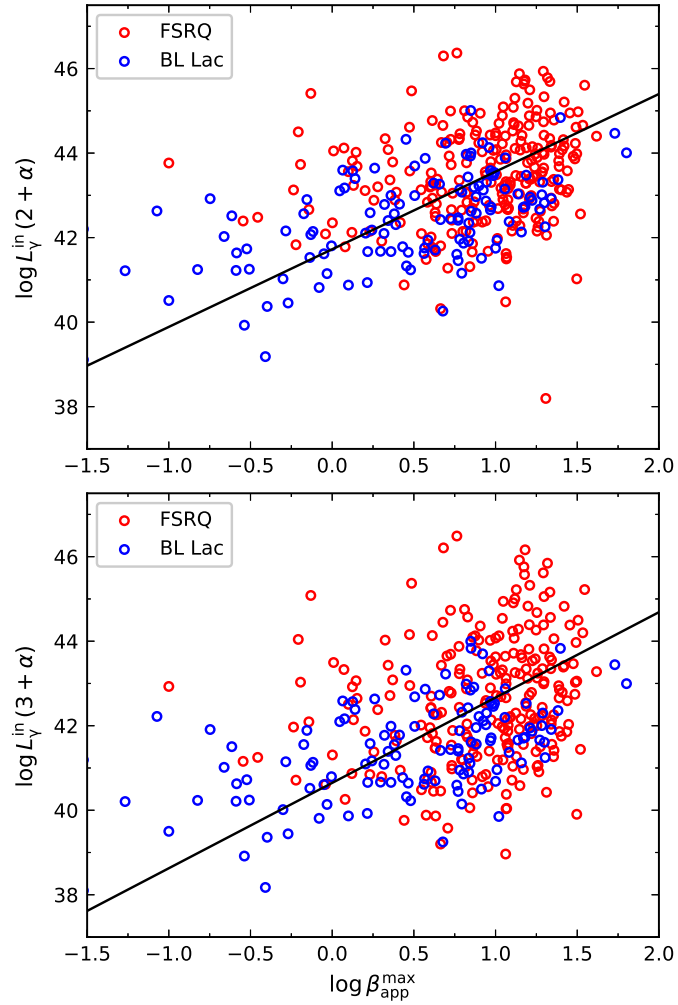


Figure 5.3: The correlations between the intrinsic logarithmic luminosities, denoted as $\log L_{\gamma}^{\text{in}}$, and the maximum apparent velocities, $\log \beta_{\text{app}}^{\text{max}}$, for the VFBs. The FSRQs are represented using red markers, whereas the BL Lacs are depicted with blue markers.

Table 5.2: The correlation of $\log P_{\text{rad}}$ and $\log L_{5\text{GHz}}^{\text{ext}}$ against $\log \beta_{\text{app}}^{\text{max}}$

Type	class	$(a \pm \Delta a)$	$(b \pm \Delta b)$	N	r	p
$\log P_{\text{rad}}$ vs. $\log \beta_{\text{app}}^{\text{max}}$	All	1.78 ± 0.12	43.62 ± 0.10	249	0.29	5.0×10^{-6}
	FSRQs	1.34 ± 0.12	44.12 ± 0.11	207	0.19	6.9×10^{-3}
	BL Lacs	1.59 ± 0.19	43.20 ± 0.07	42	0.34	0.03
$\log L_{5\text{GHz}}^{\text{ext}}$ vs. $\log \beta_{\text{app}}^{\text{max}}$	All	2.16 ± 0.12	40.97 ± 0.09	270	0.27	1.0×10^{-5}
	FSRQs	1.27 ± 0.15	42.10 ± 0.14	176	0.12	0.11
	BL Lacs	1.91 ± 0.16	40.52 ± 0.07	88	0.25	0.02

Tavecchio, 2010; Ghisellini et al., 2014; Xiong & Zhang, 2014; Zhang et al., 2020; Xiao et al., 2022a). Secondly, the bolometric jet luminosity $L_{\text{bol}}^{\text{jet}}$ is represented by the inverse Compton peak luminosity L_{IC} instead of the gamma-ray luminosity L_{γ} , to avoid degeneracy between L_{γ} and δ in Eq. 5.15. L_{IC} is expressed as $L_{\text{IC}} = 4\pi d_L^2 F_{\text{IC}}$, with F_{IC} being the inverse Compton peak flux obtained from Paliya et al. (2021).

Cavagnolo et al. (2010) proposed that the kinetic power of jets, P_{kin} , can be gauged from the energy required to inflate X-ray cavities in various astronomical systems, including giant elliptical and cD galaxies. However, this method's applicability is currently limited to a select number of sources. An alternative approach links the extended region's radio luminosity, less affected by Doppler boosting, to the jet's kinetic power (Rawlings & Saunders, 1991; Willott et al., 1999; Cavagnolo et al., 2010; Meyer et al., 2011), as shown by

$$P_{\text{rad}} = \eta (L_{5\text{GHz}}^{\text{ext}})^{\kappa}. \quad (5.16)$$

The parameters κ and η vary in the literature, reflecting differences in sample sizes and source types (Cavagnolo et al., 2010; Meyer et al., 2011). Nevertheless, a logarithmic scaling between $\log L_{5\text{GHz}}^{\text{ext}}$ and $\log P_{\text{rad}}$ is observed.

Radio flux densities at various frequencies (5 GHz, 8.4 GHz, and 15 GHz) are sourced from literature (Taylor et al., 1996; Piner & Edwards, 2014; Lister et al., 2018) and converted to 5 GHz using

$$S_{5\text{GHz}}^{\text{core}} = S_{\nu}^{\text{core}} \quad \text{and} \quad S_{5\text{GHz}}^{\text{ext}} = S_{\nu}^{\text{ext}} \left(\frac{\nu}{5\text{GHz}} \right)^{\alpha_{\text{ext}}}, \quad (5.17)$$

where $\alpha_{\text{ext}} = 0.75$ and $\alpha_{\text{core}} = 0$ (Fan et al., 2011; Pei et al., 2016, 2019, 2020). Incorporating the radio core dominance parameter at 5 GHz, collected from Pei et al. (2020),

$$R = \left(\frac{S^{\text{core}}}{S^{\text{ext}}} \right) (1+z)^{\alpha_{\text{core}} - \alpha_{\text{ext}}}, \quad (5.18)$$

we deduce $S_{5\text{GHz}}^{\text{ext}}$ and subsequently compute $\log L_{5\text{GHz}}^{\text{ext}}$. Our analysis, as demonstrated in Fig. 5.4 and Tab. 5.2, reveals positive correlations between $\log P_{\text{rad}}$ and $\log \beta_{\text{app}}^{\text{max}}$, as well as between $\log L_{5\text{GHz}}^{\text{ext}}$ and $\log \beta_{\text{app}}^{\text{max}}$ in blazars. These correlations hold individually for FSRQs and BL Lacs in the case of $\log P_{\text{rad}}$ versus $\log \beta_{\text{app}}^{\text{max}}$, but only for BL Lacs in the case of $\log L_{5\text{GHz}}^{\text{ext}}$ versus $\log \beta_{\text{app}}^{\text{max}}$. This suggests a significant correlation between jet knot motion and jet radiation power in both FSRQs and BL Lacs, whereas a correlation with kinetic power is evident only in BL Lacs.

Our investigation explored the relationship between apparent motion and key astrophysical parameters: black hole mass, accretion disk luminosity, and normalized disk luminosity (see Fig. 5.2). The

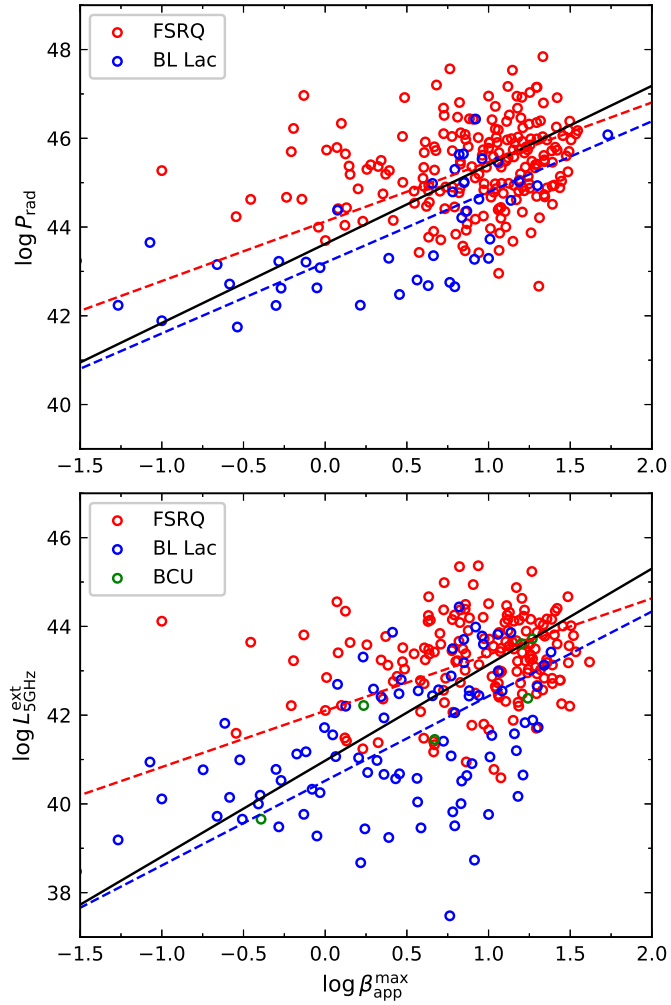


Figure 5.4: the relationship between the logarithmic values of radiative power (\log, P_{rad}) and extended luminosity at 5 GHz ($\log, L_{5\text{GHz}}^{\text{ext}}$) with the logarithmic maximal apparent beta ($\log, \beta_{\text{app}}^{\text{max}}$) of VFBs. FSRQs represented in red, BL Lacs in blue, and BCUs in green.

analysis revealed that the maximum apparent velocity logarithm ($\log \beta_{\text{app}}^{\text{max}}$) does not significantly correlate with the black hole mass logarithm ($\log (M_{\text{BH}}/M_{\odot})$), evidenced by a p -value of 0.05. However, there are moderate positive correlations between $\log \beta_{\text{app}}^{\text{max}}$ and both $\log L_{\text{Disk}}$ and $\log (L_{\text{Disk}}/L_{\text{Edd}})$. These correlations suggest that accretion disks with higher luminosity and accretion rates are associated with faster apparent velocities in jet knots. Consequently, our findings imply that the mass of the central black hole may not significantly influence the generation and acceleration of jet knots. In contrast, the accretion disk's power and rate appear more critical. Prior research has linked blazar knot propagation disturbances with γ -ray flares and overall jet activity (Jorstad et al., 2001; Marscher et al., 2010; Jorstad et al., 2013; Tanaka et al., 2015; Jorstad & Marscher, 2016; Jorstad et al., 2017), often observing γ -ray outbursts concurrent with superluminal knot transits through the radio core (Jorstad & Marscher, 2016). Our model proposes a mechanism whereby an abrupt increase in the accretion rate of a potent accretion disk could generate and accelerate knots, thereby driving jet activities and triggering outbursts.

5.4.2 Criteria for Selecting VFB Candidates from *Fermi* Blazars

Blazars exhibit a range of observational characteristics, notably high γ -ray luminosity. In the GeV γ -ray band, the spectrum represents a composite of various radiation processes, including synchrotron self-Compton scattering (SSC), scattering of seed photons from the accretion disk (Dermer & Schlickeiser, 1993c), ultraviolet seed photons from the BLR (Sikora et al., 1994d), and seed photons from the dusty torus (DT) (Błażejowski et al., 2000c; Arbeiter et al., 2002; Sokolov & Marscher, 2005). These processes are mediated by inverse Compton scattering by relativistic electrons. Consequently, blazars should exhibit spectral differences depending on the dominant γ -ray emission mechanisms.

A study by Ackermann et al. (2015d) revealed that flat-spectrum radio quasars (FSRQs) demonstrate higher γ -ray luminosity L_{γ} and a larger photon index α_{ph} compared to BL Lacs, based on analysis of the third catalog of active galactic nuclei (AGNs) detected by *Fermi*-LAT (3LAC). Blazar variability is another distinctive trait; Yang et al. (2022b) observed that FSRQs exhibit greater $\log VI$ than BL Lacs. Our analysis indicates a positive correlation between $\log \beta_{\text{app}}^{\text{max}}$ and both α_{ph} and $\log VI$, as depicted in the middle and bottom panels of Fig. 5.1. This suggests that blazars with faster apparent motion tend to have a softer γ -ray spectrum and higher variability magnitude.

In this study, we investigate the differences between VLBI-detected *Fermi* blazars and other *Fermi* blazars, focusing on three parameters: $\log L_{\gamma}$, α_{ph} , and $\log VI$. Our findings indicate that VFBs exhibit higher values in these parameters compared to other *Fermi* blazars.

Lister et al. (2015), using the complete MOJAVE 1.5 Jy sample of AGNs, found that 23% of these AGNs were not detected by *Fermi* due to instrumental selection effects and lower Doppler boosting factors. Xiao et al. (2019) proposed that blazars not detected by *Fermi*-LAT but having large apparent velocities could be potential *Fermi* blazar candidates. This hypothesis was substantiated by Xiao et al. (2020) following the release of the 4FGL. While VLBI-detected blazars are likely *Fermi* γ -ray source candidates, the question remains whether additional *Fermi* catalog blazars are VLBI candidates. Out of the 3437 identified blazars in 4FGL_DR2, 407 (11.8%) are associated with VLBI detections, primarily by the MOJAVE program. However, this number is likely an underestimate, given MOJAVE's focus on northern sky sources and *Fermi*-LAT's all-sky observation capabilities.

To identify more VFB candidates, we explore criteria based on $\log L_{\gamma}$, α_{ph} , and $\log VI$. We employ GMM from the *sklearn* machine learning library in Python (Fraley & Raftery, 2002) to decompose the distributions of these parameters, as shown in the left panel of Fig. 5.5. The GMM, a probabilistic model assuming data points are generated from a mixture of Gaussian distributions, uses the EM algorithm (see Chap. 3) for parameter estimation. The algorithm iteratively performs 'E' (Expecta-

tion) and 'M' (Maximization) steps until convergence of the estimated parameters. We implement stratified sampling at a 95% scale and repeat the GMM process 100 times to enhance our model's generalizability. The Bayesian Information Criterion (BIC) is used to evaluate the GMM model and determine the number of Gaussian components, with lower BIC values indicating preferred models (Kass & Raftery, 1995).

The BIC analysis suggests a three-component GMM for $\log L_\gamma$ with the lowest BIC of 4943.0 and $\Delta\text{BIC}_{32} = 12.8$, as illustrated in the top-right panel of Fig. 5.5. Similarly, a two-component model is favored for both α_{ph} and $\log VI$, with respective BIC values of 1370.1 ($\Delta\text{BIC}_{32} = 35.$) and 3585.0 ($\Delta\text{BIC}_{32} = 17.4$), as shown in the middle-right and bottom-right panels of Fig. 5.5.

The analysis demonstrates that *Gaussian*[1] and *Gaussian*[2] predominantly encapsulate the VFBs for $\log L_\gamma$. In contrast, *Gaussian*[0] almost exclusively encompasses the VFBs for α_{ph} and fully envelops them for $\log VI$. We postulate that the dotted-curve demarcates the 'VFB-like' sources, the dash-dotted curve delineates the 'transitional' sources, and the dashed-curve encompasses the 'non-VFB-like' sources. The intersection of the dotted and dashed curves is proposed as the boundary for selecting VFB candidates. Our computations yield the coordinates of this junction point as $\log L_\gamma = 45.40 \pm 0.22$, $\alpha_{\text{ph}} = 2.24 \pm 0.01$, and $\log VI = 1.71 \pm 0.07$. A total of 228 sources, surpassing these boundary values in luminosity, photon index, and variability index, are predicted as VFB candidates and enumerated in Tab. 5.3, where col.(1) presents the identifier as per the 4FGL catalog; col. (2) details the type or category of the object; col. (3) lists the cosmic redshift value; col. (4) enumerates the integrated photon flux ranging from 1 to 100 GeV; col. (5) records the γ -ray luminosity measurement; col. (6) denotes the index of the photon spectrum; col. (7) specifies the index quantifying variability.

It is observed that 79 VFBs (19.4%) exhibit $\log L_\gamma < 45.40$, 144 VFBs (35.4%) have $\alpha_{\text{ph}} < 2.24$, and 142 VFBs (34.9%) display $\log VI < 1.71$, which corresponds to a VI of 51.29. Therefore, it is imperative to acknowledge the considerable presence of VFBs with $\log L_\gamma$, α_{ph} , and/or $\log VI$ below the proposed boundary. Our methodology and criteria are intended to identify more promising VFB candidates.

Of the 228 identified VFB candidates, 88 (38.6%) are located in the northern sky, while 140 (61.4%) are in the southern hemisphere. This distribution should assist VLBI facilities in locating more blazars exhibiting jet 'knot' structures and in monitoring their kinematics, particularly for the MOJAVE program.

5.5 Conclusion

In this study, we employed a sample of VLBI-detected *Fermi* blazars to investigate the properties of blazar jets and accretion disks. Our findings can be summarized as follows:

1. A significant correlation was observed between the maximum apparent velocity ($\log \beta_{\text{app}}^{\text{max}}$) and the γ -ray luminosity ($\log L_\gamma$). This correlation persisted even after removing the effects of Doppler beaming, suggesting its intrinsic nature.
2. The relationship between $\log L_\gamma$ and $\log \beta_{\text{app}}^{\text{max}}$ indicates that the apparent motion of knots is linked to jet power. A higher $\log \beta_{\text{app}}^{\text{max}}$ is indicative of a more potent jet. This correlation was evident in both FSRQs and BL Lacs. However, only in BL Lacs was $\log \beta_{\text{app}}^{\text{max}}$ found to correlate with the 5 GHz extended region luminosity ($\log L_{5\text{GHz}}^{\text{ext}}$), a proxy for jet kinetic power.
3. No correlation was found between $\log \beta_{\text{app}}^{\text{max}}$ and black hole mass ($\log(M_{\text{BH}}/M_\odot)$). However, moderate correlations were observed with the accretion disk luminosity ($\log L_{\text{Disk}}$) and the nor-

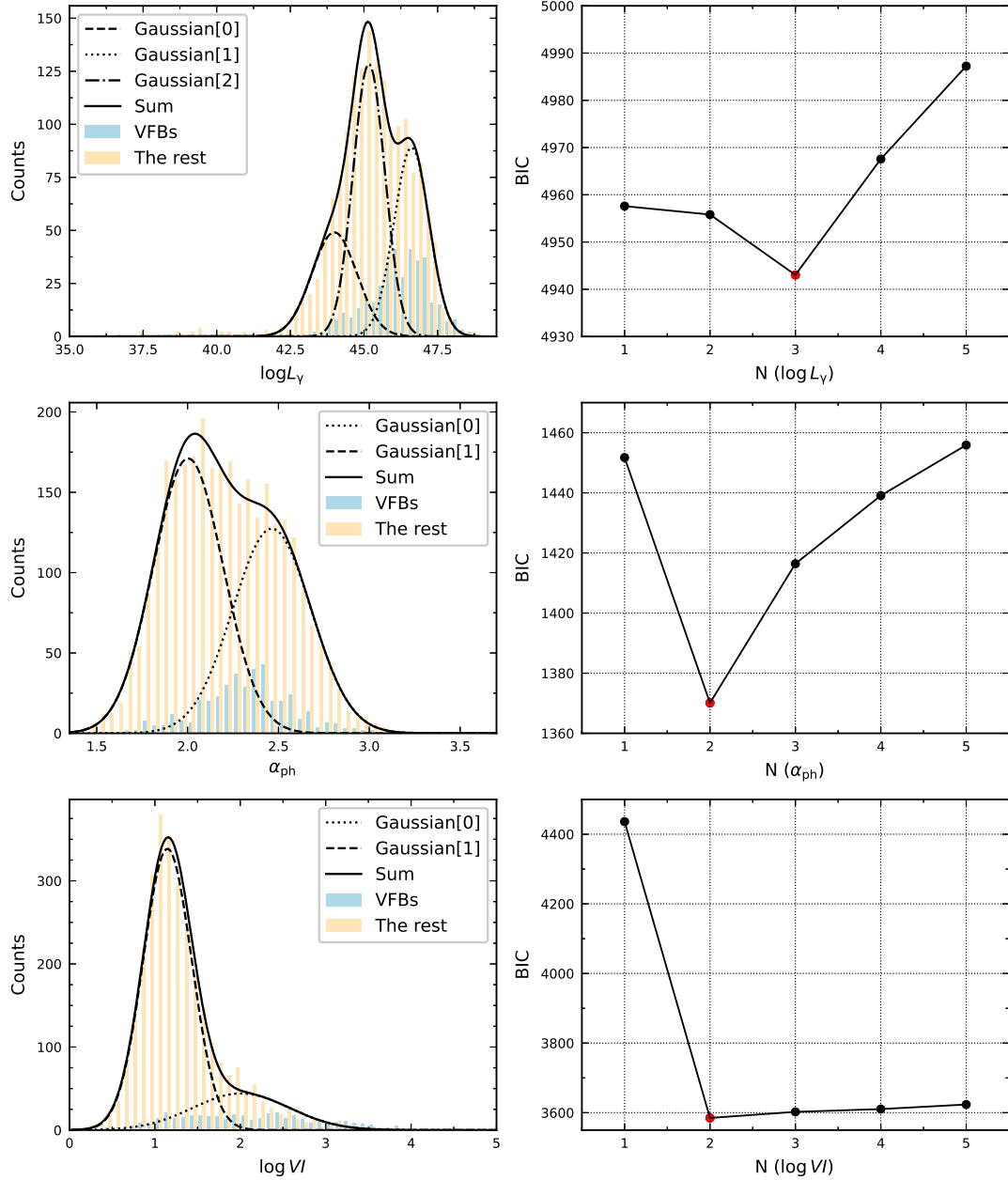


Figure 5.5: Distributions for the parameters: γ -ray luminosity (denoted as $\log L_\gamma$), photon index (represented by α_{ph}), and the variability index (expressed as $\log VI$), were modeled GMM. The VFBs are depicted with blue bars, while the remaining *Fermi* blazars are illustrated using orange bars.

Table 5.3: Potential Candidates of Superluminal Blazars Detected by *Fermi*

4FGL Name	Class	z	Flux1000 photon \cdot cm $^{-2}$ \cdot s $^{-1}$	$\log L_\gamma$ erg \cdot cm $^{-2}$ \cdot s $^{-1}$	α_{ph}	VI
(1)	(2)	(3)	(4)	(5)	(6)	(7)
4FGL J0001.5+2113	fsrq	1.106	1.36E-09	46.74	2.66	1564.42
4FGL J0004.4-4737	fsrq	0.88	4.36E-10	45.99	2.37	139.12
4FGL J0010.6-3025	fsrq	1.19	3.49E-10	46.24	2.46	91.59
4FGL J0011.4+0057	fsrq	1.492	4.29E-10	46.60	2.33	71.98
4FGL J0016.2-0016	fsrq	1.57631	4.17E-10	46.68	2.74	82.14
...
...
...

The full details of this table can be found in Tab. 9.1 (see Appendix 9).

malized accretion disk luminosity ($\log(L_{\text{Disk}}/L_{\text{Edd}})$). These results suggest that both the power of the accretion disk and the accretion rate are crucial for generating and accelerating blazar knots. We propose a mechanism where knots are formed by sudden increases in the accretion rate of a potent accretion disk, leading to jet activities and observable outbursts.

Furthermore, our analysis revealed:

1. VLBI-detected *Fermi* blazars (VFBs) differ from the rest of the *Fermi* blazars in terms of γ -ray luminosity ($\log L_\gamma$), γ -ray photon index (α_{ph}), and variability index ($\log VI$). The average values for VFBs are $\langle \log L_\gamma^{\text{VFB}} \rangle = 46.20 \pm 1.07$, $\langle \alpha^{\text{VFB}} \rangle = 2.33 \pm 0.25$, and $\langle \log VI^{\text{VFB}} \rangle = 2.14 \pm 0.84$, while those for the rest of the *Fermi* blazars are $\langle \log L_\gamma^{\text{R}} \rangle = 45.24 \pm 1.42$, $\langle \alpha^{\text{R}} \rangle = 2.21 \pm 0.30$, and $\langle \log VI^{\text{R}} \rangle = 1.30 \pm 0.50$.
2. Employing a Gaussian Mixture Model (GMM) method, we established criteria for identifying VFB candidates among other *Fermi* blazars. A blazar with $\log L_\gamma > 45.40$, $\alpha_{\text{ph}} > 2.24$, and $\log VI > 1.71$ is likely a VFB candidate. Utilizing this criterion, we identified 228 potential VFBs.

Chapter 6

Investigating TeV blazar candidates of *Fermi* blazars

6.1 Introduction

As discussed in Chap. 1, Blazars, including BL Lacertae objects, can be categorized based on the peak frequency of their synchrotron peak frequency: HSP/HBL, ISP/IBL, LSP/LBL or LBL. HSPs and HBLs are frequently identified as potential sources of very high energy (VHE, above 300 GeV) emission. Notably, sources with smaller redshifts have a significant proportion of TeV photons observable. The emission of VHE photons with energies > 300 GeV from blazars unveils novel phenomena. Particularly, photons in the TeV band pose challenges to particle acceleration theories in jets. They are also pivotal for indirectly measuring the extragalactic background light, estimating the intergalactic magnetic field, and exploring the potential origins of high-energy extragalactic neutrinos (Foffano et al., 2019). For example, observations from the Very Long Baseline Array (VLBA) have provided 23 images of 6 TeV blazars, demonstrating that apparent jet bending is a common characteristic of TeV blazars (Piner et al., 2010).

The flux emitted from most astronomical sources in the TeV band is notably low. Additionally, the extragalactic background light (EBL) significantly absorbs the most energetic γ -ray emissions (Primack et al., 1999; Kneiske et al., 2004; Fermi-LAT Collaboration et al., 2018) through the interaction $\gamma + \gamma \rightarrow e^+ + e^-$. This interaction is fundamentally related to numerous astrophysical problems (Domínguez et al., 2019). Consequently, observations at TeV energies necessitate large collection areas, achievable only with ground-based detectors such as atmospheric Cherenkov telescopes (IACTs) and extended arrays of particle detectors (EAS arrays).

The detection techniques for γ -rays in IACTs and EAS arrays differ. IACTs detect Cherenkov photons in the atmosphere, produced by the EAS of secondary particles initiated by primary γ -rays. Prominent examples include *the High Energy Stereoscopic System*¹ (Aharonian et al. 2004, H.E.S.S), *the Major Atmospheric Gamma Imaging Cherenkov Telescopes*² (MAGIC Collaboration 2000, MAGIC), *the Very Energetic emission Imaging Telescope Array System*³ (VERITAS Collaboration et al. 2005, VERITAS), and the next-generation IACT array: *the Cherenkov Telescope Array Observatory*⁴ (The CTA Consortium et al. 2011, CTAO). In contrast, EAS arrays detect secondary

¹<https://www.mpi-hd.mpg.de/hfm/HESS/>

²<https://magic.mpp.mpg.de/>

³<https://veritas.sao.arizona.edu/>

⁴<https://www.cta-observatory.org/>

particles from cosmic rays that reach the ground, such as *the High Altitude Water Cherenkov Observatory*⁵ (Abeysekara et al. 2017b, HAWC), and *the Large High Altitude Air Shower Observatory*⁶ (Cao et al. 2019, LHAASO).

As of August 2022, approximately 90 extragalactic TeV sources have been confirmed (Wakely & Horan, 2008)⁷, among which 81 are TeV blazars. The identification of TeV blazar candidates (TBCs) is both exciting and challenging, given the limitations imposed by the EBL and the sensitivity of detectors, which restrict the number of high-redshift TeV sources. Improved sensitivity, within the constraints of the EBL, will likely expand the sample of TeV blazars in terms of both quantity and higher redshift sources. The current and next-generation IACTs and EAS arrays, sensitive to photons at TeV energies, are expected to increase the population of TeV blazars, thereby enhancing our understanding of this classification and the mechanisms of VHE emission (Costamante & Ghisellini, 2002; Massaro et al., 2013; Chang et al., 2017; Foffano et al., 2019).

Most TeV blazars are HSPs, predominantly BL Lacs, thus searches for TBCs are often focused on HSPs/BL Lacs (Costamante & Ghisellini, 2002; Massaro et al., 2013; Chang et al., 2017; Foffano et al., 2019; Chiaro et al., 2019). Typically, BL Lacs are HSPs, while FSRQs are mostly LSPs and ISPs. LSP/ISP often exhibit EC components, where electrons encounter a strong, Doppler-shifted photon field. This can lead to significant γ -ray emissions, occasionally reaching the VHE regime. However, for LSP/ISPs, interactions frequently occur in the Klein-Nishina regime, and the same strong photon field often results in substantial absorption of the produced γ -ray photons. In contrast, the VHE emission of HSPs is believed to originate from low-energy photons produced by synchrotron emission from ultrarelativistic electrons in the jet, as per the SSC process. Moreover, HSPs have higher $\log \nu_p^s$, facilitating the scattering of their synchrotron bump into the TeV band by relativistic electrons. Nonetheless, it is not evident that extreme $\log \nu_p^s$ values alone lead to TeV γ -ray emissions, as this also depends on various other properties of the emission region, such as electron distribution, magnetic field strength, internal absorption, and the source's redshift (Nievas Rosillo et al., 2022).

Lin & Fan (2016) collected a sample of 662 *Fermi* BL Lac objects, which included 47 TeV sources. Their study involved a comparative analysis of the multiwavelength observational properties between TeV and non-TeV sources. The findings indicated that TeV sources generally exhibit a lower average redshift, higher flux density, and a harder γ -ray spectrum compared to their non-TeV counterparts.

It is well-documented that flux variability across all wavebands is a prevalent characteristic among blazars (Fan et al., 2017; Majumder et al., 2019; Yang et al., 2022a,c). This variability ranges from moderate to highly violent and occurs over diverse timescales, from hours to years. Notably, blazars such as 1ES 0229+200 have been observed to exhibit moderate variability. Furthermore, there is a proportional relationship between the $\log \nu_p^s$ and the variability of blazars.

In their analysis of 50 observations of 12 LSP blazars from *XMM-Newton*, Gupta et al. (2016) discovered that LSP blazars exhibit slower variability in the X-ray bands compared to the IR/optical bands. This phenomenon is attributed to the dominance of the IC mechanism in X-ray emission for LSP blazars. Conversely, HSP blazars are anticipated to show more pronounced variability in the X-ray bands. According to the SSC model, relativistic electrons upscatter X-rays into the TeV region. This correlation between X-ray and TeV emissions has been confirmed by various observations (Costamante & Ghisellini, 2002; Singh et al., 2019; Osorio et al., 2019). For instance, during flares observed in the TeV band, the synchrotron peak frequencies of Markarian 501 and 1ES 1959+650 in the X-ray band shifted to higher energy regions (Sambruna et al., 2000; Kapanadze et al., 2018; Singh

⁵<https://www.hawc-observatory.org/>

⁶<http://english.ihep.cas.cn/lhaaso/>

⁷<http://TeVcat.uchicago.edu>

et al., 2019). However, Foffano et al. (2019) presented a counterexample by observing two groups of EHBLs with the same range of $\log \nu_p^s$ but opposite spectral slopes in the TeV band, suggesting independence between $\log \nu_p^s$ and TeV emission.

Since its launch in 2008, the *Fermi* Large Area Telescope (*Fermi*-LAT) has significantly advanced the study of blazars, progressively dominating the field of high-energy γ -ray astronomy. *Fermi*-LAT has detected thousands of blazars and reported them in various publications (e.g., Abdollahi et al. 2020). The *Fermi*-LAT 10-year source catalog (4FGL-DR2, Abdollahi et al. 2020; Ballet et al. 2020), along with the *Fermi*-LAT 10-year AGN catalog (4LAC-DR2, Ajello et al. 2020; Lott et al. 2020), listed 80 out of the 81 known TeV blazars (hereafter referred to as *Fermi* blazars). Consequently, there is a strong basis to believe that the *Fermi* catalogs likely contain numerous TBCs.

From a broader perspective, this chapter endeavors to expand the pool of TBCs, transcending the constraints of confining the search solely to High Synchrotron Peaked sources (HSPs). Employing machine learning techniques, we aim to identify the physical characteristics that fundamentally differentiate TeV blazars from their non-TeV counterparts. This involves quantifying the efficacy of the demarcation criteria used to distinguish between these two categories. Furthermore, we calculate the likelihood of a source being classified as a TeV source based on these physical attributes. This approach enables us to extract potential TBCs from a pool of non-TeV sources. The data for this analysis have been collated from four comprehensive catalogs: the 4FGL-DR2 / 4LAC-DR2, the Third Catalog of Hard *Fermi*-LAT Sources (3FHL, Ajello et al. (2017)), the Third Catalog of High Synchrotron Peaked Blazars (3HSP, Chang et al. (2019)), and the Second Brazil-ICRANet Gamma-ray Blazars (2BIGB, Arsioli et al. (2020)). Subsequently, we assess the feasibility of detecting these candidates using ground-based Cherenkov telescopes.

6.2 SML

In this chapter, binary labels are used to distinguish TeV blazars from non-TeV blazars. The model is derived from training sets and applied to the entire dataset. As a result, non-TeV sources that are assigned the same predicted labels as TeV sources are considered candidates for TeV blazars. A candidate is deemed to be a high-confidence candidate if the probability of being a TeV source exceeds 80%.

6.2.1 Data Preprocessing

Refer to Chapter 1 for a detailed review of the data preprocessing employed. This process involves transforming or modifying data to create a more suitable representation for downstream models. The steps involved in data preprocessing include:

1. Data Discretization

This step involves breaking down a feature represented across multiple columns into distinct components. For instance, in the 4FGL-DR2 dataset, the feature *Flux_{band}* is spread across seven columns, each representing the integral photon flux in a specific spectral band. These are discretized into individual features: *Flux_{band_1}*, *Flux_{band_2}*, ..., *Flux_{band_7}*.

2. Data Cleaning

This step addresses the removal or correction of erroneous or irrelevant data, which can negatively impact model performance. It involves discarding features with errors, non-numeric (string) features, and those with a significant number of missing values.

3. Data Standardization

To eliminate the influence of different scales across features, data standardization is applied. This process utilizes tools like *sklearn.preprocessing.StandardScaler* to transform the features into a standard normal distribution, thereby making them dimensionless and comparable.

4. Data Partitioning

The dataset is randomly split into training and testing sets in a 4:1 ratio. This partitioning is done using the *StratifiedKFold.split* method from the *sklearn.model_selection* package in Python. The process ensures that each set maintains a proportional representation of each class. This step is repeated five times with different random seeds (0, 1, 2, 3, 4), resulting in five distinct pairs of training and testing sets, labeled as training 1 / testing 1, training 2 / testing 2, and so on.

6.2.2 Selection of the SML model

In this study, we focus on the classification of TeV blazars versus non-TeV blazars using SML methods. Our choice of the SML model is guided by two primary considerations:

1. We require a model that provides high interpretability and operates efficiently without extensive computational resources. The model should delineate a clear linear boundary in the feature space to distinguish between TeV and non-TeV blazars.
2. Upon acquiring the physical properties (features) of a blazar, we aim to compute the conditional probability that the blazar is a TeV one, given these features.

Several advanced SML classification methods exist, such as LR, SVM, ANN, and random forests (RF) (Tin Kam Ho, 1995). Among these, we select the LR model for our analysis, primarily for two reasons:

- **Mathematical Rationale**

We hypothesize that the binary labels of blazars follow a Bernoulli distribution. Given that the logistic function is the expectation of a Bernoulli distribution, LR emerges as the most natural method to map the linear combination of a blazar's physical properties into this distribution.

- **Algorithmic Simplicity**

Compared to SVM, RF, and ANN, LR is less complex, leading to lower computational costs and enhanced model interpretability. While SVM requires additional processing through an LR model to yield conditional probabilities and ANN and RF lack guarantees for a linear boundary, LR directly provides the desired output.

For the implementation, we utilized *scikit-learn* (*sklearn*, Pedregosa et al. (2011a)), a machine learning library in Python known for its extensive range of tools. Specifically, we employed the *sklearn* API (Buitinck et al., 2013), namely *sklearn.linear_model.LogisticRegression*, which facilitates the LR model. This model was instrumental in training our dataset.

Subsequently, we applied the LR model to classify the dataset. This involved computing the *logit* values for each data point and determining the conditional probability denoted as *logistic*, which indicates the likelihood of a source emitting TeV radiation. It is important to note that TeV sources are relatively scarce in our sample. To address this imbalance and enhance the representation of TeV sources, we adjusted the *class_weight* parameter to *balance* within the *sklearn.linear_model.LogisticRegression* function.

6.2.3 Evaluating performance

In the evaluation of model performance, accuracy, defined as $\frac{TP+TN}{FP+FN+TP+TN}$, is a commonly used metric. However, this measure can be misleading in scenarios where there is a significant imbalance between classes. In our context, non-TeV sources heavily outnumber TeV sources, influencing the effectiveness of accuracy as a performance metric. To address these challenges, we propose the use of the *Area Under the Curve* (AUC) of the *Receiver Operating Characteristic* (ROC) curve as a more robust metric. The ROC curve plots the true positive rate (TPR) against the false positive rate (FPR), where $TPR = \frac{TP}{TP+FN}$ and $FPR = \frac{FP}{FP+TN}$ Fawcett (2006).

Additionally, we incorporate Youden's J statistic, also known as Youden's index Youden (1950), defined as $TPR - FPR$. This index assists in identifying the optimal threshold value (p_{thre}) for classification. The maximum Youden index corresponds to a scenario where, for a fixed TPR, the FPR is minimized. This approach ensures a stringent screening process for identifying TBCs. In practical terms, a non-TeV blazar is classified as a TeV source if its logistic score exceeds the optimal p_{thre} threshold. This method balances the need to identify as many TeV sources as possible while minimizing the misclassification of non-TeV sources.

6.3 Optimization of Model Performance through Training and Tuning

- **Training Models on Datasets**

The process begins with training models on five distinct training sets. This phase is crucial for establishing a baseline from which further optimization can be achieved.

- **Enhancing Model Performance**

Two primary strategies are employed to refine the model's performance:

- **Feature Selection**

The phenomenon known as the *curse of dimensionality* (Taylor, 2019) suggests that using all available features may impair model performance. This issue arises because high-dimensional spaces tend to be sparse, making it challenging to find similarities between data points and thus complicating data processing.

To address this, the *SequentialFeatureSelector* from python's *sklearn.feature_selection* module (abbreviated as SFS) is utilized. SFS systematically eliminates non-essential features until the remaining feature subset reaches the desired size.

- **Hyperparameters Optimization**

Certain model parameters, known as hyperparameters, cannot be derived directly from training. These must be predefined before training.

For this purpose, *sklearn.grid_search.GridSearchCV* (abbreviated as GS) is an effective tool. It exhaustively searches through a predefined range of hyper-parameter values to identify the optimal combination. In the context of *sklearn.linear_model.LogisticRegression*, six hyperparameters⁸ are optimized: *penalty*, *C*, *solver*, *tol*, *max_iter*, *l1_ratio*.

⁸https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.LogisticRegression.html

The optimization strategy is inspired by *k-fold cross-validation*, which involves dividing the training set into k subsets. In each iteration, $k - 1$ subsets are used for training, and the remaining subset serves as the validation set. This process repeats k times.

Building upon this, *nested cross-validation* (Raschka, 2015b; Varma & Simon, 2006) incorporates a two-loop system. In the outer loop, the dataset is split into training and testing sets. Within the inner loop, GS conducts k -fold cross-validation to select optimal hyperparameters. Subsequently, SFS in the outer loop identifies the best features, and the testing set evaluates the model's generalizability. This nested cross-validation is implemented on five dataset partitions, yielding five models with optimal features and hyperparameters.

A common challenge in machine learning is overfitting, where models perform well on training data but poorly on unseen data. This is often due to high variance. To mitigate this, for each of the five dataset partitions, the gap between the training set's AUC (Area Under the Curve) and the testing set's AUC is calculated. The model with the smallest positive gap is selected as the optimal model, balancing performance and generalizability.

6.3.1 Constructing Linear Boundaries

In this phase, we establish linear boundaries to differentiate between TeV-emitting and non-TeV-emitting sources. A logistic regression (LR) model is employed, where a source is classified as a TeV source if its logistic value exceeds a certain threshold (p_{thre}). Correspondingly, a threshold value for the logit function, denoted as $\text{logit}_{\text{thre}}$, is associated with p_{thre} . Hence, we can represent the linear boundary as $\text{logit}' = 0$, where $\text{logit}' = \text{logit} - \text{logit}_{\text{thre}}$.

6.3.2 Predicting Labels

To enhance our understanding of the label prediction process for unknown datasets, we introduce several transitional concepts. The dataset that has only undergone *data discretization* is termed the *initial set*. In contrast, the *learning set* refers to the dataset that has completed the entire preprocessing phase. The optimal model is derived from the *learning set*, and it also determines the optimal features. Within the *initial set*, a subset containing these optimal features is referred to as the *prediction set*. Unlike the *learning set*, which requires all features to be complete and without missing values, the *prediction set* only necessitates the presence of the optimal parameters. This inclusion of more samples in the *prediction set* aids in identifying additional high-confidence TBCs. Consequently, we can compute the logistic value for each blazar in the *prediction set*. Non-TeV blazars with a logistic value of $\geq 80\%$ are considered high-confidence TBCs.

6.4 LR Model for Distinguishing TeVs from Non-TeV

6.4.1 Sample Data Sources

Our study utilized data from four catalogs: 4FGL-DR2 / 4LAC-DR2, 3FHL, 3HSP, and 2BIGB, to distinguish between TeV-emitting and non-TeV-emitting blazars.

4FGL-DR2 / 4LAC-DR2. The 4FGL-DR2 catalog, based on a decade of Fermi-LAT observations, includes over 5700 sources, of which 3511 are AGNs, with 3436 classified as blazars. The

4LAC-DR2 catalog supplements 4FGL-DR2 with critical parameters like synchrotron peak frequency ($\log \nu_p^s$), peak intensity ($\log \nu_p^s f_{\nu_p}^s$), and redshift (z). The combined 4FGL-DR2 / 4LAC-DR2 dataset comprises 3436 blazars, including 80 TeV blazars, with 88 features. For machine learning, 861 blazars (71 TeV) with 32 features were used for training and testing, while 1459 blazars (73 TeV) with 5 features were set aside for prediction.

3FHL. The 3FHL catalog, derived from seven years of Fermi-LAT data, lists 1556 VHE sources. Our study included 1207 blazars from this catalog, with 74 identified as TeV blazars. The learning set comprised 506 blazars (63 TeV) with 18 features, and the prediction set included 540 blazars (67 TeV) with 2 features.

3HSP. The 3HSP catalog contains 2013 high synchrotron peaked (HSP) blazars, with a unique feature: the peak flux normalized to the faintest HSP blazar detected in TeVCat. Our initial dataset included 2013 blazars (57 TeV) with 13 features. The learning set had 1411 blazars (53 TeV) with 5 features, and the prediction set included 1771 blazars (54 TeV) with 2 features.

2BIGB. The 2BIGB catalog, resulting from a gamma-ray likelihood analysis of 1160 3HSP sources, provided photon flux data in the range of 500 MeV to 500 GeV. The initial dataset comprised 1160 blazars (57 TeV) with 31 features. The learning set included 1040 blazars (54 TeV) with 7 features, and the same set was used for prediction with 3 features.

6.4.2 TeV Source Identification

TeV sources were identified using the TeVCat, an online catalog of VHE sources, primarily detected by IACTs and EAS arrays. As of the latest update, TeVCat lists 251 sources, including at least 91 extragalactic sources, with 80 blazars also detected by Fermi-LAT. This overlap was used to label blazars in our datasets as TeV blazar (1) or non-TeV blazar (0).

6.4.3 Analysis of LR in Identifying TBCs

In this study, we applied SML methods to identify high-confidence TBCs. The LR models were evaluated based on their performance in classifying blazars from various catalogs. The primary condition for selection was a logistic regression performance of at least 80%. This criterion led to the identification of 40 high-confidence candidates from a pool of 150 candidates in the 4FGL-DR2 / 4LAC-DR2 catalogs. The main results of the LR approach are summarized as follows:

4FGL-DR2 / 4LAC-DR2 Candidates The logistic regression model for the 4FGL-DR2 / 4LAC-DR2 candidates is given by:

$$\text{logit}' = 4.808\Gamma + 2.809V_F + 3.889 \log f_7^{\text{ph}} - 3.34z + 0.857 \log \nu_p^s + 20.244, \quad (6.1)$$

where $\text{logit}' = 0$ corresponds to an ideal threshold probability (p_{thre}) of 40%. Based on the training set 3, the optimal LR model achieved an Area Under the Curve (AUC) of 97% on training and 96% on testing. The model identified 150 TBCs from 1459 blazars, with an AUC of 98%, a False Positive Rate (FPR) of 11%, and a True Positive Rate (TPR) of 99%. Only one TeV blazar was misclassified as a non-TBC.

3FHL Candidates The logistic regression model for the 3FHL candidates is described by:

$$\text{logit}' = 0.116 \log f_2^{\text{ph}} - 0.628z + 1.028, \quad (6.2)$$

with $\text{logit}' = 0$ being ideal for a p_{thre} of 52%. Based on the training set 5, the optimal LR model achieved an AUC of 89% on training and 87% on testing. This model identified 126 TBCs from 540 blazars, with an AUC of 88%, FPR of 27%, and TPR of 88%. Only eight TeV blazars were misclassified as non-TBCs.

3HSP Candidates The logistic regression model for the 3HSP candidates is given by:

$$\text{logit}' = 0.306FOM - 3.861z - 0.173, \quad (6.3)$$

where $\text{logit}' = 0$ is ideal for a p_{thre} of 61%. Based on the training set 5, the optimal LR model achieved an AUC of 99% on training and 94% on testing. It identified 40TBCs from 1771 blazars, with an AUC of 98%, FPR of 2%, and TPR of 91%. Only five TeV blazars were misclassified as non-TBCs.

2BIGB Candidates The logistic regression model for the 2BIGB candidates is:

$$\text{logit}' = -0.016E_p + 0.248FOM - 4.395z - 0.122 \quad (6.4)$$

with $\text{logit}' = 0$ being ideal for a p_{thre} of 48%. The optimal LR model, based on training set 4, achieved an AUC of 97% on both training and testing. This model identified 83 TBCs from 1040 blazars, with an AUC of 97%, FPR of 8%, and TPR of 96%. Only two TeV blazars were misclassified as non-TBCs.

We present a portion of our results through various figures:

- Fig. 6.1 illustrates the feature selection process.
- Fig. 6.2 displays the AUC for both training and testing datasets.
- Fig. 6.3 depicts the ROC curve, along with the AUC and the threshold probability (p_{thre}) for the prediction sets.
- Fig. 6.4 visualizes the linear boundary among 3FHL, 3HSP, and 2BIGB categories.
- Fig. 6.5 presents the confusion matrix.

6.5 Potential Targets for Ground-Based Cherenkov Detectors

Within the scope of this study, we have identified 40 high-confidence TBCs as potential targets for IACTs and EAS arrays throughout the year 2023. The criteria for determining the detectability of a source are twofold: (i) the source's flux density must exceed the sensitivity threshold of the detector at energy bands $\geq 1\text{TeV}$; (ii) the source must traverse the detectable sky region of the detectors.

6.5.1 SEDs Considering EBL Correction

Our approach involves initially correcting the data by accounting for the EBL and subsequently fitting the SEDs to the intrinsic data. The SED characteristically exhibits two distinct peaks: a low-energy bump spanning from the infrared to soft X-ray range, attributable to synchrotron emission, and a high-energy bump ranging from hard X-ray to γ -ray, associated with IC emission. We utilized an online

Table 6.1: Extraction of Features from Four Catalogs for Learning Sets.

Catalog (1)	Column (2)	Feature (3)	Unit (4)	Description (5)
4FGL-DR2 / 4LAC-DR2	1	Pivot_Energy (E_p)	GeV	Energy at which error on differential flux is minimal
	1	Flux ($\log F^{\text{ph}}$)	$\text{cm}^{-2} \cdot \text{s}^{-1}$	Integral photon flux from 1 to 100 GeV
	1	Energy_Flux ($\log F$)	$\text{erg} \cdot \text{cm}^{-2} \cdot \text{s}^{-1}$	Energy flux from 100 MeV to 100 GeV
	1	PL_Flux_Density ($\log f_1$)	$\text{cm}^{-2} \cdot \text{MeV}^{-1} \cdot \text{s}^{-1}$	Differential flux at Pivot_Energy in PowerLaw fit
	1	PL_Index (α_1)		Photon index when fitting with PowerLaw
	1	LP_Flux_Density ($\log f_2$)	$\text{cm}^{-2} \cdot \text{MeV}^{-1} \cdot \text{s}^{-1}$	Differential flux at Pivot_Energy in LogParabola fit
	1	LP_Index (α_2)		Photon index at Pivot_Energy when fitting with LogParabola
	1	LP_beta (β)		Curvature parameter when fitting with LogParabola
	1	PLEC_Flux_Density ($\log f_3$)	$\text{cm}^{-2} \cdot \text{MeV}^{-1} \cdot \text{s}^{-1}$	Differential flux at Pivot_Energy in PLSuperExpCutoff fit
	1	PLEC_Index (Γ)		Low-energy photon index when fitting with PLSuperExpCutoff
	1	PLEC_Expfactor (a)		Exponential factor when fitting with PLSuperExpCutoff
	1	PLEC_Exp_Index (b)		Exponential index when fitting with PLSuperExpCutoff
	7	Flux_Band ($\log f_1 \sim \log f_7$)	$\text{cm}^{-2} \cdot \text{s}^{-1}$	Integral photon flux in each spectral band
	1	nuFnu_Band ($\log f_1^{\text{ph}} \sim \log f_7^{\text{ph}}$)	$\text{erg} \cdot \text{cm}^{-2} \cdot \text{s}^{-1}$	Spectral energy distribution over each spectral band
	1	Variability_Index (V)		Likelihood difference between the flux fitted in each time interval and the average flux
	1	Frac_Variability (V_f)		Fractional variability computed from the fluxes in each year
	3FHL	1	Redshift (z)	
1		nu_syn ($\log \nu_p^s$)	Hz	Synchrotron-peak frequency in observer frame
1		nuFnu_syn ($\log \nu_p^s f_p^s$)	$\text{erg} \cdot \text{cm}^{-2} \cdot \text{s}^{-1}$	Spectral energy distribution at synchrotron-peak frequency
1		Highest_energy (E_{H})	GeV	Highest energy among events probably coming from the source
1		Pivot Energy (E_p)	GeV	Energy at which error on differential flux is minimal
1		Flux Density ($\log f$)	$\text{cm}^{-2} \cdot \text{GeV}^{-1} \cdot \text{s}^{-1}$	Differential flux at Pivot Energy
1		Flux ($\log F^{\text{ph}}$)	$\text{cm}^{-2} \cdot \text{s}^{-1}$	Integral photon flux from 10 GeV to 1 TeV $\text{erg} \cdot \text{cm}^{-2} \cdot \text{s}^{-1}$
1		Energy Flux ($\log F$)	$\text{erg} \cdot \text{cm}^{-2} \cdot \text{s}^{-1}$	Energy flux from 10 GeV to 1 TeV
1		PowerLaw Index (α_1)		Photon index when fitting with power law
1		Spectral Index (α_2)		Photon index at Pivot Energy when fitting with LogParabola
1		beta (β)		Curvature parameter when fitting with LogParabola
4		Flux Band ($\log f_1 \sim \log f_4$)	$\text{cm}^{-2} \cdot \text{s}^{-1}$	Integral photon flux in each spectral band
4		nuFnu_Band ($\log f_1^{\text{ph}} \sim \log f_4^{\text{ph}}$)	$\text{erg} \cdot \text{cm}^{-2} \cdot \text{s}^{-1}$	Spectral energy distribution over each spectral band
1		HEP energy (E_{H})	GeV	Highest energy among events probably coming from the source
1		Redshift (z)		Redshift
1		NuPeak obs ($\log \nu_p^s$)	Hz	Synchrotron-peak frequency in observer frame
3HSP		1	radio flux density ($\log f_R$)	mJy
	1	X-ray flux flux density ($\log f_X$)	μJy	X-ray flux density at 1keV
	1	nu_syn ($\log \nu_p^s$)	Hz	Synchrotron-peak frequency in observer frame
	1	Redshift (z)		Redshift
	1	FOM (FOM)		The figure of merit parameter, which is related to the likelihood of GeV / TeV detectability
2BIGB	1	N0 ($\log f$)	$\text{cm}^{-2} \cdot \text{MeV}^{-1} \cdot \text{s}^{-1}$	Differential flux at Pivot Energy in PowerLaw fit
	1	Gamma (α)		Photon index when fitting with PowerLaw
	1	$F_{0.5-500\text{GeV}}$ (F^{ph})		Integrated photon flux from 500 MeV to 500 GeV
	1	E0 (E_p)	GeV	Energy at which error on differential flux is minimal
	1	nu_syn ($\log \nu_p^s$)	Hz	Synchrotron-peak frequency in observer frame from 3HSP
1	Redshift (z)		Redshift from 3HSP	
1	FOM (FOM)		The same as in 3HSP	

Table 6.2: LR model and its performance on the 4FGL-DR2 / 4LAC-DR2 learning dataset

Partition (1)	AUC (2)	Overfitting (3)	p_{thre} (4)	Features (5)	Hyperparameters (6)
1	96.9% 95.9%	0.9%	53.6%	$E_p, \log F_1^{\text{ph}}, \log F_1, \Gamma, V, V_f, \log f_1^{\text{ph}}, \log f_2^{\text{ph}}, \log f_4^{\text{ph}}, \log f_6^{\text{ph}}, \log f_7^{\text{ph}}, z, \log \nu_p^s$	C: 1, max_iter: 500, penalty: l2, solver: sag, tol: 10^{-4}
2	97.2% 96.0%	1.2%	53.5%	$E_p, F_1, \alpha_1, \Gamma, V, V_f, f_7^{\text{ph}}, z, \log \nu_p^s$	C: 1, max_iter: 100, penalty: l1, solver: liblinear, tol: 10^{-6}
3	96.8% 96.0%	0.8%	53.4%	$\Gamma, V_f, \log f_7^{\text{ph}}, z, \log \nu_p^s$	C: 1, max_iter: 500, penalty: l1, solver: saga, tol: 10^{-6}
4	96.0% 98.0%	-1.8%	53.0%	$\log F_1, \alpha_1, \Gamma, a, V_f, \log f_7, \log F_7^{\text{ph}}, z, \log \nu_p^s$	C: 0.1, max_iter: 100, penalty: l2, solver: newton-cg, tol: 10^{-6}
5	97.5% 92.5%	5.0%	52.2%	$E_p, \alpha_1, \alpha_3, V_f, \log f_3, \log f_5, \log f_7, \log f_7^{\text{ph}}, z, \log \nu_p^s$	C: 1, max_iter: 100, penalty: l2, solver: newton-cg, tol: 10^{-6}

Table 6.3: LR model and its performance on the 3FHL learning dataset

Partition (1)	AUC (2)	Overfitting (3)	p_{thre} (4)	Features (5)	Hyperparameters (6)
1	87.8% 89.7%	-1.9%	53.6%	$\log f_2^{ph}, z$	C: 0.01, l1_ratio: 0.4, max_iter: 100, penalty: elasticnet, solver: saga, tol: 10^{-6}
2	88.0% 95.5%	-7.4%	53.5%	$\log f_2^{ph}, z$	C: 0.01, l1_ratio: 0.3, max_iter: 100, penalty: elasticnet, solver: saga, tol: 10^{-6}
3	88.7% 83.1%	5.5%	53.4%	$\log f_2^{ph}, z$	C: 0.01, l1_ratio: 0.4, max_iter: 100, penalty: elasticnet, solver: saga, tol: 10^{-6}
4	89.6% 86.0%	3.6%	52.9%	$\log f_2^{ph}, z$	C: 0.01, l1_ratio: 0.2, max_iter: 100, penalty: elasticnet, solver: saga, tol: 10^{-6}
5	89.2% 87.0%	2.3%	52.2%	$\log f_2^{ph}, z$	C: 0.01, l1_ratio: 0.4, max_iter: 100, penalty: elasticnet, solver: saga, tol: 10^{-6}

Table 6.4: LR model and its performance on the 3HSP learning dataset

Partition (1)	AUC (2)	Overfitting (3)	p_{thre} (4)	Features (5)	Hyperparameters (6)
1	98.3% 99.5%	-1.2%	46.9%	$\log f_X, \log v_p^s, FOM, z$	C: 0.01, l1_ratio: 0.7, max_iter: 100, penalty: elasticnet, solver: saga, tol: 10^{-6}
2	98.4% 98.9%	-0.5%	60.2%	FOM, z	C: 0.01, max_iter: 100, penalty: l2, solver: liblinear, tol: 10^{-6}
3	98.6% 99.1%	-0.5%	47.6%	FOM, z	C: 0.001, max_iter: 100, penalty: l2, solver: newton-cg, tol: 10^{-6}
4	98.4% 99.8%	-1.4%	58.2%	FOM, z	C: 0.1, max_iter: 100, penalty: l2, solver: newton-cg, tol: 10^{-6}
5	99.2% 93.5%	5.7%	60.8%	FOM, z	C: 0.01, max_iter: 100, penalty: l2, solver: liblinear, tol: 10^{-6}

Table 6.5: LR model and its performance on the 2BIGB learning dataset

Partition (1)	AUC (2)	Overfitting (3)	p_{thre} (4)	Features (5)	Hyper-parameters (6)
1	98.6% 98.4%	0.2%	44.7%	$\alpha, F, \log v_p^s, FOM, z$	C: 0.01, max_iter: 100, penalty: l2, solver: newton-cg, tol: 10^{-6}
2	97.7% 91.7%	6.0%	52.4%	E_p, FOM, z	C: 0.01, max_iter: 100, penalty: l2, solver: newton-cg, tol: 10^{-6}
3	98.2% 91.8%	6.3%	49.1%	FOM, z	C: 0.01, max_iter: 100, penalty: l2, solver: newton-cg, tol: 10^{-6}
4	97.2% 97.1%	0.1%	48.1%	E_p, FOM, z	C: 0.01, l1_ratio: 0.1, max_iter: 100, penalty: elasticnet, solver: saga, tol: 10^{-6}
5	96.7% 99.7%	-3.0%	50.3%	FOM, z	C: 0.01, max_iter: 100, penalty: l2, solver: newton-cg, tol: 10^{-6}

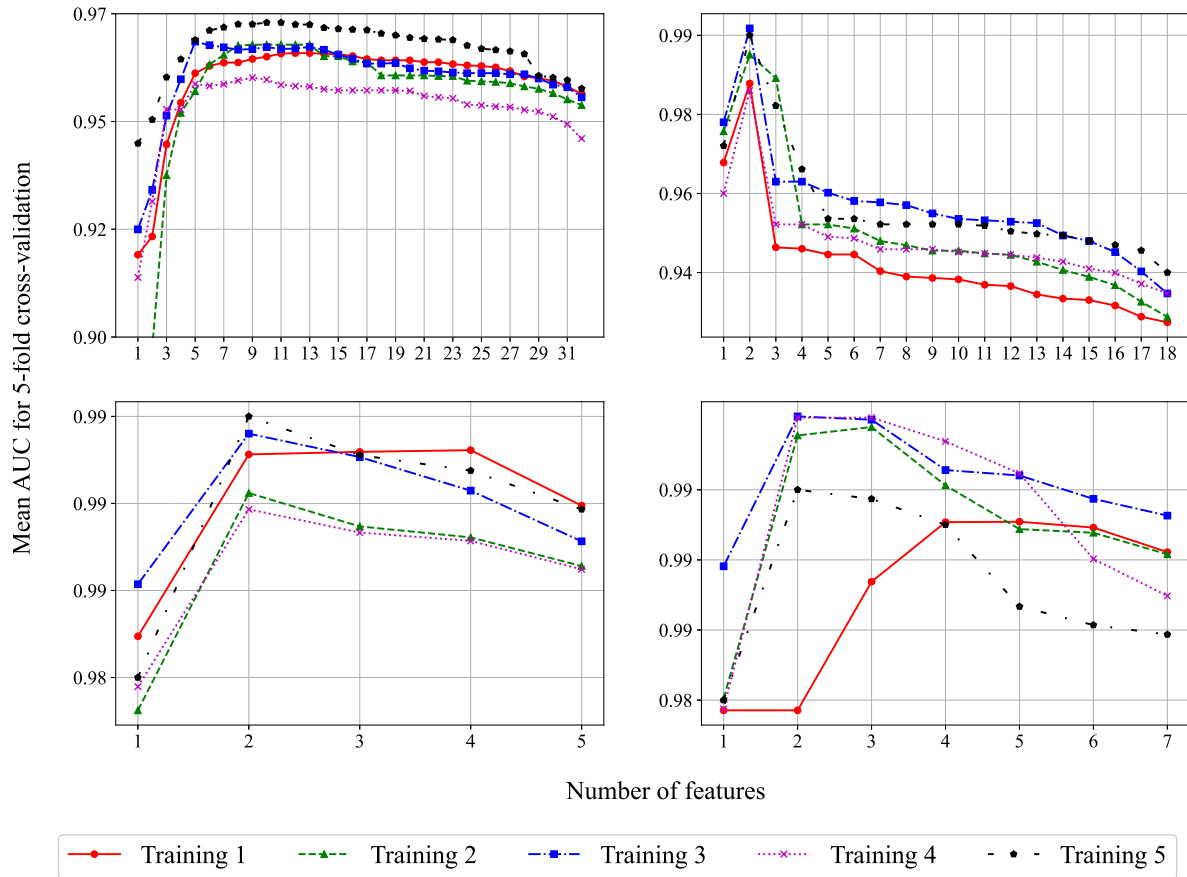


Figure 6.1: SFS analysis for Four Learning Sets. The graphs are structured with the number of parameters on the X-axis and the concentration of AUC on the Y-axis. Each learning set is represented in a separate panel, with five dotted lines in various colors indicating the training iterations from Training 1 to Training 5. *Top Left Panel:* This panel illustrates the results for the 4FGL-DR2 / 4LAC-DR2 data set; *Top Right Panel:* This panel shows the SFS analysis for the 3FHL data set; *Bottom Left Panel:* In this panel, the results for the 3HSP data set are displayed; *Bottom Right Panel:* This panel presents the analysis for the 2BIGB data set.

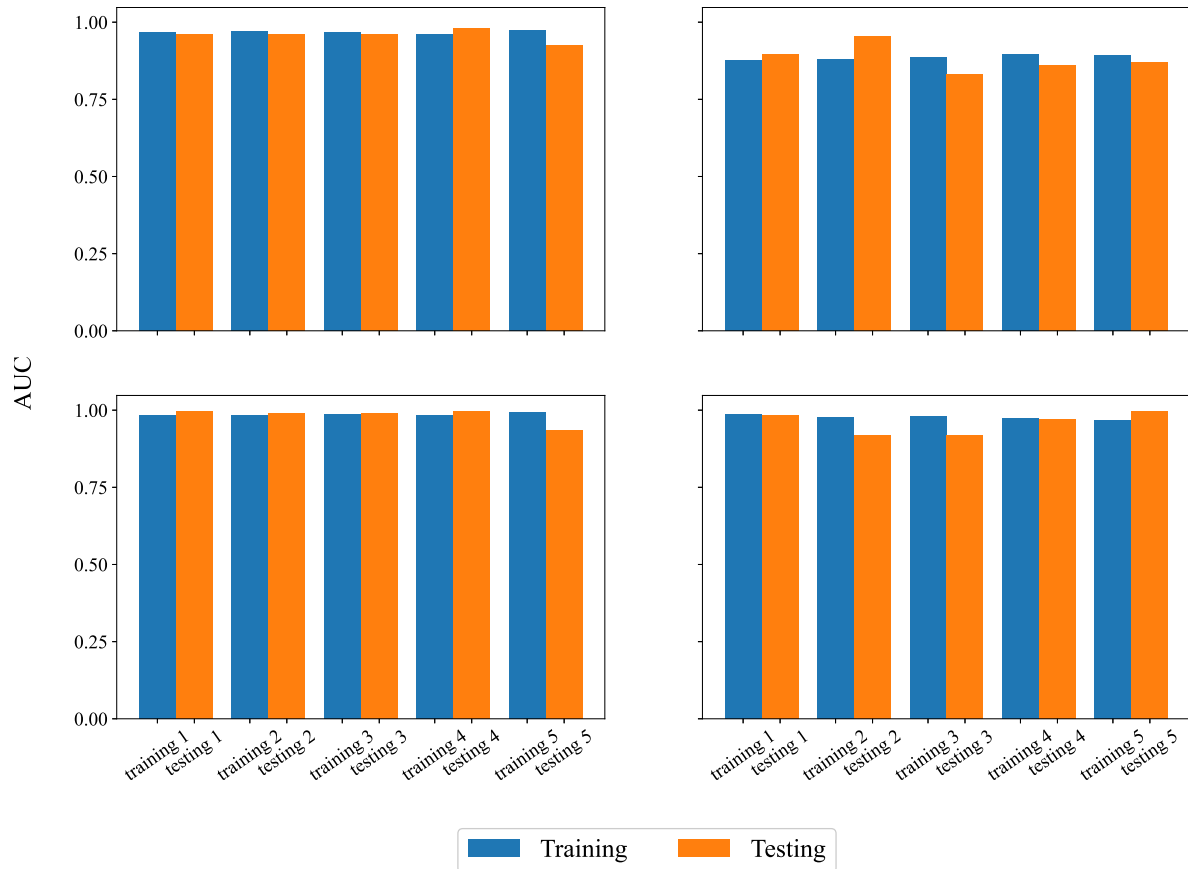


Figure 6.2: The AUC values for four distinct learning sets are presented, each characterized by blue histograms for training sets and orange histograms for testing sets. *Top Left Panel:* 4FGL-DR2 / 4LAC-DR2 - This panel likely represents data or results associated with the 4FGL-DR2 and 4LAC-DR2 datasets; *Top Right Panel:* 3FHL - This panel is dedicated to the 3FHL dataset; *Bottom Left Panel:* 3HSP - This panel focuses on the 3HSP dataset; *Bottom Right Panel:* 2BIGB - This panel displays information related to the 2BIGB dataset.

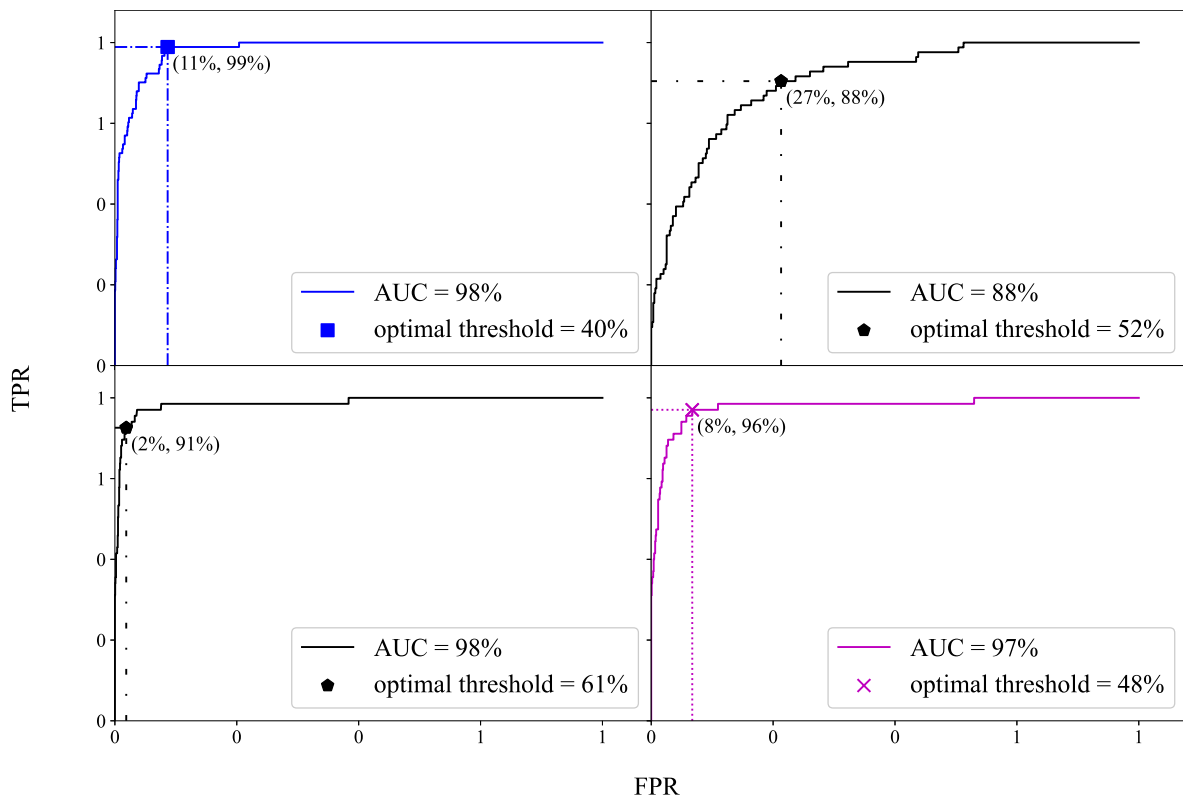


Figure 6.3: The analysis involves the ROC curve for four distinct prediction datasets, arranged as follows: The 4FGL-DR2 and 4LAC-DR2 datasets are displayed in the *top left panel*; The 3FHL dataset is shown in the *top right panel*; The 3HSP dataset is featured in the *bottom left panel*; The 2BIGB dataset is presented in the *bottom right panel*.

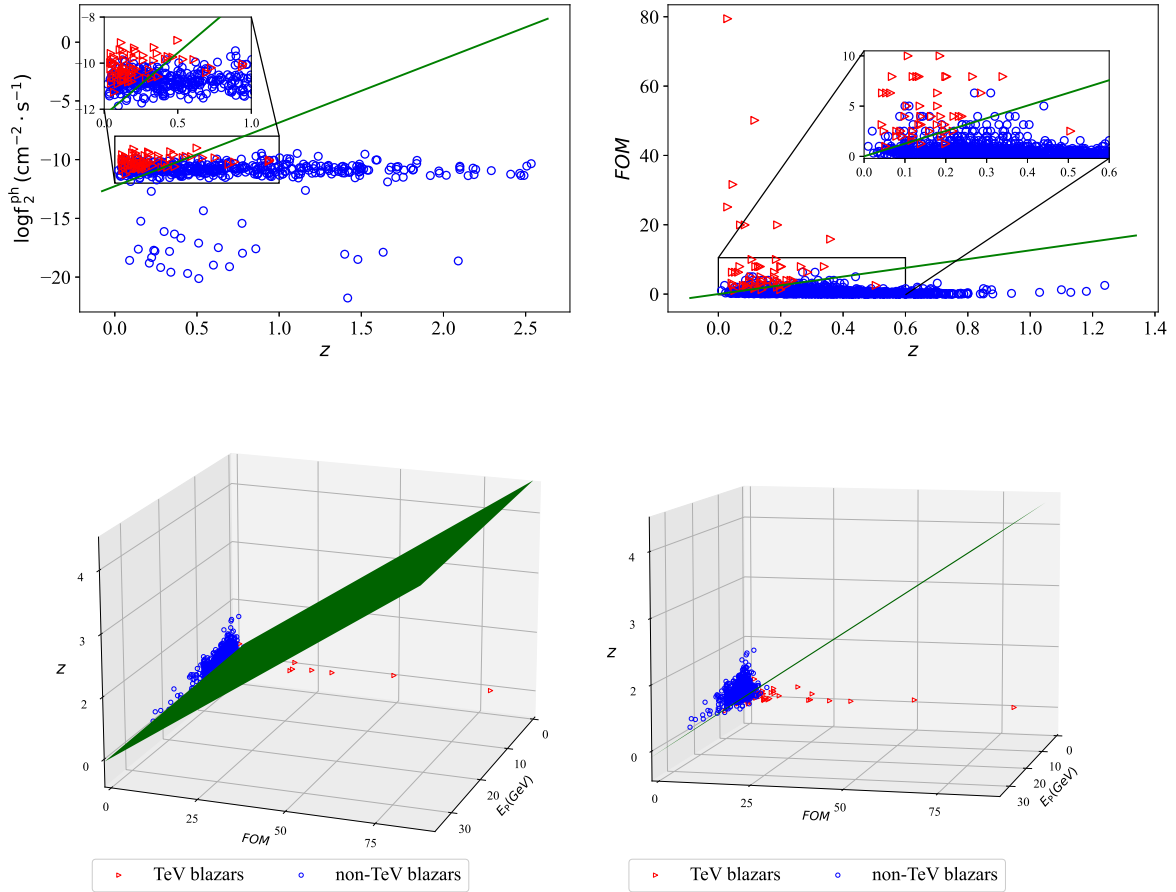


Figure 6.4: Describe the classification and prediction data sets for 3FHL, 3HSP, and 2BIGB. In these data sets, red triangles represent TeV blazars, blue circles indicate non-TeV blazars and the green line or plane is the classification boundary. The arrangement is as follows: the *top left panel* shows 3FHL, the *top right panel* displays 3HSP, the *bottom left panel* presents 2BIGB from the perspective of non-TeV blazars, and the *bottom right panel* illustrates 2BIGB as viewed along the edge of the green classification plane.

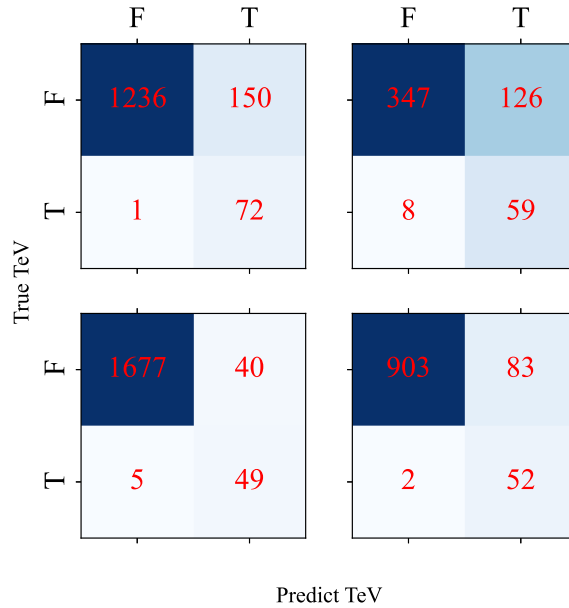


Figure 6.5: For each of the four learning sets, the confusion matrix is presented with the following four panels. *Top Left Panel:* 4FGL-DR2 / 4LAC-DR2; *Top Right Panel:* 3FHL; *Bottom Left Panel:* 3HSP; *Bottom Right Panel:* 2BIGB. In the context of coordinate values, ‘T’ represents ‘True’, and ‘F’ signifies ‘False’.

tool⁹ provided by the Space Science Data Center (SSDC) of the Italian Space Agency to aggregate data across a spectrum extending from radio to TeV frequencies, incorporating data from various missions, experiments, catalogs, and archival sources (Aharonian et al., 2001; Giommi et al., 2002; Amenomori et al., 2003; Aharonian et al., 2003, 2009b; Daniel et al., 2005; Schroedter et al., 2005; Acciari et al., 2008, 2011; Godambe et al., 2008; Chandra et al., 2010, 2012; H. E. S. S. Collaboration et al., 2010; HESS Collaboration et al., 2013; Aliu et al., 2011, 2015; Bartoli et al., 2011, 2012; Archambault et al., 2013, 2014; Arlen et al., 2013; Abramowski et al., 2013, 2015; Biteau & Williams, 2015; Sharma et al., 2015).

In the SED fitting process, we exclude data bins where the flux error exceeds the flux upper limits. Notably, TeV photon data from TeV candidates, which exhibit zero error, are also omitted since these photons are not verified by IACTs or EAS arrays (refer to Fig. 6.6). Additionally, data in the lower radio energy spectrum, specifically below $\log \nu (\text{Hz}) < 9$, are disregarded due to observed asymmetries in the synchrotron bump of some sources.

This study does not consider the simultaneity of data observations. The interaction of Extragalactic Background Light (EBL) photons with Very High Energy (VHE) photons through pair production leads to the attenuation of observed gamma-ray spectra from blazars at cosmological distances. This interaction significantly affects the SED in the Inverse Compton (IC) bump.

Assuming that the EBL optical depth ($\tau(E, z)$) varies with photon energy (E) and redshift (z), Saldana-Lopez et al. (2021) provided a two-dimensional grid of $\tau(E, z)$, ranging from 0.001 TeV to 100 TeV for energy, and from 0 to 6 for redshift¹⁰.

Following this, the dual bumps observed in the SED were independently fitted using a log-

⁹<http://tools.asdc.asi.it/SED/>

¹⁰<https://www.ucm.es/blazars/eb1>

parabolic model as outlined by Fan et al. (2016):

$$\log(\nu f_\nu) = c(\log \nu - \log \nu_p)^2 + \log \nu_p f_{\nu_p}, \quad (6.5)$$

where $|2c|$ represents the spectral curvature, $\log \nu_p$ denotes the logarithm of the peak frequency, $\log \nu_p f_{\nu_p}$ signifies the logarithm of the peak flux, and $\log(\nu f_\nu)$ corresponds to the flux density.

The synchrotron bump was successfully fitted for all 40 TBCs, while the IC bump was fitted for 20 candidates. The fitting errors for $|2c|$, $\log \nu_p$ (in units of Hz), and $\log \nu_p f_{\nu_p}$ (in units of $\text{erg cm}^{-2} \text{s}^{-1}$) were calculated. The fitting process was conducted using the `scipy.optimize.curve_fit` function in Python. The results are presented in Tab. 6.6, where Col. (1) lists the 4FGL names, Cols. (2) to (4) detail the parameters and errors for the synchrotron bump fitting and Cols. (5) to (7) provide the parameters and errors for the IC bump fitting. The right side of the slash represents data unabsorbed by EBL. The SED fitting results for each high-confidence TBC are illustrated in Fig. 6.6.

Table 6.6: Results of SED fitting in the observed frame for candidates with an energy of 40 TeV.*

4FGL name (1)	$ c^s $ (2)	$\log \nu_p^s$ (3)	$\log \nu_p^s f_{\nu_p}^s$ (4)	$ c^{\text{IC}} $ (5)	$\log \nu_p^{\text{IC}}$ (6)	$\log \nu_p^{\text{IC}} f_{\nu_p}^{\text{IC}}$ (7)
J0037.8+1239	0.11 ± 0.02	15.35 ± 0.46	-11.17 ± 0.07	0.02 ± 0.02	23.15 ± 2.42	-11.74 ± 0.1
J0051.2-6242	0.16 ± 0.01	15.72 ± 0.06	-11.23 ± 0.03	0.08 ± 0.07	25.57 ± 1.2	-11.37 ± 0.13
J0110.1+6805	0.1 ± 0.01	14.99 ± 0.29	-10.89 ± 0.09			
J0110.7-1254	0.05 ± 0.01	17.76 ± 1.09	-11.81 ± 0.22			
J0115.8+2519	0.09 ± 0.01	15.67 ± 0.26	-11.54 ± 0.04			
...
...
...

The full details of this table can be found in the comprehensive listing provided in Zhu et al. (2023).

6.5.2 Visibility Analysis of TBCs for Ground-Based Telescopes

Ground-based telescopes, unlike space detectors with all-sky scanning capabilities, are limited in their sky coverage due to geographical constraints. IACTs require dark skies for observations and have a relatively narrow field of view (FoV) of 3° to 5° . Additionally, their operational time is limited to clear, moonless nights, and their sensitivity decreases with increasing zenith angle, which also raises the energy threshold. Consequently, this restricts the sky area available for observation.

The CTAO, representing the next generation of IACTs, comprises two arrays situated in the Northern ($28^\circ 45' \text{ N}$) and Southern ($24^\circ 41' \text{ S}$) Hemispheres, aiming to cover a larger portion of the sky. However, like current IACTs, CTAO's observations are less effective at high zenith angles. In contrast, EAS arrays, such as LHAASO and HAWC, operate under various weather conditions and have a large FoV (approximately 2 sr), enabling them to continuously monitor a substantial fraction of the sky each day.

We have compiled sensitivity curves for various IACTs and EAS arrays from sources including van Eldik et al. (2015); Aleksić et al. (2016); Cao et al. (2019); Abeysekara et al. (2017c); Abeysekara et al. (2017) and online databases for CTAO¹¹, H.E.S.S.¹², and VERITAS¹³. The FoV parameters for the sensitivity curves of various instruments are specified as follows: CTAO North (north site, $0^\circ \leq Z \leq 20^\circ$), CTAO South (south site, $0^\circ \leq Z \leq 20^\circ$), H.E.S.S. at zenith ($Z \approx 0^\circ$), H.E.S.S. at

¹¹<https://www.cta-observatory.org/science/ctao-performance/>

¹²chrome-extension://efaidnbmninnipcbajpcgiclfndmkaj/https://www.mpi-hd.mpg.de/hfm/HESS/pages/home/proposals/sc_sens.pdf

¹³<https://veritas.sao.arizona.edu/about-veritas/veritas-specifications>.

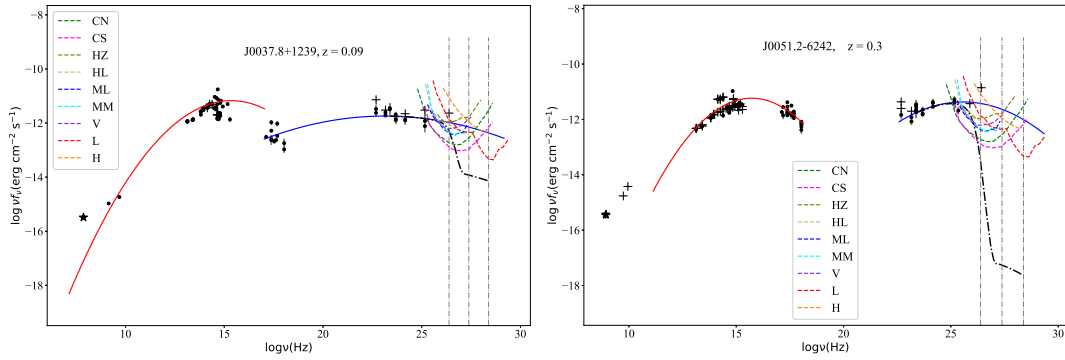


Figure 6.6: SEDs for candidates with energies around 40 TeV are presented, with only two items displayed for brevity, the complete version can be found at (Zhu et al., 2023). A comprehensive version of this figure is accessible in the online publication. Data points are denoted by grey error bars. It is important to note that the photons in the TeV band have not been certified by IACTs and EAS arrays. The optimal-fitting synchrotron and IC bumps are represented by red and solid blue curves, respectively, while the black dash-dot curves illustrate the IC bumps absorbed by EBL. Three vertical dash-dot lines correspond to $\log \nu$ (Hz) at 1 TeV, 10 TeV, and 100 TeV. The sensitivities of various observatories are indicated by different colored dotted curves: green for CTAO north (north site, $0^\circ \leq Z \leq 20^\circ$), magenta for CTAO south (south site, $0^\circ \leq Z \leq 20^\circ$), olive for H.E.S.S. zenith ($0^\circ \leq Z \leq 5^\circ$), tan for H.E.S.S. low ($12^\circ \leq Z \leq 22^\circ$), blue for MAGIC low ($0^\circ \leq Z \leq 30^\circ$), cyan for MAGIC medium ($30^\circ \leq Z \leq 45^\circ$), blue-violet for VERITAS ($0^\circ \leq Z \leq 20^\circ$), red for LHAASO, and dark-orange for HAWC. Additionally, the redshifts (z) for each source are also provided. Furthermore, data points with flux errors exceeding flux upper limits are marked with pentagons, and crosses represent data in the low radio energy range ($\log \nu (\text{Hz}) < 9$).

medium zenith angles ($12^\circ \leq Z \leq 22^\circ$), MAGIC at low zenith angles ($0^\circ \leq Z \leq 30^\circ$), MAGIC at medium zenith angles ($30^\circ \leq Z \leq 45^\circ$), VERITAS ($0^\circ \leq Z \leq 20^\circ$), LHAASO ($-11^\circ \leq Dec \leq 69^\circ$), and HAWC ($-20^\circ \leq Dec \leq 60^\circ$). Regarding exposure times, HAWC is observed over 507 days, LHAASO over one year, H.E.S.S. near the zenith for 25 hours, and the remaining instruments for 50 hours each.

To assess the visibility of a given source, we use two different approaches for EAS arrays and IACTs. For EAS arrays, a source is considered observable if its declination falls within the most observable sky region. For IACTs, we determine the observable months during dark skies using the H.E.S.S. online visibility tool¹⁴ and the `astroplan` Python package. These tools provide visibility information based on the source's celestial coordinates and the IACTs' observation parameters, as illustrated in Fig. 6.7 and Tab. 6.7.

The observable months for 40 high-confidence TBCs, determined using these tools, are listed in Tab. 6.7. The table includes the source name and observable months for IACTs and EAS arrays, with months represented numerically (e.g., 1' for January, 2' for February, etc.).

6.5.3 High-confidence TBCs

We analyzed 40 potential TeV blazar candidates, examining their energy (E) range from 0.001 TeV to 100 TeV to determine their $\tau(E, z)$ values and the impact of EBL on their IC bumps. This analysis led to the identification of 7 high-confidence TeV blazar candidates based on their detectability criteria in comparison with the sensitivity curves of IACTs and EAS arrays within the 1 TeV to 100 TeV energy range. The detection capabilities of various TeV facilities for these candidates are detailed in Tab. 6.8, including 6 targets for both CTAO North and South, 2 for H.E.S.S., 5 for MAGIC, 1 for VERITAS, 1 for LHAASO, and none for HAWC.

The specifics of the 40 high-confidence TeV blazar candidates are presented in Tab. 6.8, organized as follows:

Col. (1) the name of the source in the 4FGL catalog. Col. (2) and Col. (3) Galactic Coordinates in degrees. Col. (4) synchrotron-Peak Frequency ($\log \nu_p^s$). Col. (5) redshift (z). Col. (6) flux Density at 1 TeV ($f_{1:\text{TeV}}$): Calculated using the IC bump parameters and Formula. (6.5). Col. (7) to Col. (10) logistic likelihood in 4 Catalogs: The probability of a source being a TeV candidate based on its features. Col. (11) SED class in 4FGL-DR2 / 4LAC-DR2: Classifications include BL Lac (bll), FSRQ (fsrq), and blazar candidate of uncertain type (bcu). Comparison with 3FHL, 3HSP, and 2BIGB Catalogs: Indicated with 'Y' for candidates in common. Col. (12) to Col. (14) comparison with Other Literature: Including works by Costamante & Ghisellini (2002), Massaro et al. (2013), Foffano et al. (2019), and Chiaro et al. (2019). Col. (15) detectability in TeV Range (1 TeV to 100 TeV): Indicated for various IACTs and EAS arrays. Additionally, we sought to determine the maximum redshift at which these sources remain detectable in the TeV energy band. This involved comparing the EBL-absorbed IC bumps of 20 blazars with the sensitivity curves of detectors at 1 TeV, 10 TeV, and 100 TeV. The redshift upper limits for each IACT and EAS array are listed in Table 6.9, which includes: Col. (1) 4FGL Name. Col. (2) to Col. (10) The redshift Upper Limits for Each Detector: Presented at 1 TeV, 10 TeV, and 100 TeV. A value of '0' indicates non-detectability at the specified energy. Col. (11) actual Redshifts. Col. (12) detectability for Cherenkov Detectors. This analysis demonstrates that higher redshifts generally reduce the likelihood of a source being a TeV candidate, as indicated by the inverse relationship between logistic likelihood and redshift in Formulas (6.1) to (6.4).

¹⁴<https://www.mpi-hd.mpg.de/hfm/HESS/pages/home/visibility/>

Table 6.7: Observation periods for IACTs and EAS arrays.*

4FGL name (1)	CTAO North (2)	CTAO South (3)	H.E.S.S. zenith (4)	H.E.S.S. low (5)	MAGIC low (6)	MAGIC medium (7)	VERITAS (8)
J0037.8+1239	7, 8, ..., 12				1, 7, ..., 12	1, 6, ..., 12	7, 8, ..., 12
J0051.2-6242							
J0110.1+6805						1, 7, ..., 12	
J0110.7-1254		7, 8, ..., 12		6, 7, ..., 12		1, 7, ..., 12	
J0115.8+2519	1, 7, ..., 12				1, 7, ..., 12	1, 2, 6, ..., 12	1, 7, ..., 12
...
...
...

The full details of this table can be found in the comprehensive listing provided in Zhu et al. (2023).

‘h’ signifies that the month is exclusively observable using the H.E.S.S. instrument.

‘a’ signifies that the month is exclusively visible in the astroplan context.

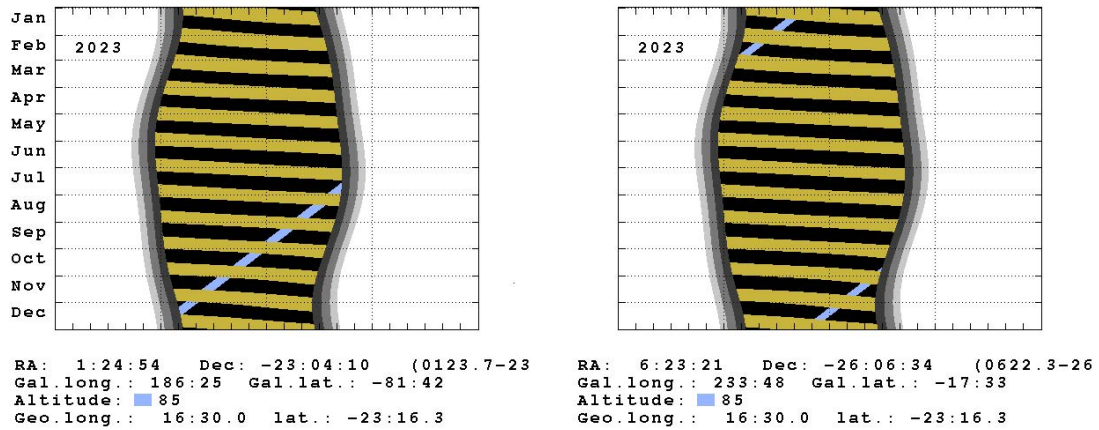


Figure 6.7: In 2023, four sources including 4FGL J0123.7-2311 and 4FGL J0622.3-2605 will be observable by H.E.S.S. near its zenith. Visibility details for other IACTs and EAS arrays are in the online supplementary data. The H.E.S.S. online tool generated these visibility charts, where white indicates daylight, grey shades represent different twilights, yellow marks moon presence or twilight, and blue shows when the object is above certain altitudes. The RA/Dec coordinates are for the current epoch, not the J2000 standard.

Table 6.8: Forecasting outcomes for 40 potential TeV blazar candidates*

4FGL name (1)	GLON (2)	GLAT (3)	$\log y_p^b$ (4)	z (5)	$\log f_{TeV}$ (6)	<i>logistic</i>				<i>Other catalogs</i>				Common (15)	Detectability (16)	
						4FGL (7)	3FHL (8)	3HSP (9)	2BIGB (10)	Class (11)	3FHL (12)	3HSP (13)	2BIGB (14)			
J0037.8+1239	117.77	-50.08	15.05	0.09	-12.32	82.50%	54.60%	CN, CS, ML, MM
J0051.2-6242	302.96	-54.43	15.96	0.3	-12.91	84.10%
J0110.1+6805	124.69	5.29	14.85	0.29	...	81.80%
J0110.7-1254	141.53	-75.08	17	0.23	...	82.40%
J0115.8+2519	129.85	-37.21	15.75	0.36	...	97.60%
...
...
...

The full details of this table can be found in the comprehensive listing provided in Zhu et al. (2023), where all items are thoroughly documented. The source is detectable within the energy ranges of 1 TeV to 10 TeV and 10 TeV to 100 TeV.

Table 6.9: Upper redshift boundaries for the 20 blazars as determined for each IACT and EAS array.*

4FGL name (1)	z_{uplimit} of CN (2)	z_{uplimit} of CS (3)	z_{uplimit} of HZ (4)	z_{uplimit} of HL (5)	z_{uplimit} of ML (6)	z_{uplimit} of MM (7)	z_{uplimit} of V (8)	z_{uplimit} of L (9)	z_{uplimit} of H (10)	z (11)	Detectability (12)
J0037.8+1239	...	0.2, 0.05, 0.01	...	0.02, 0.01, 0.01	...	0.09, 0.01, 0.01	...	0, 0.01, 0.01	0, 0, 0.01	0.09	CN, CS, ML, MM
J0051.2-6242	...	0.29, 0.07, 0.01	...	0.14, 0.03, 0.01	0.20, 0.02, 0.01	0.19, 0.04, 0.01	0.17, 0.01, 0.01	0.1, 0.04, 0.01	0, 0.02, 0.01	0.3	
J0110.1+6805	0.29	
J0110.7-1254	0.23	
J0115.8+2519	0.36	
...
...
...

The full details of this table can be found in the comprehensive listing provided in Zhu et al. (2023), where all items are thoroughly documented.

6.6 Discussions and conclusions

6.6.1 Discussions

The SFS process identified five key features in the 4FGL-DR2 / 4LAC-DR2 datasets that are crucial for differentiating TeV from non-TeV blazars. These features include flux density, FOM, synchrotron peak frequency, redshift, spectral index, and variability. This aligns with the findings of Lin & Fan (2016). Additionally, the potential influence of X-ray emissions and shifts in the synchrotron peak frequency during flare events warrants further investigation, as blazars not typically detected at TeV energies might become detectable during such flaring states.

This suggests the possibility that all blazars are capable of emitting TeV photons, with the detection of TeV emissions being contingent on the flux density in the TeV band being sufficiently high, particularly during flaring periods. Notably, the 4FGL-DR2 dataset includes features like *VariabilityIndex* and *FracVariability*, which reflect flux variability over time and are integral to our LR model. This ensures that our LR model is sensitive to sources experiencing shifts in their synchrotron peak during flaring states.

Regarding the classification of blazars, according to Abdo et al. (2010b) or Fan et al. (2016), sources with $\log \nu_p^s > 15$ (or $\log \nu_p^s > 15.3$) are categorized as High Synchrotron Peaked (HSP) blazars. However, by focusing solely on HSP/HSP BL Lacs, we would overlook over 11% of TeV sources. Our approach, which does not impose stringent initial filters based on SED class or other properties, offers an alternative method.

To evaluate the effectiveness of two criteria – HSPs ($\log \nu_p^s \geq 15.3$ Hz as per Fan et al. (2016)) and the LR model – we analyzed metrics and confusion matrices on the testing sets of 4FGL-DR2 / 4LAC-DR2 and 3FHL. For 4FGL-DR2 / 4LAC-DR2, the AUC of 73% and TPR of 62% are both lower than those predicted by the LR model, resulting in up to 27 TeV blazars being misclassified as non-TeV. Similarly, for 3FHL, excluding four sources without $\log \nu_p^s$ data, the AUC of 68% and TPR of 63% are also lower than those of the LR model, leading to the misclassification of up to 25 TeV blazars as non-TeV. The confusion matrices for these comparisons are illustrated in Fig. 6.8 and Fig. 6.9. These results demonstrate the superior effectiveness of our selection criterion.

In our study, we shifted from the traditional approach of selecting TeV candidates based on their high predicted TeV fluxes. Instead, we focused on identifying targets that, according to their broadband features, are likely to become promising TeV candidates in specific activity states. These targets might have been overlooked due to their current status or the absence of suitable instruments and observation strategies. Our method relies heavily on ML, which is inherently influenced by factors such as the composition of the sample, the ratio of the training set to the test set, and the value of k in cross-validation. Our goal is to ensure the generalizability of our results based on the available data.

However, there is a potential bias in our approach. A significant number of sources in TeVCat, a catalog of TeV sources, are blazars detected by Fermi-LAT. This is because Fermi-LAT catalogs are often used to identify potential candidates for Imaging Atmospheric Cherenkov Telescopes (IACTs), and Fermi-LAT monitoring triggers Target of Opportunity programs. Additionally, our observation proposal does not take into account factors like weather conditions or the availability of detectors. A more practical observation proposal would require collaboration with observatories and consider a broader range of influencing factors.

Detecting extragalactic photons in the TeV band is challenging due to two major limitations: the Extragalactic Background Light (EBL) and the performance of detectors (sensitivity, effective area, field of view, etc.). The relationship between improved detector sensitivity and the detection of more TeV sources is complex. While better sensitivity does aid in detecting more sources, the attenuation of

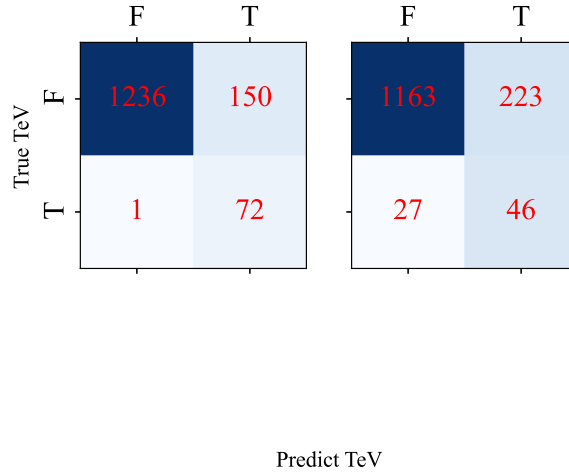


Figure 6.8: Two confusion matrices for SML and HSPs. The *left panel* uses data from the 4FGL-DR2 / 4LAD-DR2 catalog with five FOFs: Γ , V_F , $\log f_7^{\text{ph}}$, z , $\log \nu_p^s$, while the *right panel* focuses on sources with $\log \nu_p^s \geq 15.3$ Hz. Both matrices display true positives (TP), false negatives (FN), false positives (FP), and true negatives (TN) in their respective quadrants.

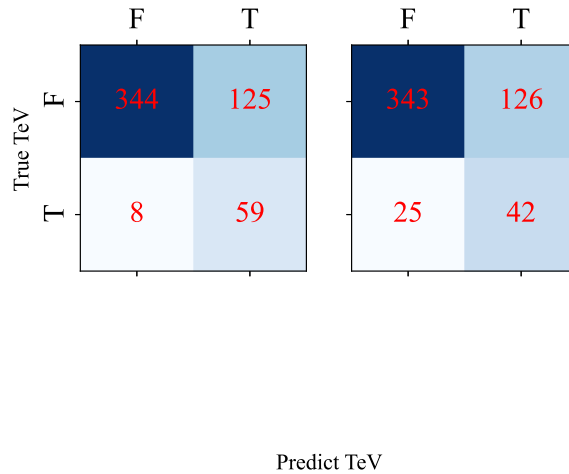


Figure 6.9: Two confusion matrices for SML and HSPs. The *left panel* uses data from the 3FHL catalog with five FOFs: $\log f_2^{\text{ph}}$, z , while the *right panel* focuses on sources with $\log \nu_p^s \geq 15.3$ Hz. Both matrices display true positives (TP), false negatives (FN), false positives (FP), and true negatives (TN) in their respective quadrants.

signals due to EBL increases with energy and redshift, leading to a modest overall effect (as discussed in Gilmore et al. 2012).

For the 81 blazars listed in TeVCat, we obtained their redshifts from four catalogs and other literature. Notably, for 4FGL J2243.9+2021 (or RGB J2243+203), we used the upper limit of the redshift range ($0.75 \leq z \leq 1.1$) as reported by Sahu et al. (2019). We then calculated the optical depth $\tau(E, z)$ at 1 TeV for each blazar and evaluated $\tau(E, z)$ for all 81 blazars. These results are detailed in Tab. 6.10, where Col. 1) and Col. 2) list the TeVCat and 4FGL-DR2 names, Columns (3) and (4) provide the celestial coordinates, Col. 5) shows the redshift, and Col. 6) gives the $\tau(E, z)$ values at 1 TeV, based on the model by Saldana-Lopez et al. (2021).

Table 6.10: Blazars at 81 TeV exhibit $\tau(E, z)$ when E equals 1 TeV.*

TeVcat Name (1)	4FGL Name (2)	Ra (3)	Dec (4)	z (5)	τ (6)
J0013-188	J0013.9-1854	3.47	-18.89	0.095	0.92
J0033-193	J0033.5-1921	8.4	-19.35	0.61	7.61
J0035+598	J0035.9+5950	8.82	59.79	0.467	5.7
J0112+227	J0112.1+2245	18.02	22.74	0.265	2.96
J0136+391	J0136.5+3906	24.14	39.1	0.75	9.35
...
...
...

The full details of this table can be found in the comprehensive listing provided in Zhu et al. (2023).

The validation of TeV candidates heavily relies on both IACTs and EAS arrays, each possessing its own set of strengths and limitations. IACTs excel in terms of energy and angular resolution, and they can detect energies in the GeV range. In contrast, EAS arrays typically become effective at energies above approximately 10 TeV. However, at around 30 TeV, the sensitivity of the LHAASO rivals that of the CTAO, and LHAASO demonstrates superior sensitivity at around 100 TeV.

EAS arrays also offer advantages in terms of observation time and FoV. Unlike IACTs, which are limited to moonless nights, EAS arrays can operate continuously, resulting in significantly longer exposure times. For example, while IACTs typically report exposure times of around 50 hours, EAS arrays can accumulate data over periods spanning 1 to 5 years. Although IACTs are more adept at capturing transient events, the extensive FoVs and prolonged observation periods of EAS arrays help bridge this gap.

The complementary nature of IACTs and EAS arrays is further highlighted by their distinct operational modes and performance characteristics. Furthermore, advancements in these technologies, particularly with LHAASO and the next-generation CTAO, have improved sensitivity by an order of magnitude compared to existing systems. These developments have extended the detectable range in the TeV band to beyond 100 TeV, paving the way for a more comprehensive exploration of the TeV sky in the future.

6.6.2 Conclusions

This study utilized a machine learning approach, specifically Logistic Regression, to identify potential TeV blazar candidates from various catalogs including 4FGL-DR2 / 4LAC-DR2, 3FHL, 3HSP, and

2BIGB. Additionally, this method aided in selecting prospective observational targets for Imaging Atmospheric Cherenkov Telescopes (IACTs) and Extensive Air Shower (EAS) arrays. Key findings are summarized as follows:

1. Application of Logistic Regression across the four catalogs yielded a formulaic approach for pinpointing blazars likely to be observable at TeV energies, with a cutoff criterion of $\text{logit}' \geq 0$. For 4FGL-DR2 / 4LAC-DR2:

$$\text{logit}' = 4.808\Gamma + 2.809V_F + 3.889 \log f_7^{ph} - 3.34z + 0.857 \log \nu_p^s + 20.244,$$

showcasing an AUC of 98%, FPR of 11%, TPR of 99%, and a threshold probability (p_{thre}) of 40%. Remarkably, 40 out of 150 non-TeV blazars were predicted with high confidence ($\geq 80\%$) as TeV candidates.

For 3FHL:

$$\text{logit}' = 0.116 \log f_2^{ph} - 0.628z + 1.028,$$

with metrics indicating an AUC of 88%, FPR of 27%, TPR of 88%, and a p_{thre} of 52%. 24 TeV candidates were overlapping with the high-confidence group from 4FGL-DR2 / 4LAC-DR2.

For 3HSP:

$$\text{logit}' = 0.306FOM - 3.861z - 0.173,$$

achieving an AUC of 98%, FPR of 2%, TPR of 91%, and a p_{thre} of 61%. Eleven TeV candidates were common with the high-confidence ones from 4FGL-DR2 / 4LAC-DR2.

For 2BIGB:

$$\text{logit}' = -0.016E_p + 0.248FOM - 4.395z - 0.122,$$

demonstrating an AUC of 97%, FPR of 8%, TPR of 96%, and a p_{thre} of 48%. Fourteen TeV candidates were shared with the high-confidence ones from 4FGL-DR2 / 4LAC-DR2.

2. Among the 40 high-confidence TeV blazar candidates, we performed independent fitting of the Spectral Energy Distributions (SEDs) and evaluated their visibility for EAS arrays and IACTs in 2023. This resulted in identifying 7 prime targets: 6 for CTAO, 2 for H.E.S.S., 5 for MAGIC, 1 for VERITAS, 1 for LHAASO, and none for HAWC.
3. Comparative analysis with previous studies revealed commonalities and distinct findings: 2 sources overlapped with Costamante & Ghisellini (2002), 9 with Massaro et al. (2013), 4 with Foffano et al. (2019), and none with Chiaro et al. (2019).

Chapter 7

Chasing the neutrino candidates of *Fermi* blazars

7.1 Introduction

In the realm of astrophysics, the study of neutrinos, first hypothesized by Wolfgang Pauli in 1930 during his work on β decay, holds paramount importance. The continuous energy spectrum of electrons produced in this decay hinted at the existence of a third, neutral, and nearly massless particle, later termed the *neutrino* by Enrico Fermi. Investigating high-energy neutrinos is pivotal in understanding the origins and acceleration mechanisms of high-energy cosmic rays. This is largely due to the isospin symmetry between u and d quarks, leading to a similar symmetry between protons and neutrinos in production.

Blazars are known for their ability to produce high-energy neutrinos, a process predominantly governed by hadronic interactions within their relativistic jets. These jets, propelled by the accretion disks surrounding supermassive black holes, are capable of accelerating particles to extremely high energies. The primary mechanisms for neutrino production in blazars are as follows:

1. Proton-Photon Interactions (p γ interactions)

In this process, protons accelerated to high energies within the jet interact with photons (originating from the accretion disk, the jet itself, or external light sources). These interactions lead to the production of charged pions (π^+ and π^-), which subsequently decay into neutrinos and antineutrinos. For instance, a π^+ decays into a positive muon and a muon neutrino, and the positive muon further decays into a positron, an electron neutrino, and a muon antineutrino.

2. Proton-Proton Interactions (pp interactions)

Here, high-energy protons in the jet collide with protons in the surrounding medium, producing pions. These pions also decay into neutrinos.

3. Decay of Neutral Pions

Alongside charged pions, neutral pions (π^0) are also produced in pp and p γ interactions. These neutral pions decay into high-energy gamma rays. Although this process does not directly produce neutrinos, the presence of high-energy gamma rays is indicative of hadronic processes, indirectly signaling neutrino production.

These processes occur within the relativistic jets of blazars, where particle acceleration mechanisms can boost protons to energies high enough to facilitate the aforementioned interactions. Neutrinos, being weakly interacting particles, can escape dense regions and travel across the universe, eventually being detected by Earth-based observatories like the IceCube Neutrino Observatory.

Historically, extraterrestrial neutrino observations were confined to solar neutrinos and those from supernova 1987A, detected by the Kamiokande II detector (Hirata et al., 1987). These neutrinos had energies in the tens of MeV range. A breakthrough came in 2013 when the IceCube Neutrino Observatory, or IceCube (IceCube Collaboration, 2005), reported the discovery of extraterrestrial neutrinos with energies surpassing 30 TeV, significantly higher than previous observations. Between May 2010 and May 2012, IceCube detected 28 events ranging from 30 TeV to 1.2 PeV, presenting an isotropic distribution without significant clustering in time or space (IceCube Collaboration, 2013a).

Subsequent analyses of six years of IceCube data (Aartsen et al., 2016) negated the possibility of atmospheric origins for these neutrinos and found no significant correlation with the Galactic Plane, reinforcing the isotropic nature of the high-energy neutrino flux. Moreover, no correlation was observed between the arrival directions of neutrinos with energies above 200 TeV and known γ -ray sources, although this could be attributed to limited data.

The potential link between blazars and high-energy neutrinos gained traction when IceCube announced a high-energy neutrino event (IceCube-170922A) coinciding with a γ -ray flare from blazar TXS 0506+056, observed by the *Fermi* Large Area Telescope (*Fermi*-LAT, Atwood et al., 2009b). While the association had a significance of less than 3σ , it was somewhat supported by an excess of neutrino events towards TXS 0506+056 detected between September 2014 and March 2015, with a significance of 3.5σ (IceCube Collaboration et al., 2018a,b).

7.2 Neutrino Detection in IceCube

The IceCube Neutrino Observatory, located at the South Pole, marked a new era in high-energy neutrino astronomy with its first detection of extraterrestrial high-energy neutrinos in 2013 (IceCube Collaboration, 2013a). IceCube, a gigaton-scale neutrino detector, comprises over 5000 photomultiplier tubes (PMTs) distributed across 86 strings within a cubic kilometer of ice, as depicted in Fig. 7.1. Neutrinos passing through the detector may interact with atomic nuclei in the ice, producing charged particles. The PMTs detect Cherenkov radiation emitted by these charged particles, indirectly capturing information about the neutrinos. IceCube classifies neutrino events into two types: muon tracks and showers. Muon tracks result from charged current interactions of (anti-)muon neutrinos, producing muons with significant penetrating power due to their larger mass and lower bremsstrahlung radiation efficiency. These muons travel considerable distances in the ice, emitting Cherenkov radiation captured by the PMTs, as shown in Fig. 7.2 left panel. This allows for accurate reconstruction of the neutrino's arrival direction, with an error margin as small as one degree. Shower events, on the other hand, are caused by charged current reactions of (anti-)electron or (anti-)tau neutrinos and neutral current interactions of any neutrino type. Electrons, being less massive, lose energy rapidly through bremsstrahlung, resulting in shallow penetration and electromagnetic cascades that form a blob-like structure (middle panel of Fig. 7.2). While shower events provide less precise directional information, they facilitate more accurate neutrino energy measurements as most of the neutrino's energy is deposited within the detector. The detection principle for (anti-)tau neutrinos, illustrated in Fig. 7.2 right panel, involves the rapid decay of the produced tau particle into secondary charged particles, triggering cascades and exhibiting a "double-bang" structure. To date, (anti-)tau neutrinos have not been detected.

Fig. 7.3 shows the neutrino energy spectrum observed by IceCube in its first eight years of operation for neutrinos above 60 TeV. The blue solid line represents the estimated flux of conventional atmospheric neutrinos, produced by the decay of charged pions and kaons generated in cascades initiated by cosmic rays interacting with atmospheric nuclei. The green solid line indicates the subdominant component of predicted prompt neutrinos from the atmosphere, originating from the decay of heavier mesons in atmospheric showers. The atmospheric neutrino flux decreases rapidly with increasing energy, indicating that neutrinos above 300 TeV are of astrophysical origin (Aartsen et al., 2013; IceCube Collaboration, 2013a).

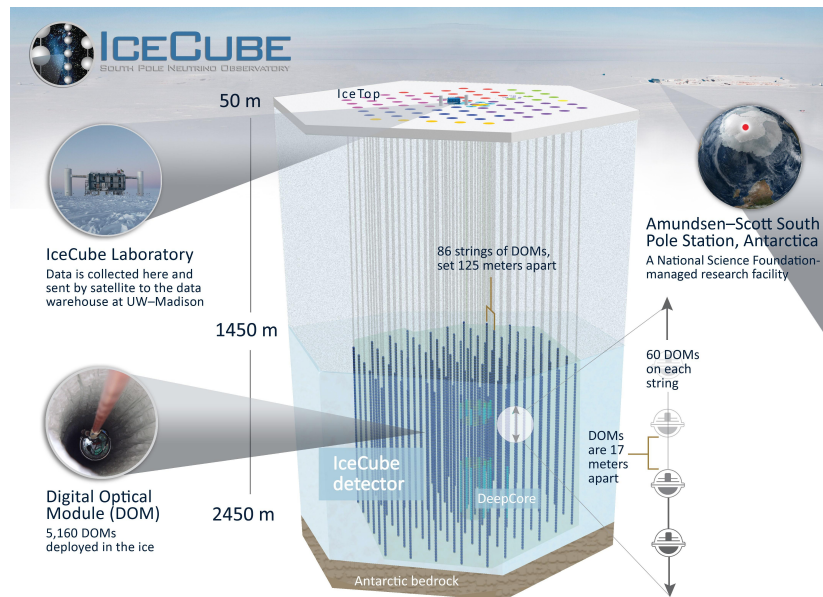


Figure 7.1: Schematic of the IceCube detector. Image source: <https://icecube.wisc.edu/science/icecube/>.

Blazars significantly contribute to the extragalactic γ -ray background (Ajello et al., 2015). If hadronic interactions predominantly produce the emitted gamma-ray photons, blazars could also significantly contribute to the diffuse neutrino background (Atoyan & Dermer, 2001a; Padovani et al., 2016; Palladino et al., 2019a). The arrival directions of muon track neutrinos, which can be accurately reconstructed, allow for the search of gamma-ray point sources correlated with neutrinos. As depicted in Fig. 7.4, the arrival directions of muon neutrinos with energies above 200 TeV detected by IceCube over eight years are approximately isotropic, suggesting an extragalactic origin for these high-energy neutrinos. Despite the simultaneous emission of photons in the hadronic interactions that produce neutrinos, there is no significant correlation between the arrival directions of these neutrinos and any known gamma-ray point sources (Halzen, 2016; Haack et al., 2017a). Comparing the arrival locations of muon neutrinos with the positions of various types of high-energy astronomical objects can effectively constrain their contribution to the diffuse neutrino background. Stacking analyses indicate no apparent association between gamma-ray bursts (GRBs), star-forming galaxies, starburst galaxies, tidal disruption events, and AGNs with the diffuse neutrino background (Stecker et al., 1991; Halzen & Zas, 1997; Atoyan & Dermer, 2001a; He et al., 2013; Tamborra et al., 2014; Aartsen et al., 2017a; Murase et al., 2020). For instance, blazars contribute at most 15% to the neutrinos detected by IceCube (Aartsen et al., 2017b; Hooper et al., 2019; Smith et al., 2021; Yuan et al., 2020). Zhou et al.

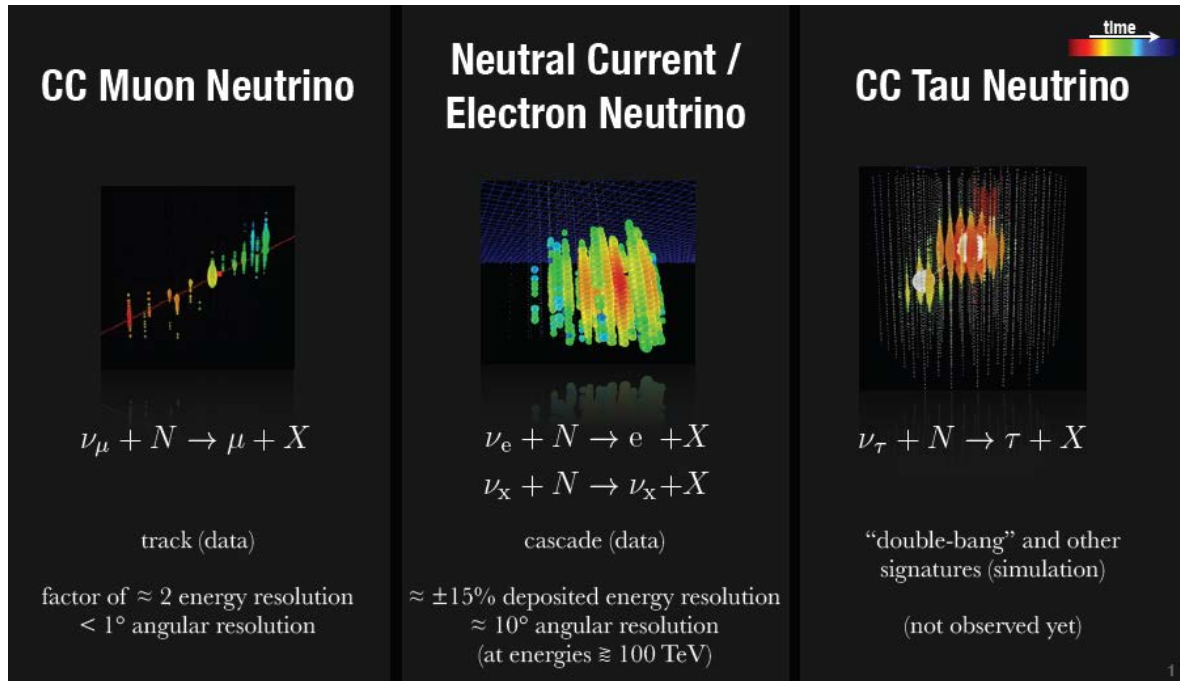


Figure 7.2: Principles of detecting different flavors of neutrinos in IceCube. Image source: chrome-extension://efaidnbmninnbpcajpcgclclefindmkaj/https://www2.physics.ox.ac.uk/sites/default/files/2012-03-27/antonin_vacheret.pdf.18644.pdf.

(2021) suggests that radio-bright AGNs contribute no more than 30% to neutrinos. Studies focusing on multiple Fermi-LAT source catalogs, including 2LAC, 3FHL, the 12-year 4FGL, and 4LAC, indicate that each catalog contributes no more than 36% to the neutrino background (Aartsen et al., 2017b; Li et al., 2022a). These results imply that most neutrino events do not originate in radio-bright AGNs or blazars.

Aartsen et al. (2020); Abbasi et al. (2021) searched a catalog of 110 strong gamma-ray sources and identified four AGNs with a 3σ level excess of neutrino flux: TXS 0506+056, NGC 1068, PKS 1424+240, and GB6 J1542+6129. Besides TXS 0506+056, which was the first confirmed neutrino source, NGC 1068, PKS 1424+240, and GB6 J1542+6129 are also considered the most promising neutrino emitters.

7.3 Neutrino Emission from Blazars and Seyfert Galaxies

- **TXS 0506+056**

On September 22, 2017, the IceCube Neutrino Observatory detected a muon neutrino event, IceCube 170922A, with an energy of approximately 290 TeV. This track-like event coincided spatially and temporally with a gamma-ray flare from the blazar TXS 0506+056, as observed by the *Fermi* Large Area Telescope (*Fermi*-LAT). Following the IceCube alert, multi-wavelength observations across radio, optical, X-ray, and very-high-energy (VHE) bands were conducted, revealing significant flux increases in all bands (see Fig. 7.5). This event is considered the first multi-messenger observation of high-energy neutrinos and photons from an astronomical source, although the association’s significance was below 3σ . Subsequent analysis of 9.5

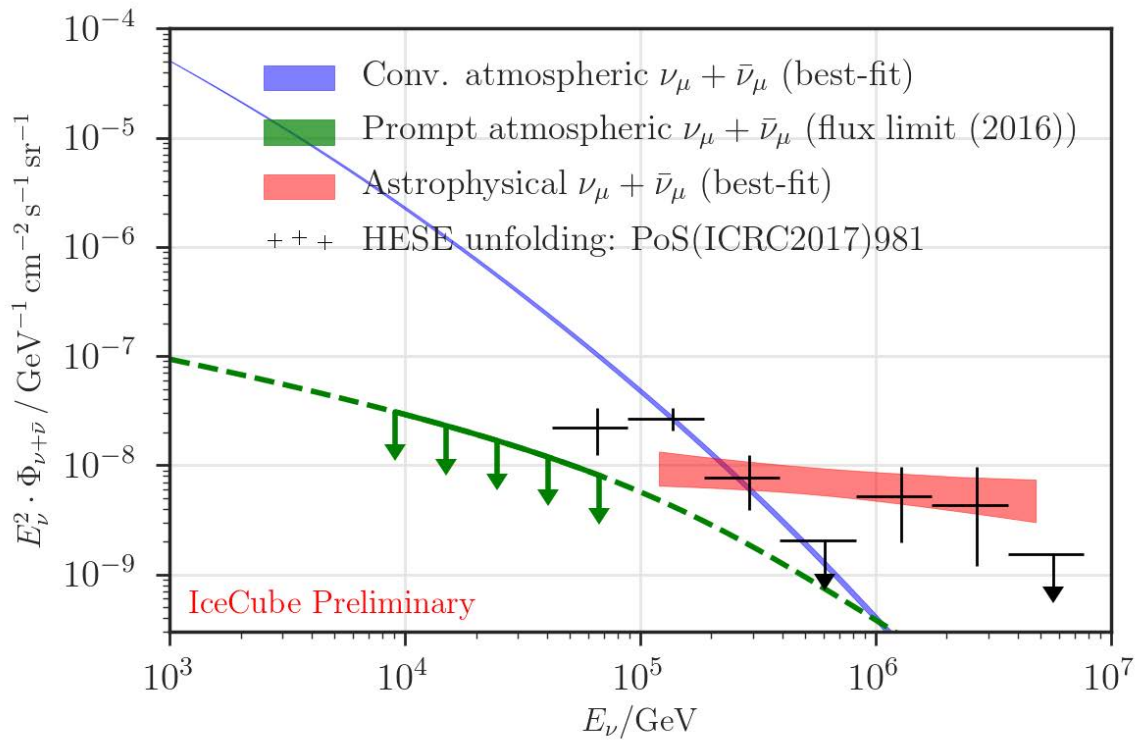


Figure 7.3: Neutrino energy spectrum observed by IceCube in its first eight years for muon neutrinos above 200 TeV. Image source: (Haack et al., 2017a).

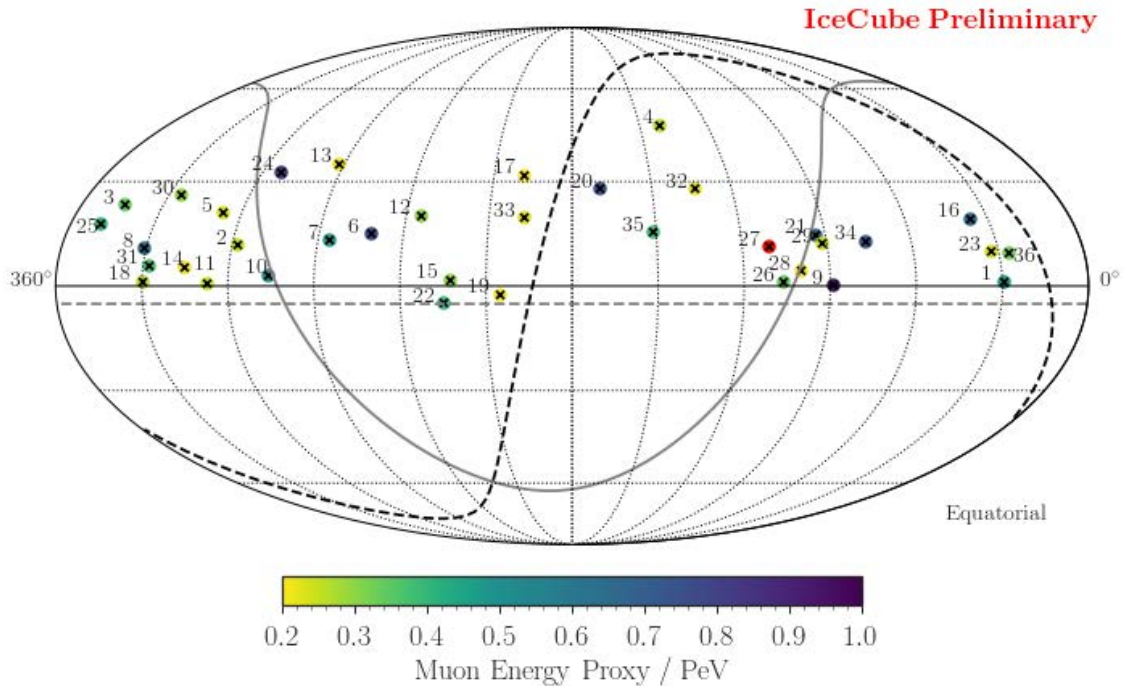


Figure 7.4: Reconstructed arrival directions of muon neutrinos with energies above 200 TeV observed by IceCube over eight years. Different colors represent different energies. The gray solid line indicates the Galactic plane and the black dashed line represents the supergalactic plane (Haack et al., 2017a).

years of IceCube data in the direction of TXS 0506+056 revealed a neutrino burst during 2014–2015, with 13 ± 5 neutrinos exceeding atmospheric background (see Fig. 7.6). However, no corresponding electromagnetic flare was detected during this period, challenging traditional blazar jet models.

- **NGC 1068**

NGC 1068 was identified as a significant neutrino source in a 10-year survey by Aartsen et al. (2020). The observed neutrino flux from NGC 1068 exceeded its GeV gamma-ray flux, suggesting distinct origins and significant attenuation of GeV gamma-rays near the neutrino production site. This implies a dense concentration of X-ray photons in the neutrino production region capable of attenuating γ -rays above 100 MeV. Despite not being the brightest X-ray Seyfert galaxy, NGC 1068’s obscured nature, due to materials including neutral hydrogen, suggests its intrinsic X-ray power is considerably higher, potentially making it a prominent source of VHE neutrinos.

- **PKS 1424+240**

PKS 1424+240, similar to TXS 0506+056, is considered a masquerading BL Lac object, intrinsically a flat-spectrum radio quasar with hidden broad lines and a standard accretion disc (Padovani et al., 2019, 2022). It has been associated with IceCube Event 35, “Big Bird”, due to spatial and temporal alignment with a 2 PeV neutrino event and a significant gamma-ray flare from PKS 1424-418. The energy output from PKS 1424-418’s flare could explain the neutrino event, suggesting a direct physical link (Kadler et al., 2016; Gao et al., 2017).

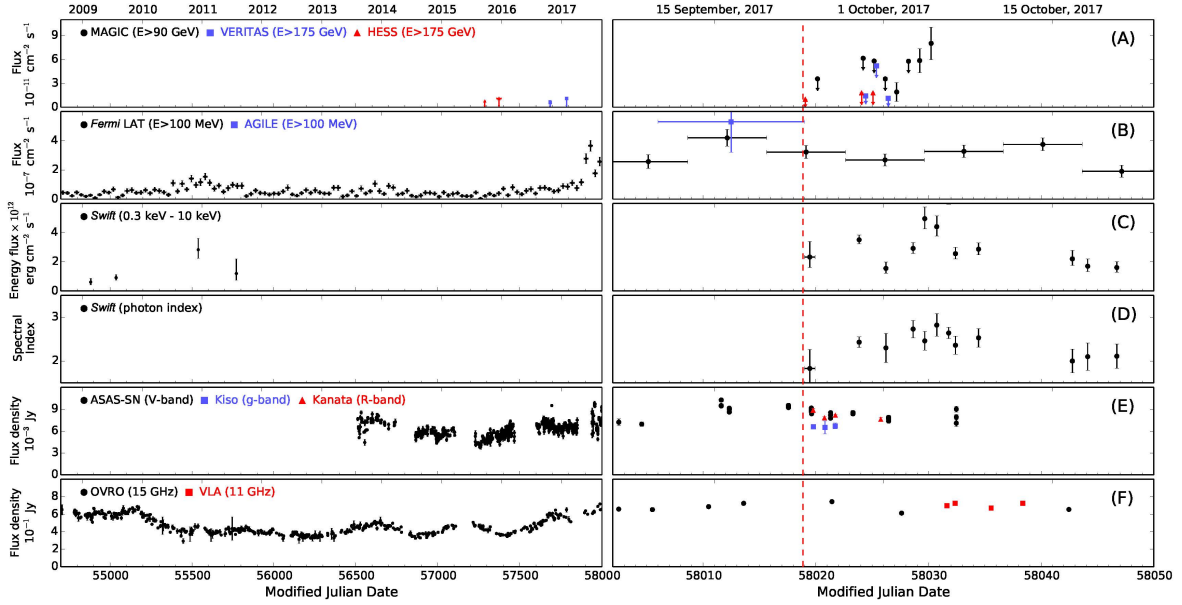


Figure 7.5: Light curves of TXS 0506+056 in various bands around the time of IceCube 170922A detection (IceCube Collaboration et al., 2018a).

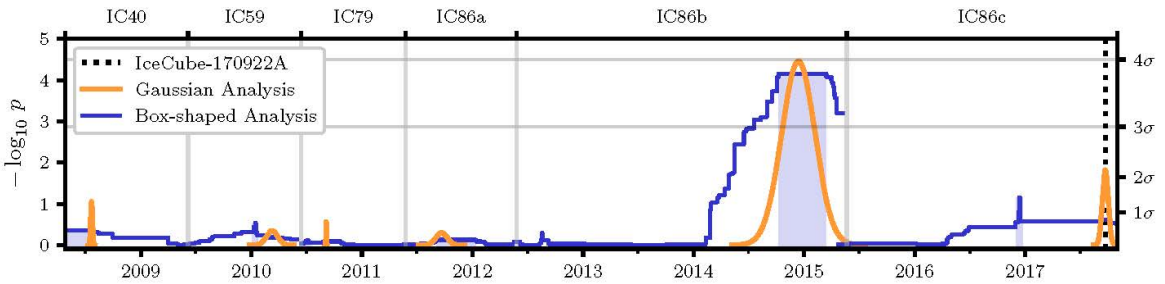


Figure 7.6: Neutrino burst from TXS 0506+056 during 2014-2015 (IceCube Collaboration et al., 2018b).

- **GB6 J1542+6129**

GB6 J1542+6129, like PKS 1424+240, is classified as a masquerading BL Lac object (Padovani et al., 2022). Its radio properties, Doppler factor, and photon spectral index are similar to those of NGC 1068 and PKS 1424+240, indicating comparable radio characteristics among these sources (Shaw et al., 2013; Padovani et al., 2022).

Other blazars, such as PKS1502+106 and NVSSJ065844+063711, have also been associated with neutrino alerts (Garrappa et al., 2022; de Menezes et al., 2022; Garrappa et al., 2021). Galván et al. (2022) reported 23 neutrino blazar candidates (NBCs) by spatially correlating IceCube neutrino events with γ -ray blazars from the *Fermi*-LAT catalog.

The debate on whether blazars contribute significantly to the high-energy neutrino diffuse flux and the lepton-to-hadron ratio in blazar jets remains open. If neutrinos are indeed produced in blazar jets, a stronger correlation between IceCube neutrinos and *Fermi*-LAT catalog sources could be expected, potentially increasing the number of identified NBCs and supporting the hadronic scenario.

In the present study, we utilize the TL to identify high-energy NBCs within the 4FGL catalog. This approach involves comparing known neutrino blazars (NBs) with other blazars listed in the 4FGL, aiming to extract a subset of potential NBCs.

7.4 Methodology

In this study, we aim to develop an ANN model to transform the initial feature space of a dataset into an embedding feature space, to segregate Non-Neutrino Blazar (non-NB) samples and NBCs into two distinct clusters. This transformation is based on identifying inherent patterns or similarities within the data. Our approach involves using a clustering algorithm to facilitate data exploration and pattern recognition, thereby grouping samples into clusters. In the embedding feature space, if a non-NB source falls within the 99% confidence error ellipse of the NB cluster, it is identified as an NB candidate. This method allows for the effective identification of potential NBs based on their proximity to the established NB cluster in the transformed feature space, thereby enhancing the selection process for NBCs.

Clustering, an essential task in data analysis, lacks a universally defined algorithm and instead encompasses a variety of methods. In this research, we employ DL techniques, particularly ANNs, to execute the clustering task. The core principle of DL in clustering involves training the ANN to discover a function that effectively maps input features to their corresponding output targets, as elucidated in the works of Bengio (2009); LeCun et al. (2015); Ahmad et al. (2019).

ANNs, the backbone of DL, operate by introducing hidden layers between the input and output layers. These hidden layers represent the input data through multiple levels of abstraction. Each layer in the network abstracts and processes the information received from its preceding layer, thereby creating a hierarchical structure of data representation. The depth of these hidden layers, along with the number of artificial neurons (ANs) in each layer, can be fine-tuned to achieve desired levels of data abstraction. This hierarchical structure is often constructed using a greedy, layer-by-layer approach, as highlighted in Bengio et al. (2006). This method has demonstrated superior learning capabilities compared to other machine learning techniques (Bengio et al., 2013b). By identifying and emphasizing more impactful features from the original dataset, DL enhances the overall effectiveness and efficiency of the machine learning process (Bengio et al., 2013b).

TL is an ML technique where knowledge gained while solving one problem is applied to a different but related problem. As mentioned in Chap. 3, the formal definition of TL involves the concepts

of domains and tasks:

- **Domain**

A domain \mathcal{D} is defined by two components: a feature space \mathcal{X} and a marginal probability distribution $P(X)$, where $X = x_1, x_2, \dots, x_n \in \mathcal{X}$. Essentially, a domain represents the space in which the input data resides, along with its distribution.

- **Task**

A task \mathcal{T} , associated with a domain, consists of two elements: a label space \mathcal{Y} and a predictive function $f(\cdot)$. The function $f : \mathcal{X} \rightarrow \mathcal{Y}$ is learned from the training data, which includes pairs of input and output (x_i, y_i) , where $x_i \in \mathcal{X}$ and $y_i \in \mathcal{Y}$. The task represents the specific problem or learning objective within the domain.

- **Transfer Learning Process**

In TL, you have a source domain \mathcal{D}_S and a corresponding source task \mathcal{T}_S , as well as a target domain \mathcal{D}_T and a target task \mathcal{T}_T . The key idea is to apply the knowledge (features, models, patterns) learned in \mathcal{D}_S and \mathcal{T}_S to improve the learning of the predictive function $f_T(\cdot)$ in \mathcal{D}_T for \mathcal{T}_T . This is particularly useful when the target domain has limited labeled data.

Advantages in Imbalanced Clustering Tasks

TL is beneficial in scenarios where the target domain has insufficient or imbalanced data. By leveraging the knowledge from a well-represented source domain, TL can enhance the learning process in the target domain, aiding in tasks like classification, regression, or clustering, even when the available data in the target domain is scarce or skewed.

7.5 Experiments

Blazars, recognized as intense extragalactic entities, are anticipated to be significant sources of neutrinos. The successful identification of neutrinos originating from these blazars offers profound implications for our understanding of astrophysical phenomena. It sheds light on the mechanisms at play in these distant celestial objects and contributes to unraveling the mysteries surrounding the genesis of cosmic rays. This detection not only enhances our comprehension of the energetic processes occurring in blazars but also aids in piecing together the broader cosmic ray puzzle, which remains one of the fundamental challenges in astrophysics.

7.6 Blazars as Neutrino Emitters

7.6.1 Sample Selection

- *Blazar Samples*

As delineated in Chapter 3, the *Fermi*-LAT collaboration has recently unveiled the incremental version of the 12-year *Fermi*-LAT Gamma-ray Source Catalog (4FGL-DR3, Abdollahi et al. 2022). This latest edition of the catalog encompasses a total of 3743 blazars, referred to as 4FGL-DR3 blazars. This group includes 794 FSRQs, 1456 BL Lacs, and 1493 BCUs. The catalog furnishes vital astrophysical data for each source, facilitating in-depth exploration of these high-energy γ -ray sources. This compilation serves as a foundational resource for identifying

NBCs within the blazar population. Additionally, the *blazar samples* are categorized into *NB samples* and *non-NB samples*.

A series of studies have been undertaken to pinpoint blazar counterparts of neutrinos, primarily utilizing the blazar catalogs provided by *Fermi*-LAT (Nolan et al., 2012a; Ackermann et al., 2011; Ajello et al., 2017; Ajello et al., 2020; Lott et al., 2020; Abdollahi et al., 2020; Abdollahi et al., 2022). Our analysis incorporates findings from eight significant publications (e.g., Kadler et al. 2016; Garrappa et al. 2019, 2021, 2022; de Menezes et al. 2022; Galván et al. 2022; Liao et al. 2022; Li et al. 2022b) that emerged from these investigations. These studies collectively highlight 35 blazars from the 4FGL-DR3 catalog, exhibiting a pronounced temporal and spatial association with neutrino events. These 35 blazars constitute our *NB samples*, and the 4FGL-DR3 blazars closely resembling them are identified as NBCs. Tab. 7.1 presents the essential information of this *NB samples*, where Column (1) lists the names in 4FGL-DR3, Column (2) indicates the SED class, Column (3) shows the classification results of BCUs following Fan et al. (2022), and Column (4) cites the relevant references. The remaining blazars in our study are classified as *non-NB samples*.

- *Non-Blazar Samples*

Excluding the 3743 blazars from 4FGL-DR3 leaves 2916 sources. After further removing 2157 unassociated sources and 134 sources of unknown classification, we are left with 625 non-blazar entities with definitive classifications. These non-blazars, detailed in Tab. 5 of Abdollahi et al. 2022, provide a contrasting dataset to the blazar samples.

7.6.2 Feature engineering

As discussed in Chapter 3, before the utilization of the samples mentioned earlier for TL, it is imperative to perform data preprocessing, commonly referred to as feature engineering. This process transforms the original features of the samples into a format that is more suitable for TL. Feature engineering encompasses the following steps:

1. *Feature Discretization*

Similar to 4FGL-DR2, a single feature in 4FGL-DR3 may be represented by multiple columns. For example, in the case of `Flux_band`, there are eight columns corresponding to the integral photon flux within each spectral band. These columns are denoted as `Flux_band_1`, `Flux_band_2`, ..., `Flux_band_8`.

2. **Feature Cleaning**

In cases where samples have missing values, we fill these gaps with the mean value of the feature across similar samples. For example, if a BL Lac lacks a redshift (z) value, we substitute it with the average z of other BL Lacs.

3. **Feature Selection**

Features presenting NAN values or string features are excluded. Additionally, features with strong correlations are evaluated, and one is chosen for efficiency. For instance, in 4FGL-DR3, integrated photon flux (*Flux_Band*) and Spectral Energy Distribution (SED) (*nuFnu_Band*) per energy interval are closely related, with the latter being preferred despite 4FGL-DR3 offering three spectral types (*PowerLaw*, *LogParabola*, and *PLSuperExpCutoff*), we standardize all

Table 7.1: 35 NB samples in 4FGL-DR3.

4FGL Name (1)	Class (2)	Class in Fan 2022 (3)	Reference (4)
4FGL J0006.4+0135	BL Lacs		Li et al. 2022b
4FGL J0118.7-0848	BCU	FSRQ	Li et al. 2022b
4FGL J0148.6+0127	BLL		Galván et al. 2022
4FGL J0206.4-1151	FSRQ		Garrappa et al. 2022
4FGL J0244.7+1316	BCU	FSRQ	Galván et al. 2022
4FGL J0258.1+2030	BLL		Galván et al. 2022
4FGL J0420.3-3745	BCU	BLL	Galván et al. 2022
4FGL J0428.6-3756	BLL		Galván et al. 2022
4FGL J0509.4+0542	BLL		Galván et al. 2022
4FGL J0525.6-2008	BLL		Galván et al. 2022
4FGL J0609.5+1402	BCU	FSRQ	Galván et al. 2022
4FGL J0649.5-3139	BLL		Galván et al. 2022
4FGL J0658.6+0636	BCU	BLL	Galván et al. 2022, de Menezes et al. 2022, Garrappa et al. 2022
4FGL J0725.8-0054	BCU	BLL	Galván et al. 2022
4FGL J0738.1+1742	BLL		Garrappa et al. 2021
4FGL J0946.2+0104	BLL		Garrappa et al. 2022
4FGL J1003.4+0205	BCU	BLL	Garrappa et al. 2022
4FGL J1039.6+0535	BCU	BLL	Galván et al. 2022
4FGL J1040.5+0617	BLL		Garrappa et al. 2019, Galván et al. 2022
4FGL J1043.6+0654	BLL		Galván et al. 2022
4FGL J1210.3+3928	BLL		Li et al. 2022b
4FGL J1220.1+3432	BLL		Galván et al. 2022
4FGL J1231.5+1421	BLL		Galván et al. 2022
4FGL J1342.7+0505	BLL		Garrappa et al. 2022
4FGL J1359.1-1152	BCU	BLL	Galván et al. 2022
4FGL J1427.0+2348	BLL		Li et al. 2022b
4FGL J1427.9-4206	FSRQ		Kadler et al. 2016
4FGL J1504.4+1029	FSRQ		Garrappa et al. 2022
4FGL J1505.0-3433	BLL		Galván et al. 2022
4FGL J1543.0+6130	BLL		Li et al. 2022b
4FGL J1744.9-1727	BCU	BLL	Galván et al. 2022
4FGL J1751.6-1750	BCU	BLL	Galván et al. 2022
4FGL J1808.8+3522	BLL		Galván et al. 2022
4FGL J2113.9+1120	BCU	FSRQ	Liao et al. 2022
4FGL J2227.9+0036	BLL		Galván et al. 2022

samples to *PowerLaw* due to its superior fit to the γ -ray data. The selected features are detailed in Tab. 7.2, with Column (1) listing the feature names, Column (2) the units, and Column (3) providing descriptions.

4. Feature Normalization

To prevent the dominance of features with larger magnitudes in the model, all features are scaled to a uniform range.

Table 7.2: The 14 selected features for TL.

Feature (1)	Unit (2)	Description (3)
Pivot Energy (E_P)	GeV	Energy at which error on differential flux is minimal
Energy Flux100 ($\log F$)	$\text{erg} \cdot \text{cm}^{-2} \cdot \text{s}^{-1}$	Energy flux from 100 MeV to 100 GeV obtained by spectral fitting
PL Flux Density ($\log f$)	$\text{cm}^{-2} \cdot \text{MeV}^{-1} \cdot \text{s}^{-1}$	Differential flux at Pivot_Energy in PowerLaw fit
PL Index (α)		Photon index when fitting with PowerLaw
Variability Index (V)		Likelihood difference between the flux fitted in each time interval and the average flux
Frac Variability (V_F)		Fractional variability computed from the fluxes in each year
nuFnu Band ($\log F_1 \sim \log F_8$)	$\text{erg} \cdot \text{cm}^{-2} \cdot \text{s}^{-1}$	Spectral energy distribution over each spectral band

7.6.3 Domains and Tasks

Based on the feature-engineered samples, we have identified distinct domains and tasks. The source domain, denoted as \mathcal{D}_S , combines NB and non-blazar samples for the binary clustering task \mathcal{T}_S : distinguishing between *NB cluster* and *non-blazar cluster*. Conversely, the target domain \mathcal{D}_T comprises NB and non-NB samples, tasked with \mathcal{T}_T : separating *NB cluster* from *non-NB cluster*. The source domain \mathcal{D}_S includes 660 sources, with 625 *non-blazar samples* and 35 *NB samples*. The target domain \mathcal{D}_T , sharing the same features as \mathcal{D}_S , contains 3743 blazars, consisting of 3708 *non-NB samples* and 35 *NB samples*.

This structure ensures $\mathcal{D}_S \neq \mathcal{D}_T$ or $\mathcal{T}_S \neq \mathcal{T}_T$. The target domain \mathcal{D}_T significantly mitigates the issue of the small proportion of NB samples. Additionally, the commonality of features between \mathcal{D}_T and \mathcal{D}_S facilitates the transition from \mathcal{T}_S to \mathcal{T}_T , enhancing the model's credibility.

7.6.4 Transfer Learning Processing and Results

The TL process is executed through the following steps, culminating in the presentation of results:

1. Dataset Partitioning

The source domain \mathcal{D}_S is divided into three segments in a 6:2:2 ratio, serving distinct purposes:

- *Training Set*

Comprising 396 samples, including 21 *NB samples*, this set is utilized to train the Artificial Neural Network (ANN). The training process enables the ANN to learn patterns and make predictions.

- *Validation Set*

Containing 132 samples with 7 *NB samples*, this set assists in fine-tuning the model's hyperparameters. These are adjustable settings chosen before training that influence the model's behavior.

- *Testing Set*

Also including 132 samples with 7 *NB samples*, this set is used independently to evaluate the model’s performance and its ability to generalize to new data. The partitioning is illustrated in Fig.7.7.

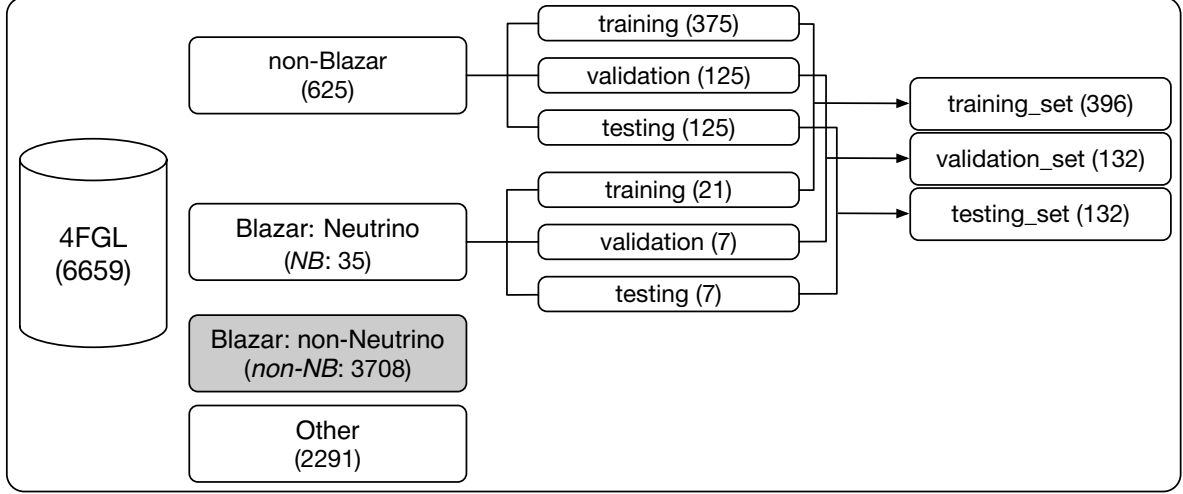


Figure 7.7: The datasets used to train or evaluate the ANN model.

2. ANN

An ANN consists of interconnected layers of ANs, facilitating information flow through input, hidden, and output layers. The ANs across these layers are interconnected via activation functions (Steele et al., 2001), enabling robust learning capabilities. The training process of the ANN is illustrated in Fig. 7.8. The input layer of the ANN comprises 14 ANs, corresponding to the 14 features of the dataset \mathcal{D}_S . For illustrative purposes, the output layer is configured with 2 ANs, mapping the original 14 features into 2 embedding features. **The ANN comprises four hidden layers. The number of nodes per layer, the learning rate of the ANN, and the activation functions between layers, among other hyperparameters, are detailed in Tab. 7.3.**

The training set samples are fed into the model, with the loss function’s objective being to learn embeddings where the anchor is closer to the positive example than to the negative example by a margin value. Ultimately, this results in the *NB samples* and *non-blazar samples* being classified into two distinct clusters. The triplet loss function (Schroff et al., 2015) is optimized and formulated in Equation 7.1, where x_a , x_p , and x_n denote the inputs of the anchor, positive, and negative samples, respectively. The function $f(\cdot)$ represents the ANN model, and m is the margin defining the distance between the two clusters.

$$\mathcal{L} = \max(0, m + \|f(x_a) - f(x_p)\|_2^2 - \|f(x_a) - f(x_n)\|_2^2) \cdot w_a. \quad (7.1)$$

To enhance performance, several strategies were employed during training. For instance, in Equation 7.1, a weight is assigned to the anchor sample if it is an NB, addressing the imbalance issue (325 *non-blazar samples* vs. 21 *NB samples*). This approach is akin to class weighting in deep learning, signaling the model to prioritize under-represented classes. Consequently,

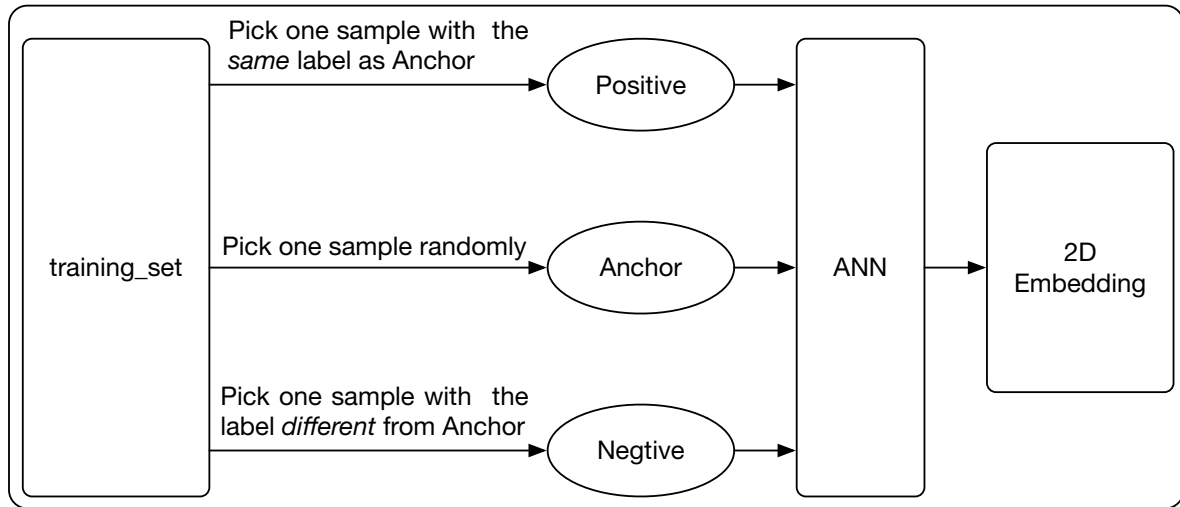


Figure 7.8: Training logic utilizing the triplet loss function.

$w_a = 1$ for non-NB anchor samples. Additionally, the learning rate was linearly adjusted during training, and Dropout (Srivastava et al., 2014) was utilized to mitigate overfitting. The hyper-parameter details are summarized in Tab.7.3.

Table 7.3: The structure of ANN model and the hyper-parameters of the training process.

Parameters or Model (1)	Description (2)	Values (3)
ANN model	the units of each layer	[14, 32, 16, 16, 8, 2]
normalization method	fit on training set	StandardScaler of sci-kit-learn (Pedregosa et al., 2011b)
batch size		128
learning rate	initial	0.001
decay factor of learning rate		0.95
epochs		300
margin		1.0
drop out probability	in each layer	0.3
w_a	weight of NB sample	8
activation function	in each layer	parametric relu

In the feature space of embeddings, we categorize clusters into NBs and non-NBs. Potential NBCs are identified within the non-NB clusters, specifically, those falling within the 99% confidence interval. The processes of training and evaluation are conducted using the well-known open-source DL framework, PyTorch.

3. Results

The analysis of the training, verification, and testing sets in the embedding feature space demonstrates the successful training of the Artificial Neural Network (ANN). The *NB samples* and *non-blazar samples* are distinctly segregated into two separate clusters within this space. The non-blazar cluster is predominantly located in the upper left corner, while the NB cluster is situated in the lower right corner, as depicted in Figure 7.9. The training set includes 3 out-cluster NBs in the upper left corner. In contrast, the validation and testing sets contain 1 and 3 outliers, respectively. Additionally, the *blazar samples* comprise 7 outliers.

Utilizing this ANN, the dataset \mathcal{D}_T is projected into the embedding feature space. The 20 most densely populated samples in the NB cluster are designated as the "core." Subsequently, *non-NB samples* located within the 99% confidence level error ellipse are classified as our Non-Blazar Candidates (NBCs). This process yields a total of 273 candidates, including 127 BL Lacs, 49 FSRQs, and 97 BCUs. These candidates are enumerated in Tab. 7.4. Here, Column (1) lists the 4FGL-DR3 names of the NBCs; Columns (2) and (3) provide the coordinates in the embedding feature space; Column (4) indicates the source class in 4FGL-DR3, and Column (5) presents the classification of BCUs according to Fan et al. (2022).

Table 7.4: 273 NBCs*.

Source name (1)	x-axis (2)	y-axis (3)	Class (4)	Class in Fan 2022 (5)
4FGL J0003.3-1928	-0.0345	-0.0297	BCU	BLL
4FGL J0004.3+4614	-0.0346	-0.0283	FSRQ	
4FGL J0019.3-8152	-0.0352	-0.0235	BLL	
4FGL J0021.9-5140	-0.0337	-0.0313	BLL	
4FGL J0022.5+0608	-0.0355	-0.0354	BLL	
...	

* The full details of this table can be found in 9.2 (see Appendix 9).

7.7 Discussion

The advent of high-energy neutrinos from extragalactic sources marks a significant milestone in neutrino astronomy (IceCube Collaboration, 2013b). It is generally accepted that these high-energy neutrinos are produced through the interaction of cosmic rays with matter or photon fields within their source, leading to the formation of charged muons and their subsequent decay into neutrinos. Extragalactic entities such as starburst galaxies (Liu et al., 2014; Chang et al., 2015), tidal disruption events (Lunardini & Winter, 2017; Senno et al., 2017), and AGN (Stecker, 2013; Murase et al., 2014; Padovani et al., 2015) are considered potential sources of these high-energy neutrinos and cosmic rays. Blazars, a primary source of the extragalactic diffuse gamma-ray background (Ajello et al., 2015), could significantly contribute to the diffuse neutrino background if their emitted gamma-rays are primarily from hadron interactions (Atoyan & Dermer, 2001b; Zhu et al., 2013; Padovani et al., 2016a; Palladino et al., 2019b). The accurate reconstruction of neutrino arrival directions through muon tracks enables the search for gamma-ray point sources associated with these neutrinos. The isotropic distribution of high-energy neutrinos, as indicated by IceCube data, suggests their predominantly extragalactic origin. However, a lack of significant correlation between neutrino arrival directions and known gamma-ray point sources has been observed (Haack et al., 2017b; Halzen, 2017). By comparing the arrival positions of high-energy neutrinos with those of various high-energy objects, the contribution of each object type to the diffuse neutrino background can be constrained. Stacking analyses have shown that gamma-ray bursts contribute at most 1% to the neutrino background (Aartsen et al., 2017a), while blazars contribute less than 30% (Aartsen et al., 2017b).

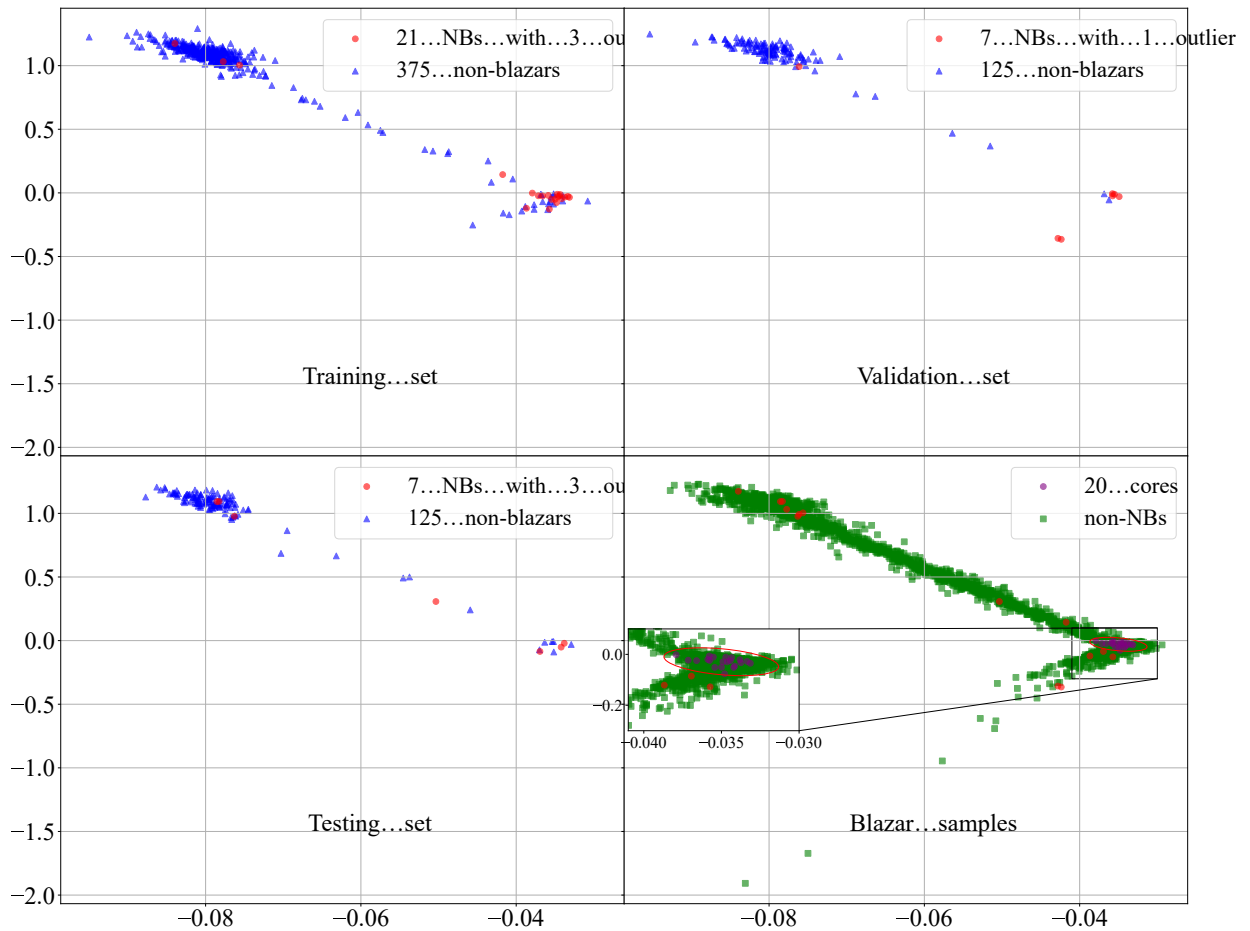


Figure 7.9: The clustering results are displayed across four panels, each representing a different dataset. *Top Left Panel:* Clustering on the training set; *Top Right Panel:* Clustering on the validation set; *Bottom Left Panel:* Clustering on the testing set; *Bottom Right Panel:* Clustering on the blazar samples. In these panels, the NB samples are represented by red circles, while the non-blazar samples are denoted by blue triangles. The 20 core samples in the densest cluster of the NB samples in the bottom right panel are highlighted by purple circles. Additionally, the non-NB samples are illustrated by green rectangles.

7.7.1 Statistical Analysis

Our study aimed to identify distributional differences in various features among three distinct non-overlapping categories: 35 neutrino NBs, 3435 non-NBs, and 273 NBCs. Building on (Yang et al., 2022b), which performed SED fitting for 2709 blazars from the 4FGL-DR3 catalog across radio to X-ray wavelengths, we included key features like the synchrotron spectral component's peak frequency ($\log \nu_p^s$) and peak luminosity ($\log L_p^s$). Additionally, (Yang et al., 2023) determined the peak frequency ($\log \nu_p^{IC}$) and peak luminosity ($\log L_p^{IC}$) of the IC component for 3743 blazars from the 4FGL-DR3. By cross-matching samples from Abdollahi et al. 2022; Yang et al. 2022b, 2023, we compiled an enhanced sample of 3734 4FGL-DR3 blazars with four novel features: \mathcal{DT} : $\log \nu_p^s$, $\log L_p^s$, $\log \nu_p^{IC}$, and $\log L_p^{IC}$.

We employed the two-sample Kolmogorov-Smirnov test (KS test, Massey 1951) and t-test, setting the significance level at $\alpha = 0.05$. We also calculated the mean and standard deviation for each sample. The p-values from the KS test and t-test, along with the mean and standard deviations of each feature, are listed in Tab. 7.5. Column 1 lists the feature names; Columns 2 to 4 provide the p-values from both tests, conducted pairwise across the three samples; Column 5 presents the mean and standard deviation for each sample.

Table 7.5: Statistical results

Features (1)	NBs vs. non-NBs (2)	NBs vs. NBCs (3)	non-NBs vs. NBCs (4)	mean \pm std (NBs, non-NBs, NBCs) (5)
E_p	$p_{KS} = 0.75, p_t = 0.67$	$p_{KS} = 0.91, p_t = 0.85$	$p_{KS} = 0.48, p_t = 0.6$	$2.8 \pm 3.19, 2.6 \pm 2.76, 2.69 \pm 3.1$
$\log F$	$p_{KS} = 0.11, p_t = 0.0$	$p_{KS} = 0.15, p_t = 0.0$	$p_{KS} = 0.39, p_t = 0.41$	$-11.16 \pm 0.72, -11.44 \pm 0.44, -11.41 \pm 0.42$
$\log f$	$p_{KS} = 0.65, p_t = 0.28$	$p_{KS} = 0.58, p_t = 0.23$	$p_{KS} = 0.84, p_t = 0.64$	$-12.73 \pm 1.23, -12.92 \pm 1.03, -12.95 \pm 0.98$
α	$p_{KS} = 0.02, p_t = 0.06$	$p_{KS} = 0.28, p_t = 0.39$	$p_{KS} = 0.01, p_t = 0.01$	$2.13 \pm 0.25, 2.22 \pm 0.3, 2.17 \pm 0.28$
V	$p_{KS} = 0.21, p_t = 0.03$	$p_{KS} = 0.07, p_t = 0.0$	$p_{KS} = 0.0, p_t = 0.23$	$1079.7 \pm 3477.07, 243.65 \pm 2316.15, 74.91 \pm 146.74$
V_F	$p_{KS} = 0.96, p_t = 0.66$	$p_{KS} = 0.03, p_t = 0.03$	$p_{KS} = 0.0, p_t = 0.0$	$0.37 \pm 0.32, 0.4 \pm 0.39, 0.48 \pm 0.26$
$\log F_1$	$p_{KS} = 0.06, p_t = 0.41$	$p_{KS} = 0.06, p_t = 0.72$	$p_{KS} = 0.01, p_t = 0.0$	$-12.97 \pm 2.13, -13.23 \pm 1.84, -12.87 \pm 1.48$
$\log F_2$	$p_{KS} = 0.03, p_t = 0.02$	$p_{KS} = 0.09, p_t = 0.08$	$p_{KS} = 0.03, p_t = 0.0$	$-12.08 \pm 1.21, -12.79 \pm 1.74, -12.46 \pm 1.18$
$\log F_3$	$p_{KS} = 0.09, p_t = 0.05$	$p_{KS} = 0.1, p_t = 0.06$	$p_{KS} = 0.8, p_t = 0.35$	$-12.03 \pm 0.96, -12.38 \pm 1.08, -12.32 \pm 0.84$
$\log F_4$	$p_{KS} = 0.1, p_t = 0.0$	$p_{KS} = 0.16, p_t = 0.01$	$p_{KS} = 0.48, p_t = 0.26$	$-11.97 \pm 0.76, -12.31 \pm 0.6, -12.35 \pm 0.84$
$\log F_5$	$p_{KS} = 0.05, p_t = 0.0$	$p_{KS} = 0.11, p_t = 0.0$	$p_{KS} = 0.0, p_t = 0.02$	$-12.06 \pm 0.79, -12.45 \pm 0.65, -12.35 \pm 0.54$
$\log F_6$	$p_{KS} = 0.0, p_t = 0.04$	$p_{KS} = 0.06, p_t = 0.17$	$p_{KS} = 0.06, p_t = 0.02$	$-12.45 \pm 1.55, -13.0 \pm 1.54, -12.77 \pm 1.28$
$\log F_7$	$p_{KS} = 0.0, p_t = 0.03$	$p_{KS} = 0.01, p_t = 0.09$	$p_{KS} = 0.22, p_t = 0.28$	$-13.23 \pm 2.2, -14.07 \pm 2.32, -13.91 \pm 2.21$
$\log F_8$	$p_{KS} = 0.08, p_t = 0.09$	$p_{KS} = 0.06, p_t = 0.06$	$p_{KS} = 0.72, p_t = 0.25$	$-15.09 \pm 3.0, -15.78 \pm 2.37, -15.95 \pm 2.43$
$\log m_{\text{syn}}$	$p_{KS} = 0.36, p_t = 0.74$	$p_{KS} = 0.56, p_t = 0.69$	$p_{KS} = 0.18, p_t = 0.04$	$14.36 \pm 1.07, 14.28 \pm 1.25, 14.47 \pm 1.33$
$\log L_{\text{syn}}^p$	$p_{KS} = 0.34, p_t = 0.16$	$p_{KS} = 0.46, p_t = 0.22$	$p_{KS} = 0.19, p_t = 0.78$	$45.56 \pm 0.86, 45.33 \pm 0.85, 45.35 \pm 0.86$
$\log m_{\text{IC}}$	$p_{KS} = 0.05, p_t = 0.12$	$p_{KS} = 0.5, p_t = 0.57$	$p_{KS} = 0.03, p_t = 0.01$	$23.16 \pm 1.08, 22.81 \pm 1.35, 23.04 \pm 1.27$
$\log L_{\text{IC}}^p$	$p_{KS} = 0.61, p_t = 0.24$	$p_{KS} = 0.29, p_t = 0.08$	$p_{KS} = 0.16, p_t = 0.06$	$45.51 \pm 1.2, 45.29 \pm 1.12, 45.15 \pm 1.11$

Here we present a detailed statistical analysis of various features across different samples, including NBs, non-NBs, and NBCs. The analysis employs both the Kolmogorov-Smirnov (KS) test and the t-test. The following attributes exhibit notable differences:

- $\log F$: Significant differences in means between NBs and non-NBs ($p_t = 0.0$), and NBs and NBCs ($p_t = 0.0$) are observed. NBs show a higher mean (-11.16) compared to non-NBs (-11.44) and NBCs (-11.41).
- α : The KS test reveals significant distribution differences between NBs and non-NBs ($p_{KS} = 0.02$), and non-NBs and NBCs ($p_{KS} = 0.01$). The t-test also indicates a significant mean difference between non-NBs and NBCs ($p_t = 0.01$).

- V : Significant distribution differences between non-NBs and NBCs ($p_{KS} = 0.0$) are noted. The t-test shows significant mean differences between NBs and non-NBs ($p_t = 0.03$), and NBs and NBCs ($p_t = 0.0$). NBs have a notably higher mean (1079.7) compared to non-NBs (243.65) and NBCs (74.91).
- V_F : The KS test indicates significant distribution differences between NBs and NBCs ($p_{KS} = 0.03$), and non-NBs and NBCs ($p_{KS} = 0.0$). The t-test also shows significant mean differences between these groups.
- $\log F_1$: Significant distribution differences between non-NBs and NBCs are observed ($p_{KS} = 0.01$, $p_t = 0.0$).
- $\log F_2$: Both KS and t-tests suggest significant differences between NBs and non-NBs ($p_{KS} = 0.03$, $p_t = 0.02$), and between non-NBs and NBCs ($p_{KS} = 0.03$).
- $\log F_4$, $\log F_5$, $\log F_6$, $\log F_7$: Significant differences are noted in these features as well, as indicated by the KS and t-tests.
- $\log \nu_{IC}$: The KS test shows significant distribution differences between NBs and non-NBs ($p_{KS} = 0.05$), and non-NBs and NBCs ($p_{KS} = 0.03$). The t-test also indicates significant mean differences between non-NBs and NBCs ($p_t = 0.01$).

In the γ -ray spectrum, our analysis reveals distinct characteristics for NBs and NBCs when compared to non-NBs. Specifically, both NBs and NBCs exhibit a higher flux and flux density in this band. Notably, NBs demonstrate a more pronounced variability in their emission. Furthermore, NBCs are characterized by a marginally higher IC peak frequency. These observations collectively suggest that NBs and NBCs are more active and emit more intensely in the gamma-ray spectrum than their non-NB counterparts.

7.7.2 Single-Zone vs. Two-Zone Models

Extensive research has been conducted on the association of IceCube-170922A and TXS 0506+056, primarily using the hybrid leptohadronic single-zone model, which combines leptonic processes with $p\gamma$ interactions (Ansoldi et al., 2018; Keivani et al., 2018; Cerruti et al., 2018; Gao et al., 2019). In this model, the blazar's multiwavelength spectrum is explained by the synchrotron and IC processes of primary relativistic electrons within the radiation region. Neutrinos are produced through pp interactions, and the decay of π^0 mesons, along with high-energy photons from synchrotron and IC scattering of secondary electron-positron pairs, can undergo $\gamma\gamma$ annihilation with soft photons in the radiation region (Mannheim et al., 1992). This process transfers the energy of high-energy electrons to the X-ray to TeV energy band (Petropoulou & Mastichiadis, 2014). Observations by *Swift* and *NuSTAR* provide strong constraints on the electromagnetic cascade (Keivani et al., 2018; Murase et al., 2018). In the single-zone model, both the electromagnetic cascade and neutrino production occur in the same photon field, leading to X-ray emission that limits the neutrino flux. Studies of IceCube's extremely high energy (EHE) track alerts suggest that the single-zone model predicts a maximum annual detection rate of neutrinos of only 0.03 (IceCube Collaboration et al., 2018a; Gao et al., 2019; Keivani et al., 2018), indicating a less than one percent probability of detecting neutrinos from this months-long blazar flare with IceCube, challenging the single-zone model's validity.

In response, a two-zone model was proposed by Zhang et al. (2019); Xue et al. (2019); Murase et al. (2018). This model necessitates an external photon field, typically provided by the broad-line

region (BLR), for neutrino production. The radiation region is thus divided into two zones: the inner zone, located inside or near the BLR, and the outer zone, situated far from the BLR. In the inner zone, relativistic protons in the jet interact with BLR clouds, using the BLR photons for both the $p\gamma$ process and IC scattering. Additionally, the BLR clouds may facilitate the pp process within the jet (Sahakyan, 2018; Liu et al., 2019). The outer zone, lacking an adequate external photon field for the $p\gamma$ process and a sufficient supply of target protons for the pp process, presents challenges in neutrino production. Thus, the two-zone model allows for separate radiation regions for the low-energy peak in the SED of blazars and neutrino emission. The inner zone can produce a neutrino flux an order of magnitude higher than the single-zone model, while the outer zone remains consistent with X-ray flux observational constraints.

The $p\gamma$ process necessitates an external photon field, a requirement irrespective of whether a single-zone or two-zone model is employed. This criterion renders FSRQs, with their higher luminosity and external photon fields, more favorable as neutrino emitters compared to BL Lacs (Atoyan & Dermer, 2003; Murase et al., 2014; Rodrigues et al., 2018). TXS 0506+056, typically classified as a BL Lac due to the detection of weak broad emission lines in its spectral analysis, might be misclassified. This misclassification could stem from the obscuration of the Broad-Line Region (BLR) by the dominant non-thermal jet emission. Indeed, Padovani et al. (2019) suggested that TXS 0506+056 is a masquerading BL Lac, harboring a concealed BLR with a luminosity of approximately 5×10^{43} erg/s and a standard accretion disk. Despite this, BL Lacs have also been proposed as potential neutrino emitters, primarily due to their lower energy requirements compared to FSRQs (Cerruti et al., 2015b, 2017; Zech et al., 2017). Several studies have supported the idea that bright, γ -loud BL Lacs could account for the observed IceCube neutrinos (Padovani & Resconi, 2014; Petropoulou et al., 2015; Righi et al., 2017). In the context of a simplified blazar view, Padovani et al. (2015) concluded that BL Lacs could explain approximately 10% of the neutrino background above a certain energy threshold, with individual BL Lacs potentially contributing up to 20% at lower energies.

In our dataset comprising 35 NBs and 273 NBCs, the NB group includes 20 BL Lacs, 3 FSRQs, and 12 BCUs. Among the 273 NBCs, there are 127 BL Lacs, 49 FSRQs, and 97 BCUs. Fan et al. (2022) employed the SVM to categorize BCUs in the 4FGL-DR3 catalog. By combining three features in pairs, they created three sets of feature spaces for BCU classification. BCUs meeting the criteria in all three feature spaces were categorized as BL Lacs/FSRQs, while those satisfying the criteria in only two feature spaces were considered possible BL Lacs/FSRQs (P-B/P-F). For our analysis, we classified both as BL Lacs/FSRQs. Consequently, out of the 12 BCUs in the NB set, 8 were classified as BL Lacs and 4 as FSRQs. Similarly, among the 97 BCUs in the NBC set, 62 were classified as BL Lacs and 25 as FSRQs. The classification results of BCUs are detailed in Tab. 7.1 and 7.4. BL Lacs dominate both the NBs and NBCs, accounting for 80% (28/35) and 69% (189/273) respectively, underscoring their prominence as neutrino emitter candidates.

TL, as a data-driven algorithm, is influenced by various factors, including updates to the 4FGL catalog, changes in hyperparameters, and different methods of training set partitioning. These factors can lead to diverse clustering outcomes. Therefore, our methods and findings should be considered heuristic. We present a novel and effective approach, distinct from traditional physical methods, for employing AI techniques in the exploration of NBCs and other rare source candidates. Even if future observations confirm blazars as sources of neutrino emitters, AI methods will continue to be invaluable in identifying NBCs. The verification of the 273 NBCs identified in our study will depend on observations from IceCube, as well as upcoming neutrino detectors like IceCube-Gen2 (Aartsen et al., 2021) and KM3Net (Adrián-Martínez et al., 2016).

7.8 Conclusion

In our exploration of the potential correlation between neutrino events detected by IceCube and electromagnetic flares from blazars, we encountered the inherent challenges associated with the limited detection of neutrinos. Despite these obstacles, our analysis successfully segregated NBs and non-blazars into distinct clusters within the embedding feature space. Utilizing TL, we effectively clustered both NBs and non-NBs, demonstrating the robustness of TL in identifying NBs and showcasing the potential of Artificial Intelligence in handling extensive astronomical datasets.

Our findings revealed that both NBs and NBCs typically exhibit higher flux values and flux densities in the γ -ray band compared to non-NBs. Specifically, NBs displayed a higher mean flux ($\log F$) of -11.16, surpassing the -11.44 mean flux of non-NBs. NBCs also showed a slightly higher mean flux of -11.41. In terms of flux density across various γ -ray bands, from $\log F_1$ to $\log F_8$, NBCs had a higher mean flux density in $\log F_1$ (-12.87) compared to non-NBs (-13.23). Furthermore, NBs consistently demonstrated higher flux densities than non-NBs in bands such as $\log F_4$ (-11.97 vs. -12.31), $\log F_5$ (-12.06 vs. -12.45), $\log F_6$ (-12.45 vs. -13.0), and $\log F_7$ (-13.23 vs. -14.07). The variability (V) observed in NBs, with a mean value of 1079.7, was significantly higher than the 243.65 observed in non-NBs. Additionally, NBCs exhibited a marginally higher $\log \nu_{\text{IC}}$ of 23.04, slightly exceeding the 22.81 of non-NBs.

Chapter 8

Conclusions and Prospect

AGN are cosmic entities characterized by their broad electromagnetic emissions spanning from radio to γ -ray bands, playing a pivotal role in our understanding of galactic evolution. Blazars, a distinct subset of AGN, stand out due to their rapid variability and pronounced polarization. Given their unique attributes and significance, blazars have naturally become the focal point in AGN research. This prominence is further underscored by the increasing deployment of sophisticated and sensitive telescopes dedicated to their observations.

Thanks to the invaluable contributions of telescopes such as *Fermi*-LAT and VLBI, the number of identified blazars has surged into the thousands. These observations have not only expanded the catalog but also enriched it by revealing a plethora of physical attributes associated with blazars. This expansive and information-rich sample has paved the way for the integration of data science, particularly on a large scale, into blazar research. This thesis is dedicated to harnessing cutting-edge data science methodology, specifically AI, to undertake the classification/clustering of blazars and subsequently identify those with rare properties. The principal conclusions derived from this endeavor are enumerated as follows:

- The debate surrounding the radio dichotomy in AGNs has been addressed by analyzing AGNs with optical B band and radio 6 cm wavelength data. A clear division in radio loudness, denoted as $\log R$, was identified at $\log R = \langle 1.37 \pm 0.02 \rangle$, differentiating RL from RQ AGNs. A dual criterion for AGN classification was proposed, merging radio luminosity and loudness. This machine learning-derived relation is expressed as $\log L_{6\text{cm}} = -2.7 \log R + 44.3$. Both RLs and RQs appear to have analogous emission sources, such as jets and mini-jets.
- In our study of 407 VLBI-detected *Fermi* blazars, we found correlations between jet knot apparent velocity and jet/accretion disc properties. γ -ray luminosity showed a positive relationship with jet power. The apparent motion of knots correlated with jet radiation and kinetic power, but not consistently for flat-spectrum radio quasars. Accretion disc luminosity and normalized accretion rate appear crucial for knot generation and acceleration. VFBs have higher γ -ray luminosity, photon index, and variability than other *Fermi* blazars. Using Gaussian mixture models, we identified 228 potential VFBs from the remaining *Fermi* blazars.
- Employing an SML technique, specifically LR, potential TeV blazar candidates were identified. The LR model, defined as $LR(y) = \frac{1}{1+e^{-(\beta_0+\beta_1 x)}}$, was applied to various catalogs, resulting in the identification of 40 high-confidence TeV candidates. Subsequent analysis indicated that 7 of these are probable detections for current IACT observatories.

- The potential of blazars as neutrino emitters was explored, focusing on the 4FGL-DR3. Utilizing the 4FGL-DR3 catalog, blazars' potential as neutrino emitters were examined using TL, leading to the identification of 273 NB candidates. This research underscores the relationship between blazars and neutrinos, offering insights into cosmic ray mechanisms. Comparative analyses revealed that both NBs and NBCs consistently exhibit higher $\log F$ and flux densities in the γ -ray band than non-NBs. Notably, NBs display a significantly elevated V value, and NBCs have a marginally higher $\log \nu_{IC}$, highlighting their distinct emission properties.

8.1 Prospect

As astronomy transitions into the era of big data, the voluminous observational data anticipated poses challenges to traditional data processing and analytical methodologies. The pressing issue lies in efficiently extracting valuable insights from this vast data pool, emphasizing the need for data mining and automation. Data mining, often heralded as the fourth paradigm of scientific research (Hey et al., 2009) following theory, experimentation, and computer simulation, has seen a surge in relevance. AI offers a plethora of algorithms for data mining, aligning perfectly with the big data trend in astronomy. In certain contexts, AI has demonstrated superior performance compared to traditional data analysis techniques. For example, White et al. (2000) employed decision trees, an ML technique, to identify quasar candidates from survey data, achieving an accuracy of 85% as opposed to the 60% accuracy of conventional methods. Similarly, Huertas-Company et al. (2008, 2009) and colleagues utilized SVM in the COSMOS survey, achieving early-late separation at $KAB = 22$ mag, twice as effective as the concentration, asymmetry, and clumpiness system (CAS, Conselice 2003).

The applications of AI in blazar research, as showcased in this thesis, represent merely the tip of the iceberg. There's immense potential for AI's integration into broader astronomical research areas, such as real-time processing, the virtual observatory, and telescope scheduling. It is anticipated that more advanced AI techniques will be incorporated into these domains in the future. While AI practitioners or data scientists might lean towards employing more sophisticated algorithms, their domain-specific knowledge might be limited. Hence, it's imperative for astronomers to pose valuable domain-specific questions and provide essential knowledge. The synergy between astronomy and AI undoubtedly serves as a significant catalyst for astronomical research and is poised to be a dominant trend in future astronomical endeavors.

Chapter 9

Appendix

Table 9.1: Potential Candidates of Superluminal Blazars Detected by *Fermi**.

4FGL Name	Class	z	Flux1000 photon \cdot cm $^{-2}$ \cdot s $^{-1}$	$\log L_\gamma$ erg \cdot cm $^{-2}$ \cdot s $^{-1}$	α_{ph}	VI
(1)	(2)	(3)	(4)	(5)	(6)	(7)
4FGL J0001.5+2113	FSRQ	1.11	1.36E-09	46.74	2.66	1564.42
4FGL J0004.4-4737	FSRQ	0.88	4.36E-10	45.99	2.37	139.12
4FGL J0010.6-3025	FSRQ	1.19	3.49E-10	46.24	2.46	91.59
4FGL J0011.4+0057	FSRQ	1.49	4.29E-10	46.60	2.33	71.98
4FGL J0016.2-0016	FSRQ	1.58	4.17E-10	46.68	2.74	82.14
4FGL J0023.7+4457	FSRQ	1.06	5.47E-10	46.30	2.44	98.55
4FGL J0028.4+2001	FSRQ	1.55	6.05E-10	46.79	2.41	81.12
4FGL J0030.3-4224	FSRQ	0.50	1.03E-09	45.68	2.52	78.61
4FGL J0030.6-0212	BCU	1.80	1.53E-09	47.37	2.40	725.42
4FGL J0036.9+1832	BCU	1.60	2.76E-10	46.48	2.39	106.28
4FGL J0038.2-2459	FSRQ	1.20	1.45E-09	46.88	2.26	1644.31
4FGL J0045.1-3706	FSRQ	1.02	4.12E-10	46.11	2.59	102.67
4FGL J0047.9+2233	FSRQ	1.16	3.88E-10	46.25	2.51	125.62
4FGL J0050.0-5736	FSRQ	1.80	3.85E-10	46.79	2.62	65.86
4FGL J0050.4-0452	FSRQ	0.92	9.44E-10	46.37	2.38	206.62
4FGL J0058.0-0539	FSRQ	1.25	4.12E-10	46.36	2.46	55.40
4FGL J0058.4+3315	FSRQ	1.37	2.18E-10	46.20	2.34	79.39
4FGL J0102.4+4214	FSRQ	0.87	3.71E-10	45.88	2.68	136.28
4FGL J0104.8-2416	FSRQ	1.75	3.31E-10	46.69	2.59	82.03
4FGL J0113.1-3553	FSRQ	1.22	4.44E-10	46.37	2.40	161.09
4FGL J0117.8-2109	FSRQ	1.49	7.05E-10	46.81	2.38	139.06
4FGL J0126.0-2221	FSRQ	0.72	3.17E-10	45.59	2.59	114.90
4FGL J0133.1-5201	FSRQ	0.93	1.72E-09	46.65	2.32	1417.04
4FGL J0137.6-2430	FSRQ	0.84	9.40E-10	46.24	2.54	343.58
4FGL J0157.7-4614	FSRQ	2.29	5.20E-10	47.19	2.45	123.17
4FGL J0209.9+7229	BLL	0.90	7.78E-10	46.27	2.28	151.23

Continued on next page

Table 9.1 – Continued from previous page

4FGL Name	Class	z	Flux1000 photon \cdot cm $^{-2}$ \cdot s $^{-1}$	$\log L_\gamma$ erg \cdot cm $^{-2}$ \cdot s $^{-1}$	α_{ph}	VI
(1)	(2)	(3)	(4)	(5)	(6)	(7)
4FGL J0210.7-5101	FSRQ	1.00	5.29E-09	47.23	2.34	1486.40
4FGL J0221.1+3556	FSRQ	0.94	7.80E-09	47.33	2.29	3494.66
4FGL J0224.9+1843	FSRQ	2.69	1.70E-10	47.04	2.95	52.21
4FGL J0229.5-3644	FSRQ	2.12	7.02E-10	47.25	2.60	267.60
4FGL J0231.2-4745	FSRQ	0.77	3.14E-10	45.65	2.76	53.11
4FGL J0236.8-6136	FSRQ	0.47	9.44E-10	45.61	2.35	792.43
4FGL J0239.7+0415	FSRQ	0.98	3.81E-10	46.04	2.78	162.94
4FGL J0242.3+1102	FSRQ	2.68	4.86E-10	47.38	2.58	93.25
4FGL J0243.2-0550	FSRQ	1.81	1.66E-10	46.46	2.77	57.28
4FGL J0245.4+2408	FSRQ	2.24	3.81E-10	47.08	2.68	107.84
4FGL J0245.9-4650	FSRQ	1.39	2.65E-09	47.30	2.42	1149.33
4FGL J0253.2-5441	FSRQ	0.54	4.58E-10	45.42	2.53	176.19
4FGL J0253.9+5103	FSRQ	1.73	7.91E-10	47.04	2.47	111.14
4FGL J0303.6-6211	FSRQ	1.35	8.69E-10	46.78	2.45	153.06
4FGL J0309.9-6058	FSRQ	1.48	1.94E-09	47.24	2.49	781.66
4FGL J0312.8+0134	FSRQ	0.66	5.16E-10	45.75	2.33	84.17
4FGL J0315.9-1033	FSRQ	1.57	5.66E-10	46.78	2.48	123.41
4FGL J0325.7+2225	FSRQ	2.07	1.15E-09	47.45	2.65	176.59
4FGL J0342.2+3858	FSRQ	0.95	3.15E-10	45.94	2.26	97.20
4FGL J0343.2-2529	FSRQ	1.42	4.09E-10	46.52	2.54	194.48
4FGL J0350.6-3226	BCU	0.93	1.41E-10	45.56	2.31	94.64
4FGL J0401.7+2112	FSRQ	0.83	6.14E-10	46.07	2.42	122.28
4FGL J0407.0-3826	FSRQ	1.29	2.02E-09	47.09	2.36	343.00
4FGL J0413.1-5332	FSRQ	1.02	3.58E-10	46.07	2.41	161.40
4FGL J0416.5-1852	FSRQ	1.54	6.63E-10	46.82	2.39	159.13
4FGL J0433.6-6030	FSRQ	0.93	9.35E-10	46.37	2.44	175.02
4FGL J0440.3-4333	FSRQ	2.85	5.44E-10	47.53	2.64	626.21
4FGL J0451.8-4651	FSRQ	0.60	4.16E-10	45.52	2.45	174.41
4FGL J0455.7-4617	FSRQ	0.86	9.78E-10	46.29	2.52	167.05
4FGL J0505.8-0419	FSRQ	1.48	5.12E-10	46.66	2.40	62.17
4FGL J0507.7-6104	FSRQ	1.09	6.70E-10	46.41	2.52	69.84
4FGL J0509.4+1012	FSRQ	0.62	7.35E-10	45.81	2.41	80.22
4FGL J0526.2-4830	FSRQ	1.30	2.66E-09	47.23	2.30	731.63
4FGL J0529.3-7243	BCU	1.34	2.81E-10	46.29	2.37	57.47
4FGL J0536.4-3401	FSRQ	0.68	1.19E-09	46.10	2.62	422.54
4FGL J0539.9-2839	FSRQ	3.10	7.56E-10	47.80	2.73	107.26
4FGL J0540.8-5415	FSRQ	1.19	3.46E-10	46.23	2.63	115.19
4FGL J0601.1-7035	FSRQ	2.41	1.64E-09	47.75	2.42	578.06
4FGL J0608.9-5456	BCU	2.20	3.63E-10	47.03	2.68	76.69
4FGL J0620.5-2512	BCU	1.90	5.34E-10	47.00	2.58	96.88
4FGL J0633.4-2222	FSRQ	1.51	4.21E-10	46.61	2.60	99.51

Continued on next page

Table 9.1 – Continued from previous page

4FGL Name	Class	z	Flux1000 photon \cdot cm $^{-2}$ \cdot s $^{-1}$	$\log L_\gamma$ erg \cdot cm $^{-2}$ \cdot s $^{-1}$	α_{ph}	VI
(1)	(2)	(3)	(4)	(5)	(6)	(7)
4FGL J0635.6-7518	FSRQ	0.65	6.46E-10	45.77	2.70	448.18
4FGL J0641.7-0320	FSRQ	1.20	1.47E-09	46.87	2.68	357.07
4FGL J0643.3+0857	FSRQ	0.88	2.29E-09	46.69	2.55	366.73
4FGL J0648.0-3045	FSRQ	1.15	1.07E-09	46.69	2.49	422.16
4FGL J0648.4-1743	FSRQ	1.23	2.68E-09	47.16	2.45	1467.83
4FGL J0656.3-0322	FSRQ	0.63	2.23E-09	46.29	2.57	106.96
4FGL J0701.5-4634	FSRQ	0.82	1.26E-09	46.36	2.44	378.12
4FGL J0709.7-0255	FSRQ	1.47	2.21E-09	47.30	2.29	206.90
4FGL J0713.8+1935	FSRQ	0.54	1.44E-09	45.94	2.40	794.15
4FGL J0721.3+0405	FSRQ	0.67	1.05E-09	46.02	2.56	87.29
4FGL J0726.4-4727	FSRQ	1.69	1.02E-09	47.12	2.48	125.96
4FGL J0733.8+0455	FSRQ	3.01	4.66E-10	47.47	2.44	134.42
4FGL J0746.4+2546	FSRQ	2.99	2.39E-10	47.30	2.86	87.52
4FGL J0752.2+3313	FSRQ	1.94	1.22E-09	47.36	2.43	2114.82
4FGL J0805.2-0110	FSRQ	1.39	5.76E-10	46.64	2.48	92.96
4FGL J0809.5+5341	FSRQ	2.13	1.10E-09	47.42	2.29	407.46
4FGL J0823.1+4042	FSRQ	0.87	2.72E-10	45.75	2.46	58.80
4FGL J0824.4+2440	FSRQ	1.24	4.89E-10	46.43	2.53	196.86
4FGL J0850.1-1212	FSRQ	0.57	2.10E-09	46.18	2.30	644.25
4FGL J0857.9-1949	FSRQ	0.66	6.29E-10	45.79	2.52	73.90
4FGL J0904.9-5734	BCU	0.70	4.35E-09	46.73	2.28	767.83
4FGL J0909.7-0230	FSRQ	0.96	2.13E-09	46.78	2.29	151.09
4FGL J0910.6+2247	FSRQ	2.66	5.67E-10	47.40	2.34	151.02
4FGL J0912.2+4127	FSRQ	2.56	5.60E-10	47.37	2.46	356.20
4FGL J0916.7+3856	FSRQ	1.27	4.54E-10	46.43	2.46	66.22
4FGL J0922.6+0434	BCU	1.33	5.43E-10	46.56	2.49	157.60
4FGL J0923.5+3852	BCU	1.58	3.44E-10	46.59	2.69	91.15
4FGL J0924.0+2816	FSRQ	0.74	3.69E-10	45.70	2.51	95.94
4FGL J0939.3-1732	BCU	1.83	4.14E-10	46.82	2.26	67.54
4FGL J0943.7+6137	BCU	0.79	2.65E-10	45.62	2.71	59.01
4FGL J0946.6+1016	FSRQ	1.01	1.82E-09	46.76	2.43	291.94
4FGL J0957.3-1348	FSRQ	1.32	6.27E-10	46.62	2.47	115.00
4FGL J1001.1+2911	BLL	0.56	7.24E-10	45.71	2.27	211.58
4FGL J1006.7-2159	FSRQ	0.33	2.21E-09	45.61	2.35	1322.41
4FGL J1007.6-3332	FSRQ	1.84	1.17E-09	47.28	2.44	311.46
4FGL J1012.7+2439	FSRQ	1.81	2.02E-09	47.49	2.31	731.24
4FGL J1018.3-3124	FSRQ	0.79	9.90E-10	46.20	2.54	211.20
4FGL J1023.9-3236	FSRQ	1.57	7.28E-10	46.89	2.50	371.97
4FGL J1027.2+7427	BCU	0.88	7.52E-10	46.21	2.42	199.39
4FGL J1031.6+6019	FSRQ	1.23	5.22E-10	46.45	2.61	121.08
4FGL J1037.7-2822	FSRQ	1.07	1.55E-09	46.75	2.42	374.78

Continued on next page

Table 9.1 – Continued from previous page

4FGL Name	Class	z	Flux1000 photon \cdot cm $^{-2}$ \cdot s $^{-1}$	$\log L_{\gamma}$ erg \cdot cm $^{-2}$ \cdot s $^{-1}$	α_{ph}	VI
(1)	(2)	(3)	(4)	(5)	(6)	(7)
4FGL J1038.8-5312	FSRQ	1.45	9.25E-10	46.92	2.74	723.98
4FGL J1045.8-2928	FSRQ	2.13	3.37E-10	46.93	2.54	166.51
4FGL J1049.8+1429	BCU	1.63	6.22E-10	46.87	2.59	85.40
4FGL J1050.1+0432	FSRQ	1.22	5.90E-10	46.49	2.61	286.93
4FGL J1058.6-8003	BLL	0.57	1.95E-09	46.17	2.26	485.42
4FGL J1103.0+1157	FSRQ	0.91	1.22E-09	46.47	2.41	505.80
4FGL J1106.0+2813	FSRQ	0.84	7.74E-10	46.19	2.37	127.71
4FGL J1107.0-4449	FSRQ	1.60	6.43E-10	46.88	2.70	123.58
4FGL J1112.5+3448	FSRQ	1.95	7.78E-10	47.18	2.52	196.50
4FGL J1119.0+1235	FSRQ	2.13	4.09E-10	47.02	2.59	55.32
4FGL J1127.8+3618	FSRQ	0.88	1.08E-09	46.38	2.37	144.52
4FGL J1128.0+5924	FSRQ	1.80	3.96E-10	46.79	2.47	82.83
4FGL J1136.2+3407	FSRQ	1.34	3.57E-10	46.39	2.55	182.52
4FGL J1139.0+4033	BCU	2.36	1.97E-10	46.86	2.72	52.96
4FGL J1152.3-0839	FSRQ	2.37	8.54E-10	47.45	2.41	168.16
4FGL J1154.0+6018	FSRQ	1.12	2.65E-10	46.04	2.70	86.22
4FGL J1159.2-2227	BCU	0.57	8.31E-10	45.75	2.45	156.04
4FGL J1159.3-2142	FSRQ	0.62	1.43E-09	46.09	2.43	94.67
4FGL J1207.7-0106	FSRQ	1.01	3.57E-10	46.06	2.34	145.23
4FGL J1208.9+5441	FSRQ	1.35	1.12E-09	46.89	2.59	356.94
4FGL J1209.8+1810	FSRQ	0.85	5.80E-10	46.06	2.41	128.08
4FGL J1215.0+1656	FSRQ	1.13	6.50E-10	46.45	2.41	219.92
4FGL J1220.1+7105	FSRQ	0.45	2.03E-09	45.91	2.32	1132.39
4FGL J1223.9+5000	FSRQ	1.06	5.55E-10	46.30	2.54	148.27
4FGL J1225.0+0330	FSRQ	0.96	4.35E-10	46.06	2.63	235.11
4FGL J1228.7+4858	FSRQ	1.72	3.41E-10	46.67	2.56	87.10
4FGL J1238.5-1201	FSRQ	1.32	4.70E-10	46.50	2.66	138.06
4FGL J1257.8+3228	FSRQ	0.81	1.11E-09	46.30	2.32	362.65
4FGL J1258.6-1759	FSRQ	1.96	1.11E-09	47.33	2.39	208.44
4FGL J1258.8-2219	FSRQ	1.30	1.98E-09	47.10	2.39	508.70
4FGL J1311.0+3233	FSRQ	1.64	8.85E-10	47.02	2.40	96.70
4FGL J1312.8-0425	FSRQ	0.82	1.52E-09	46.45	2.40	573.55
4FGL J1317.6+3428	FSRQ	1.05	4.10E-10	46.15	2.53	118.94
4FGL J1322.6-0936	FSRQ	1.86	6.71E-10	47.08	2.62	371.39
4FGL J1324.9+4748	FSRQ	2.26	2.13E-10	46.83	2.67	99.02
4FGL J1326.9+2210	FSRQ	1.40	1.27E-09	46.99	2.49	178.39
4FGL J1332.6-1256	FSRQ	1.49	3.20E-09	47.47	2.42	670.33
4FGL J1333.7+5056	FSRQ	1.36	9.71E-10	46.85	2.34	340.38
4FGL J1339.0-2400	BCU	0.66	7.35E-10	45.89	2.30	61.17
4FGL J1340.4+6926	BCU	2.26	1.10E-10	46.52	2.53	139.49
4FGL J1341.8-2053	FSRQ	1.58	3.10E-10	46.55	2.73	106.78

Continued on next page

Table 9.1 – Continued from previous page

4FGL Name	Class	z	Flux1000 photon \cdot cm $^{-2}$ \cdot s $^{-1}$	$\log L_{\gamma}$ erg \cdot cm $^{-2}$ \cdot s $^{-1}$	α_{ph}	VI
(1)	(2)	(3)	(4)	(5)	(6)	(7)
4FGL J1345.8+0706	FSRQ	1.09	6.57E-10	46.41	2.44	210.77
4FGL J1347.6-3751	FSRQ	1.30	6.27E-10	46.60	2.34	51.92
4FGL J1351.0+0029	FSRQ	2.08	2.51E-10	46.77	2.44	92.55
4FGL J1351.7-2912	BCU	1.03	4.54E-10	46.19	2.36	75.44
4FGL J1359.1+5544	FSRQ	1.01	6.47E-10	46.32	2.43	66.20
4FGL J1419.4-0838	FSRQ	0.90	2.51E-09	46.80	2.24	495.73
4FGL J1433.0-1801	FSRQ	2.33	3.17E-10	47.03	2.60	114.81
4FGL J1434.7+1950	FSRQ	1.38	1.96E-09	47.17	2.30	626.17
4FGL J1438.9+3710	FSRQ	2.40	1.04E-09	47.55	2.41	334.80
4FGL J1450.4+0910	FSRQ	2.61	3.13E-10	47.12	2.39	76.04
4FGL J1457.4-3539	FSRQ	1.42	3.49E-09	47.45	2.32	2853.58
4FGL J1513.4-3231	FSRQ	1.15	8.05E-10	46.56	2.46	147.72
4FGL J1514.8-4748	FSRQ	1.55	1.47E-09	47.18	2.39	61.42
4FGL J1553.6-2422	FSRQ	0.33	1.56E-09	45.48	2.30	215.51
4FGL J1617.3-5849	FSRQ	1.41	1.06E-09	46.93	2.54	117.21
4FGL J1617.9-7718	FSRQ	1.71	1.54E-09	47.33	2.61	361.51
4FGL J1618.0+5139	FSRQ	2.56	1.57E-10	46.86	2.67	90.73
4FGL J1628.8-6149	FSRQ	2.58	7.14E-10	47.52	2.64	87.52
4FGL J1632.8-1048	BCU	1.40	3.55E-10	46.44	2.63	153.51
4FGL J1635.6+3628	FSRQ	3.65	2.45E-10	47.50	2.67	175.36
4FGL J1639.2+4129	FSRQ	0.69	4.15E-10	45.68	2.43	62.66
4FGL J1648.2+4232	BCU	2.50	4.98E-10	47.28	2.43	148.05
4FGL J1656.3-3301	FSRQ	2.40	7.79E-10	47.50	2.79	131.51
4FGL J1703.6-6213	FSRQ	1.76	2.85E-09	47.61	2.43	379.79
4FGL J1717.6-5154	FSRQ	1.16	9.38E-10	46.63	2.59	286.08
4FGL J1722.6+6104	FSRQ	2.06	1.72E-10	46.66	2.84	57.70
4FGL J1759.1-4822	BCU	3.86	6.83E-10	47.90	2.32	201.91
4FGL J1802.6-3940	FSRQ	0.30	4.84E-09	45.83	2.37	3170.66
4FGL J1816.9-4942	FSRQ	1.70	8.14E-10	47.03	2.39	98.55
4FGL J1821.6+6819	BCU	1.69	5.52E-10	46.89	2.72	149.15
4FGL J1829.2-5813	FSRQ	1.53	2.42E-09	47.38	2.55	2352.28
4FGL J1833.6-2103	FSRQ	2.51	1.43E-08	48.76	2.53	2473.93
4FGL J1839.6-7107	FSRQ	1.36	2.00E-10	46.18	2.86	75.43
4FGL J1913.0-8009	FSRQ	1.76	1.43E-09	47.31	2.45	355.87
4FGL J1913.4-3629	BCU	0.80	8.40E-10	46.17	2.32	73.80
4FGL J1934.3+6541	FSRQ	1.69	2.51E-10	46.56	2.86	404.36
4FGL J1937.2-3958	FSRQ	0.97	6.07E-10	46.22	2.66	55.94
4FGL J1941.7+7218	BCU	1.10	1.77E-10	45.85	2.77	81.95
4FGL J1955.2+1358	FSRQ	0.74	9.29E-10	46.11	2.41	177.69
4FGL J1959.1-4247	FSRQ	2.17	8.01E-10	47.32	2.46	381.51
4FGL J2005.8+6424	FSRQ	1.57	4.47E-10	46.68	2.37	53.47

Continued on next page

Table 9.1 – Continued from previous page

4FGL Name	Class	z	Flux1000 photon \cdot cm $^{-2}$ \cdot s $^{-1}$	$\log L_\gamma$ erg \cdot cm $^{-2}$ \cdot s $^{-1}$	α_{ph}	VI
(1)	(2)	(3)	(4)	(5)	(6)	(7)
4FGL J2005.9-2309	FSRQ	0.83	9.53E-10	46.24	2.50	269.44
4FGL J2024.6-3252	FSRQ	1.47	8.17E-10	46.86	2.48	83.91
4FGL J2038.7+5117	FSRQ	1.69	1.08E-09	47.15	2.57	80.07
4FGL J2040.5-1705	BCU	2.00	4.11E-10	46.94	2.54	63.36
4FGL J2052.2-5533	BCU	1.50	1.09E-09	47.01	2.30	481.53
4FGL J2056.2-4714	FSRQ	1.49	4.66E-09	47.63	2.45	1325.44
4FGL J2110.3+0808	FSRQ	1.58	4.74E-10	46.71	2.40	79.11
4FGL J2119.6-1105	FSRQ	1.84	3.88E-10	46.80	2.37	145.17
4FGL J2126.3-4605	FSRQ	1.67	8.08E-10	47.01	2.55	625.13
4FGL J2135.3-5006	FSRQ	2.18	8.54E-10	47.35	2.43	59.93
4FGL J2141.7-6410	BCU	0.96	2.86E-09	46.91	2.33	1711.78
4FGL J2145.0-3356	FSRQ	1.36	1.39E-09	47.00	2.30	293.96
4FGL J2147.3-7536	FSRQ	1.14	4.47E-09	47.29	2.44	2763.30
4FGL J2151.8-3027	FSRQ	2.35	1.00E-09	47.60	2.85	963.78
4FGL J2153.8-1137	FSRQ	1.58	6.60E-10	46.86	2.55	51.98
4FGL J2156.3-0036	FSRQ	0.50	4.75E-10	45.40	2.25	65.22
4FGL J2157.5+3127	FSRQ	1.49	2.96E-09	47.43	2.31	1583.81
4FGL J2201.5-8339	FSRQ	1.87	1.53E-09	47.43	2.55	392.48
4FGL J2206.8-0032	BLL	1.05	8.77E-10	46.51	2.25	99.20
4FGL J2207.5-5346	FSRQ	1.22	6.30E-10	46.52	2.44	82.58
4FGL J2212.9-2526	FSRQ	1.83	5.42E-10	46.94	2.26	77.52
4FGL J2231.0-4416	FSRQ	1.33	4.47E-10	46.47	2.43	154.95
4FGL J2235.3-4836	FSRQ	0.51	7.49E-10	45.57	2.51	105.23
4FGL J2249.4-1300	BLL	0.50	8.22E-10	45.61	2.41	247.64
4FGL J2311.0+3425	FSRQ	1.82	4.11E-09	47.81	2.35	2009.58
4FGL J2311.7+2604	BCU	1.75	3.19E-10	46.66	2.43	63.17
4FGL J2318.2+1915	BCU	2.16	3.75E-10	47.02	2.67	55.20
4FGL J2323.6-0617	FSRQ	2.14	5.54E-10	47.14	2.40	195.63
4FGL J2325.4-3559	FSRQ	0.36	1.54E-09	45.55	2.32	162.47
4FGL J2328.3-4036	FSRQ	1.97	1.85E-09	47.56	2.38	246.80
4FGL J2329.3-4955	FSRQ	0.52	1.17E-08	46.84	2.28	13180.00
4FGL J2331.0-2147	FSRQ	0.56	1.00E-09	45.82	2.44	194.84
4FGL J2335.4-0128	FSRQ	1.18	4.00E-10	46.29	2.47	52.62
4FGL J2336.6-4115	FSRQ	1.41	7.87E-10	46.79	2.33	130.21
4FGL J2338.0-0230	FSRQ	1.07	7.92E-10	46.47	2.49	130.12
4FGL J2357.8-5311	FSRQ	1.01	3.91E-10	46.08	2.64	224.58
4FGL J2358.3-1021	FSRQ	1.64	6.09E-10	46.87	2.59	136.97
4FGL J2359.2-3134	FSRQ	0.99	1.74E-10	45.71	2.60	85.62

Table 9.2: 273 NBCs*.

Source name (1)	x-axis (2)	y-axis (3)	Class (4)	Class in Fan 2022 (5)
4FGL J0003.3-1928	-0.03448	-0.02968	BCU	BLL
4FGL J0003.3-1928	-0.03448	-0.02968	BCU	BLL
4FGL J0004.3+4614	-0.03461	-0.0283	FSRQ	
4FGL J0019.3-8152	-0.03521	-0.02351	BLL	
4FGL J0021.9-5140	-0.0337	-0.03134	BLL	
4FGL J0022.5+0608	-0.03553	-0.03535	BLL	
4FGL J0028.9+3553	-0.03455	-0.02649	BLL	
4FGL J0034.0-4116	-0.03506	-0.02822	BCU	FSRQ
4FGL J0037.8+1239	-0.03539	-0.02864	BLL	
4FGL J0043.5-0442	-0.0357	-0.02947	BLL	
4FGL J0044.2-8424	-0.03359	-0.02494	FSRQ	
4FGL J0045.1-3706	-0.03463	-0.02554	FSRQ	
4FGL J0045.7+1217	-0.03607	-0.03239	BLL	
4FGL J0054.7-2455	-0.03648	-0.02594	BLL	
4FGL J0054.8-1954	-0.03333	-0.02309	BCU	BLL
4FGL J0058.3+1723	-0.031	-0.02464	BLL	
4FGL J0102.1+5846	-0.03504	-0.03366	BCU	BLL
4FGL J0102.4+0942	-0.03527	-0.03135	BLL	
4FGL J0102.4+4214	-0.03323	-0.02309	FSRQ	
4FGL J0102.6-5639	-0.03516	-0.03384	BCU	BLL
4FGL J0110.3-0431	-0.03393	-0.02522	BCU	FSRQ
4FGL J0111.5-2546	-0.03347	-0.03319	BCU	FSRQ
4FGL J0114.8+1326	-0.03456	-0.0289	BLL	
4FGL J0116.2-6153	-0.03481	-0.03127	BLL	
4FGL J0137.6-2430	-0.03318	-0.02532	FSRQ	
4FGL J0137.9+5814	-0.03654	-0.03049	BLL	
4FGL J0143.1-3622	-0.0353	-0.03531	BCU	FSRQ
4FGL J0146.0-6746	-0.03667	-0.02896	BLL	
4FGL J0153.9+0823	-0.0347	-0.03455	BLL	
4FGL J0156.5+3914	-0.03496	-0.03065	FSRQ	
4FGL J0156.8-4744	-0.03401	-0.03232	BLL	
4FGL J0158.2+2515	-0.0371	-0.03026	BLL	
4FGL J0158.5-3932	-0.0363	-0.03423	BLL	
4FGL J0202.1+3939	-0.03492	-0.02316	FSRQ	
4FGL J0204.8+1513	-0.03398	-0.02379	BCU	FSRQ
4FGL J0207.5-2402	-0.03431	-0.02921	BCU	BLL
4FGL J0207.9+0953	-0.03412	-0.02061	BCU	FSRQ
4FGL J0208.6+3523	-0.03438	-0.02416	BLL	
4FGL J0216.5+2313	-0.03676	-0.03679	BLL	
4FGL J0223.5+3912	-0.03478	-0.02085	BCU	BLL
4FGL J0226.3-1845	-0.03556	-0.03489	BCU	FSRQ

Continued on next page

Table 9.2 – Continued from previous page

Source name (1)	x-axis (2)	y-axis (3)	Class (4)	Class in Fan 2022 (5)
4FGL J0231.2-5754	-0.03642	-0.03296	BLL	
4FGL J0232.5-1118	-0.03314	-0.03022	BLL	
4FGL J0237.6+0923	-0.03523	-0.02259	BCU	BLL
4FGL J0245.4-5950	-0.03472	-0.03035	BCU	BLL
4FGL J0250.3-3422	-0.03339	-0.02396	BCU	FSRQ
4FGL J0253.9+5103	-0.03338	-0.03324	FSRQ	
4FGL J0259.4+0746	-0.0356	-0.02673	FSRQ	
4FGL J0301.0-1652	-0.03356	-0.03004	BLL	
4FGL J0303.6-6211	-0.03353	-0.02012	FSRQ	
4FGL J0304.9-0606	-0.03391	-0.03146	BLL	
4FGL J0305.1-1608	-0.03614	-0.02761	BLL	
4FGL J0312.4-3221	-0.03281	-0.02445	BCU	BLL
4FGL J0316.0-5626	-0.03261	-0.02684	BCU	BLL
4FGL J0316.2-2608	-0.03454	-0.02463	BLL	
4FGL J0322.4+6606	-0.03402	-0.02288	BCU	FSRQ
4FGL J0332.1-1123	-0.03547	-0.02585	FSRQ	
4FGL J0332.8+1557	-0.03328	-0.02795	BCU	BLL
4FGL J0334.2-4008	-0.03423	-0.02902	BLL	
4FGL J0349.4-1159	-0.03655	-0.03217	BLL	
4FGL J0350.0+0640	-0.03608	-0.03332	BCU	BLL
4FGL J0353.0+5654	-0.03607	-0.02655	BLL	
4FGL J0354.4+4643	-0.03276	-0.02082	BCU	FSRQ
4FGL J0357.6-4625	-0.03288	-0.03055	BLL	
4FGL J0420.3-6016	-0.03495	-0.02578	BLL	
4FGL J0425.3+6319	-0.03655	-0.0372	BCU	BLL
4FGL J0438.4-1254	-0.03302	-0.03137	FSRQ	
4FGL J0439.8-1859	-0.03439	-0.03462	BLL	
4FGL J0450.4+7230	-0.03637	-0.02964	BCU	BLL
4FGL J0451.8+5721	-0.03387	-0.02385	BCU	BLL
4FGL J0458.6+5509	-0.03487	-0.02579	BCU	FSRQ
4FGL J0500.6-4911	-0.03596	-0.03434	BLL	
4FGL J0505.8-0419	-0.03522	-0.02716	FSRQ	
4FGL J0510.4-1809	-0.03412	-0.03174	BCU	FSRQ
4FGL J0516.7-6207	-0.03392	-0.02548	BLL	
4FGL J0517.5+0858	-0.03203	-0.0303	FSRQ	
4FGL J0525.6-6013	-0.03475	-0.0257	BCU	BLL
4FGL J0529.1+0935	-0.03545	-0.03261	BCU	BLL
4FGL J0536.4-3343	-0.03655	-0.03053	BLL	
4FGL J0548.5-5218	-0.03311	-0.02478	BCU	BLL
4FGL J0549.0+3258c	-0.03279	-0.03157	BCU	BLL
4FGL J0550.5-3216	-0.03648	-0.03288	BLL	
4FGL J0552.8+0313	-0.03248	-0.0264	BCU	FSRQ

Continued on next page

Table 9.2 – Continued from previous page

Source name (1)	x-axis (2)	y-axis (3)	Class (4)	Class in Fan 2022 (5)
4FGL J0601.3+5444	-0.03685	-0.02521	BCU	BLL
4FGL J0608.0+6721	-0.03448	-0.02123	FSRQ	
4FGL J0614.8+6136	-0.03444	-0.03001	BCU	FSRQ
4FGL J0617.2+5701	-0.03612	-0.02645	BLL	
4FGL J0629.6+2435	-0.03411	-0.02397	BCU	BLL
4FGL J0641.7-0320	-0.03283	-0.02447	FSRQ	
4FGL J0643.3+0857	-0.03137	-0.02762	FSRQ	
4FGL J0647.0-5138	-0.03589	-0.03514	BLL	
4FGL J0647.7-6058	-0.03471	-0.03245	BLL	
4FGL J0700.5-6610	-0.03501	-0.0322	BLL	
4FGL J0712.5+3311	-0.03109	-0.02631	BCU	BLL
4FGL J0715.3-6828	-0.03416	-0.02705	BCU	BLL
4FGL J0723.5+2900	-0.03451	-0.02979	FSRQ	
4FGL J0725.5+0216	-0.03488	-0.03461	BCU	BLL
4FGL J0727.1+3734	-0.0336	-0.02298	BLL	
4FGL J0739.8-6722	-0.03048	-0.01886	BCU	BLL
4FGL J0744.1-6211	-0.03778	-0.02713	BCU	BLL
4FGL J0749.9+1823	-0.03409	-0.03424	FSRQ	
4FGL J0750.8+1229	-0.03441	-0.02863	FSRQ	
4FGL J0754.7+4823	-0.03462	-0.02272	BLL	
4FGL J0805.1+7744	-0.03461	-0.02533	BCU	FSRQ
4FGL J0805.2-0110	-0.03286	-0.02466	FSRQ	
4FGL J0812.5+0711	-0.03296	-0.02787	BCU	FSRQ
4FGL J0812.6+2821	-0.03524	-0.02323	BCU	BLL
4FGL J0812.9+5555	-0.03348	-0.0205	BLL	
4FGL J0814.2-1013	-0.03638	-0.02606	BLL	
4FGL J0814.5-2642	-0.03417	-0.03441	BCU	BLL
4FGL J0816.7-2420	-0.03422	-0.03416	BCU	FSRQ
4FGL J0820.9+2353	-0.03455	-0.03388	BLL	
4FGL J0828.6-0747	-0.03482	-0.02635	BCU	FSRQ
4FGL J0831.8+0429	-0.03428	-0.0236	BLL	
4FGL J0835.7+0936	-0.03507	-0.03432	BLL	
4FGL J0839.4+1803	-0.03485	-0.03406	BLL	
4FGL J0843.0-0853	-0.0336	-0.03347	BCU	FSRQ
4FGL J0845.4+0442	-0.03348	-0.02231	BCU	BLL
4FGL J0858.0-3130	-0.03561	-0.02651	BLL	
4FGL J0902.4+2051	-0.03547	-0.03471	BLL	
4FGL J0920.9-2256	-0.03343	-0.02727	BLL	
4FGL J0929.3+5014	-0.03577	-0.03349	BLL	
4FGL J0930.3+8612	-0.03496	-0.03366	BLL	
4FGL J0931.2-8533	-0.03296	-0.02316	BCU	FSRQ
4FGL J0947.9+1121	-0.03492	-0.03199	BLL	

Continued on next page

Table 9.2 – Continued from previous page

Source name (1)	x-axis (2)	y-axis (3)	Class (4)	Class in Fan 2022 (5)
4FGL J1002.8+2140	-0.03554	-0.03633	BCU	FSRQ
4FGL J1008.7-2909	-0.03397	-0.02555	BCU	FSRQ
4FGL J1011.3-0427	-0.03374	-0.03155	FSRQ	
4FGL J1015.6+5553	-0.03369	-0.03376	FSRQ	
4FGL J1018.4+3540	-0.03438	-0.02499	FSRQ	
4FGL J1028.3+3108	-0.03538	-0.03388	BLL	
4FGL J1030.6-2028	-0.03489	-0.0337	BLL	
4FGL J1032.7+6624	-0.03498	-0.03543	BLL	
4FGL J1033.5+4221	-0.0333	-0.0283	BLL	

Continued on next page

Table 9.2 – Continued from previous page

Source name (1)	x-axis (2)	y-axis (3)	Class (4)	Class in Fan 2022 (5)
4FGL J1034.0-2547	-0.03314	-0.03153	BCU	BLL
4FGL J1038.5+3926	-0.03389	-0.03305	BLL	
4FGL J1041.5+0607	-0.03469	-0.03198	FSRQ	
4FGL J1041.7+3902	-0.03322	-0.03185	BLL	
4FGL J1051.9+0103	-0.03522	-0.03352	BLL	
4FGL J1058.6+5627	-0.03587	-0.03017	BLL	
4FGL J1059.5+2057	-0.03529	-0.02287	FSRQ	
4FGL J1100.3+4020	-0.03473	-0.03354	BLL	
4FGL J1106.2-1048	-0.03495	-0.03151	BCU	BLL
4FGL J1118.2-4634	-0.03499	-0.02993	FSRQ	
4FGL J1124.6-0809	-0.03501	-0.02187	BCU	BLL
4FGL J1124.9+4934	-0.03359	-0.0338	BLL	
4FGL J1127.8+3618	-0.03402	-0.02493	FSRQ	
4FGL J1128.8+3757	-0.03474	-0.0307	BLL	
4FGL J1130.5-7801	-0.03594	-0.02704	BLL	
4FGL J1131.1-0944	-0.03703	-0.03137	BCU	BLL
4FGL J1136.4+6736	-0.03614	-0.02792	BLL	
4FGL J1136.4+7009	-0.03654	-0.03709	BLL	
4FGL J1143.9+1558	-0.03356	-0.03094	BLL	
4FGL J1147.2-2627	-0.03291	-0.02303	BCU	FSRQ
4FGL J1148.6+1841	-0.03384	-0.02917	BLL	
4FGL J1149.1+2819	-0.03388	-0.03048	BCU	BLL
4FGL J1155.8+6137	-0.03418	-0.03294	BLL	
4FGL J1201.1-0332	-0.0337	-0.02915	BCU	BLL
4FGL J1202.5-0528	-0.03634	-0.03133	FSRQ	
4FGL J1203.4-3925	-0.03643	-0.02704	BLL	
4FGL J1209.7+2548	-0.0329	-0.02787	FSRQ	
4FGL J1213.3-2618	-0.03325	-0.02524	BLL	
4FGL J1221.3+3010	-0.03628	-0.02934	BLL	
4FGL J1221.5+2814	-0.03537	-0.03105	BLL	
4FGL J1223.5+0818	-0.03535	-0.03039	BLL	
4FGL J1223.9+7954	-0.03443	-0.02176	BLL	
4FGL J1225.5-2851	-0.03366	-0.03132	BCU	FSRQ
4FGL J1231.6+6415	-0.03335	-0.02824	BLL	
4FGL J1232.5-3720	-0.03671	-0.02672	BCU	BLL
4FGL J1233.7-0144	-0.03569	-0.02624	BLL	

Continued on next page

Table 9.2 – Continued from previous page

Source name (1)	x-axis (2)	y-axis (3)	Class (4)	Class in Fan 2022 (5)
4FGL J1238.5-1201	-0.03356	-0.02345	FSRQ	
4FGL J1245.1+5709	-0.03309	-0.02866	BLL	
4FGL J1246.7-2548	-0.03237	-0.02328	FSRQ	
4FGL J1251.3-0201	-0.03488	-0.03163	BCU	BLL
4FGL J1254.2-2205	-0.03602	-0.03554	BCU	BLL
4FGL J1259.7-3223	-0.03272	-0.0255	BLL	
4FGL J1259.8-3749	-0.03446	-0.02526	BLL	
4FGL J1302.8+5748	-0.03327	-0.02385	BLL	
4FGL J1310.6+2449	-0.03675	-0.0375	BLL	
4FGL J1311.0+0034	-0.03507	-0.02549	BLL	
4FGL J1315.9-6349c	-0.03405	-0.02576	BCU	FSRQ
4FGL J1321.1+2216	-0.034	-0.0273	FSRQ	
4FGL J1322.6-1617	-0.03412	-0.03451	BCU	BLL
4FGL J1324.9+4748	-0.03419	-0.03383	FSRQ	
4FGL J1327.0+3154	-0.03452	-0.02631	BCU	BLL
4FGL J1327.8+2522	-0.03411	-0.0292	BLL	
4FGL J1336.2+2320	-0.03338	-0.02929	BLL	
4FGL J1339.0-2400	-0.03617	-0.02749	BCU	BLL
4FGL J1340.1+3857	-0.03321	-0.02225	BCU	FSRQ
4FGL J1354.8-1041	-0.03485	-0.02575	FSRQ	
4FGL J1400.6-5605	-0.03268	-0.02812	BCU	FSRQ
4FGL J1410.1+0202	-0.03352	-0.02568	BLL	
4FGL J1427.4-1823	-0.03549	-0.03637	BCU	BLL
4FGL J1433.7-7304	-0.0305	-0.02153	BCU	BLL
4FGL J1438.0-3128	-0.03294	-0.02179	FSRQ	
4FGL J1439.3+3932	-0.03644	-0.02889	BLL	
4FGL J1442.0+4348	-0.03353	-0.02275	BLL	
4FGL J1442.2+0622	-0.03488	-0.03321	BCU	FSRQ
4FGL J1446.7+1719	-0.03324	-0.02587	FSRQ	
4FGL J1451.5+1415	-0.03476	-0.02335	BCU	FSRQ
4FGL J1503.7-1540	-0.03413	-0.02694	BLL	
4FGL J1509.7+5556	-0.03351	-0.03342	BLL	
4FGL J1511.8-0513	-0.03487	-0.02827	BLL	
4FGL J1514.7-3617	-0.03276	-0.02968	BCU	FSRQ
4FGL J1516.9+1934	-0.03335	-0.02306	BLL	
4FGL J1517.7+6525	-0.03592	-0.02736	BLL	
4FGL J1529.7+6733	-0.03309	-0.02317	BCU	BLL
4FGL J1530.5-3026	-0.03263	-0.02997	BCU	BLL
4FGL J1533.2+3416	-0.03494	-0.02632	BLL	
4FGL J1535.0+5320	-0.03436	-0.03429	BLL	
4FGL J1558.8+5625	-0.03418	-0.02668	BLL	

Continued on next page

Table 9.2 – Continued from previous page

Source name (1)	x-axis (2)	y-axis (3)	Class (4)	Class in Fan 2022 (5)
4FGL J1608.0-2038	-0.03636	-0.035	BLL	
4FGL J1608.7+1029	-0.03225	-0.02562	FSRQ	
4FGL J1612.4-3100	-0.0348	-0.03573	BLL	
4FGL J1616.6+4630	-0.03609	-0.03529	FSRQ	
4FGL J1616.7+4107	-0.03351	-0.02325	BLL	
4FGL J1623.4+0858	-0.03247	-0.02797	BCU	BLL
4FGL J1639.2+4129	-0.03539	-0.02587	FSRQ	
4FGL J1640.3+6850	-0.03626	-0.03833	BCU	BLL
4FGL J1648.0+2221	-0.0356	-0.02663	BCU	FSRQ
4FGL J1649.4+5235	-0.03589	-0.03779	BLL	
4FGL J1656.0+2047	-0.03339	-0.0296	BCU	FSRQ
4FGL J1703.6-6213	-0.03286	-0.02955	FSRQ	
4FGL J1717.5-8114	-0.0364	-0.02835	BLL	
4FGL J1740.0+4737	-0.0345	-0.03264	FSRQ	
4FGL J1742.5+5944	-0.03513	-0.03103	BLL	
4FGL J1745.6+3950	-0.03399	-0.02175	BLL	
4FGL J1756.9+1531	-0.03437	-0.0272	BCU	BLL
4FGL J1814.0+3828	-0.03411	-0.01953	BLL	
4FGL J1837.3+1052	-0.03279	-0.03069	BCU	FSRQ
4FGL J1909.5+3511	-0.03382	-0.02533	BCU	BLL
4FGL J1918.2-4111	-0.03597	-0.02512	BLL	
4FGL J1924.8-2914	-0.03363	-0.03026	FSRQ	
4FGL J1931.1+0937	-0.03591	-0.03285	BLL	
4FGL J1945.1-4007	-0.03372	-0.03001	BCU	BLL
4FGL J1950.0-7321	-0.03309	-0.02367	BCU	BLL
4FGL J1955.1-1604	-0.03701	-0.02551	BLL	
4FGL J2017.5-4113	-0.03456	-0.0328	BCU	BLL
4FGL J2021.9+0629	-0.03376	-0.0271	BLL	
4FGL J2024.7-2328	-0.03277	-0.02797	BCU	BLL
4FGL J2040.1-4621	-0.03326	-0.03287	BCU	BLL
4FGL J2040.5-1705	-0.0339	-0.0195	BCU	FSRQ
4FGL J2053.8+2922	-0.03562	-0.03573	BLL	
4FGL J2104.7+0108	-0.03371	-0.0297	BCU	BLL
4FGL J2115.2+1218	-0.036	-0.02997	BLL	
4FGL J2131.5-0916	-0.03646	-0.02864	BLL	
4FGL J2134.2-0154	-0.03373	-0.03162	BLL	
4FGL J2134.3-6511	-0.0339	-0.02375	BCU	BLL
4FGL J2146.5-1344	-0.0359	-0.02806	BLL	
4FGL J2151.8-3027	-0.03205	-0.01617	FSRQ	
4FGL J2152.0-1205	-0.03331	-0.02061	BLL	
4FGL J2153.1-0041	-0.03638	-0.03911	BLL	

Continued on next page

Table 9.2 – Continued from previous page

Source name (1)	x-axis (2)	y-axis (3)	Class (4)	Class in Fan 2022 (5)
4FGL J2158.5-7550	-0.03427	-0.0309	BCU	BLL
4FGL J2158.8-3013	-0.03502	-0.02844	BLL	
4FGL J2204.5+3634	-0.03553	-0.02239	BLL	
4FGL J2207.1+4316	-0.03253	-0.02741	BCU	BLL
4FGL J2211.4-7040	-0.03657	-0.02979	BLL	
4FGL J2222.8+1209	-0.03405	-0.02072	BCU	FSRQ
4FGL J2225.5-1114	-0.03319	-0.01832	BLL	
4FGL J2226.6+0210	-0.03479	-0.02176	BCU	BLL
4FGL J2230.9-7815	-0.03373	-0.03147	FSRQ	
4FGL J2236.3+2828	-0.03328	-0.02517	FSRQ	
4FGL J2239.5-2440	-0.03101	-0.02464	BCU	BLL
4FGL J2241.1-4122	-0.03357	-0.02675	BCU	BLL
4FGL J2243.9+2021	-0.03591	-0.02427	BLL	
4FGL J2246.5-1204	-0.03393	-0.02195	FSRQ	
4FGL J2325.2+3957	-0.03471	-0.02189	BLL	
4FGL J2325.6+1644	-0.03211	-0.0287	BLL	
4FGL J2328.3-4036	-0.03439	-0.03093	FSRQ	
4FGL J2329.0+0832	-0.03374	-0.02114	FSRQ	
4FGL J2330.4+1230	-0.03309	-0.03269	BCU	BLL
4FGL J2336.5-7622	-0.0352	-0.02492	BLL	
4FGL J2339.2-7403	-0.03666	-0.03107	BLL	
4FGL J2359.3-2049	-0.03557	-0.02902	BLL	

Publications

1. **J. T. Zhu**, C. Lin, H. B. Xiao, J. H. FAN*, BASTIERI, and G. G. WANG. Exploring TeV Candidates of *Fermi* Blazars through Machine Learning. *The Astrophysical Journal*, 2023, 950(2): 123.
2. Hubing XIAO, **Jingtian ZHU***, Liping Fu, Shaohua ZHANG, and Junhui FAN*. The radio dichotomy of active galactic nuclei. *Publications of the Astronomical Society of Japan*, 2022, 74(2): 239-246.
3. H. B. Xiao, **J. T. Zhu***, J. H. Fan*, Z. Y. Pei, Z. J. Luo and S. H. Zhang. The jet apparent motion and central engine study of *Fermi* blazars. *Monthly Notices of the Royal Astronomical Society*. 2022, 517(3): 4202–4212.
4. Jingtian Zhu, Haitao Cao*, Hubing Xiao, Zhiyuan Pei, JunHui Fan*, and Denis Bastieri*. Chasing the neutrino blazar candidates. *The Astrophysical Journal Supplement Series*, is under review.

Acknowledgements

First and foremost, I wish to express my deepest gratitude to my wife, who has silently stood by my side, offering unwavering support throughout the many years of this journey. Her strength and love have been my pillars, and I am eternally grateful for the joy our lovely son has brought into our lives. I owe a profound debt of gratitude to my parents, who, during my years in Padova, have shouldered many of the family responsibilities.

I am immensely thankful to my supervisor, Prof. Denis Bastieri, for his rigorous scientific guidance over the past three years of my Ph.D. studies. His mentorship has been instrumental in deepening my understanding of AGNs. I am equally indebted to Prof. Junhui Fan, my co-supervisor during my master's studies, whose guidance was pivotal in introducing me to the world of astronomical research.

I would be remiss if I did not acknowledge the contributions of my esteemed colleagues and friends who have been instrumental in my scientific journey. Special thanks to Giovanni LA MURA from the Laboratory of Instrumentation and Experimental Particle Physics, Portugal; Hubing Xiao from Shanghai Normal University, China; Zhiyuan Pei from Guangzhou University, China; Xuhong Ye from the University of Padova, Italy; and Gege Wang from Northeastern University, China. Their insights, collaboration, and camaraderie have enriched my academic experience.

I extend my heartfelt appreciation to the University of Padua for providing an enriching environment and a robust scientific research platform that has been pivotal for my Ph.D. studies. I am also grateful to Guangzhou University for its joint Ph.D. training program, which afforded me the unique opportunity to pursue my doctorate at the University of Padua. This has undoubtedly been the most exhilarating chapter of my academic journey.

We extend our heartfelt gratitude to the esteemed reviewer for their invaluable comments and constructive suggestions that have significantly enhanced the quality of our work. Our research has been generously supported by several institutions and grants, including the National Natural Science Foundation of China (NSFC U1831119, NSFC U2031201, NSFC 11733001, NSFC 12203034, NSFC 11933002), Guangdong Major Project of Basic and Applied Basic Research (2019B030302001), and the Shanghai Science and Technology Fund (22YF1431500, 20070502400). We are also indebted to the Scientific and Technological Cooperation Projects (2020-2023) between the People's Republic of China and the Republic of Bulgaria.

Special acknowledgment goes to key astrophysical initiatives and observatories that have been instrumental in our research, including the Astrophysics Key Subjects of Guangdong Province and Guangzhou City, the Large High Altitude Air Shower Observatory collaboration, the Cherenkov Telescope Array, the High Altitude Water Cherenkov Observatory, the Major Atmospheric Gamma Imaging Cherenkov Telescopes, the Very Energetic Radiation Imaging Telescope Array System, and the UK Swift Science Data Centre at the University of Leicester.

The key laboratory for astrophysics in Shanghai has been a cornerstone of our research endeavors. Individual acknowledgments are extended to L.P.F for the support from STCSM grants 18590780100, 19590780100, SMEC Innovation Program 2019-01-07-00-02-E00032, and Shuguang Program 19SG41. We are grateful for the support from the Natural Science Foundation of Shanghai (20ZR1473600, 22YF1431500). contributions of Junhui Fan were made possible by the NSFC (NSFC U2031201, NSFC 11733001) and the science research grants from the China Manned Space Project with NO. CMS-CSST-2021-A06. Z.J.L's efforts were bolstered by the Shanghai Science and Technology Fund (Grant No. 20070502400). We also express our gratitude to the MOJAVE team for granting us access to their invaluable kinematic data.

Bibliography

1977, Evolution of galaxies and stellar populations

1994, SUGI 19 : proceedings of the nineteenth annual SAS Users Group International conference ; Dallas, Texas, April 10 - 13, 1994 (Cary, NC: SAS Institute;). <https://www.tib.eu/de/suchen/id/TIBKAT%3A197848265>

1998, Black-hole accretion disks

Aartsen, M. G., Abbasi, R., Abdou, Y., et al. 2013, Phys. Rev. Lett., 111, 021103, doi: [10.1103/PhysRevLett.111.021103](https://doi.org/10.1103/PhysRevLett.111.021103)

Aartsen, M. G., Abraham, K., Ackermann, M., et al. 2016, ApJ, 833, 3, doi: [10.3847/0004-637X/833/1/3](https://doi.org/10.3847/0004-637X/833/1/3)

Aartsen, M. G., Ackermann, M., Adams, J., et al. 2017a, ApJ, 843, 112, doi: [10.3847/1538-4357/aa7569](https://doi.org/10.3847/1538-4357/aa7569)

Aartsen, M. G., Abraham, K., Ackermann, M., et al. 2017b, ApJ, 835, 45, doi: [10.3847/1538-4357/835/1/45](https://doi.org/10.3847/1538-4357/835/1/45)

Aartsen, M. G., Ackermann, M., Adams, J., et al. 2017a, The Astrophysical Journal, 843, 112, doi: [10.3847/1538-4357/aa7569](https://doi.org/10.3847/1538-4357/aa7569)

Aartsen, M. G., Abraham, K., Ackermann, M., et al. 2017b, The Astrophysical Journal, 835, 45, doi: [10.3847/1538-4357/835/1/45](https://doi.org/10.3847/1538-4357/835/1/45)

Aartsen, M. G., Ackermann, M., Adams, J., et al. 2020, Phys. Rev. Lett., 124, 051103, doi: [10.1103/PhysRevLett.124.051103](https://doi.org/10.1103/PhysRevLett.124.051103)

Aartsen, M. G., Abbasi, R., Ackermann, M., et al. 2021, Journal of Physics G Nuclear Physics, 48, 060501, doi: [10.1088/1361-6471/abbd48](https://doi.org/10.1088/1361-6471/abbd48)

Abbasi, R., Ackermann, M., Adams, J., et al. 2021, ApJ, 920, L45, doi: [10.3847/2041-8213/ac2c7b](https://doi.org/10.3847/2041-8213/ac2c7b)

Abdo, A. A., Ackermann, M., Arimoto, M., et al. 2009, Science, 323, doi: [10.1126/science.1169101](https://doi.org/10.1126/science.1169101)

Abdo, A. A., Ackermann, M., Ajello, M., et al. 2009, ApJ, 700, 597, doi: [10.1088/0004-637X/700/1/597](https://doi.org/10.1088/0004-637X/700/1/597)

—. 2010a, ApJ, 715, 429, doi: [10.1088/0004-637X/715/1/429](https://doi.org/10.1088/0004-637X/715/1/429)

- Abdo, A. A., Ackermann, M., Agudo, I., et al. 2010b, *ApJ*, 716, 30, doi: 10.1088/0004-637X/716/1/30
- Abdo, A. A., Ackermann, M., Ajello, M., et al. 2010c, *ApJS*, 188, 405, doi: 10.1088/0067-0049/188/2/405
- Abdollahi, S., Acero, F., Ackermann, M., et al. 2020, *The Astrophysical Journal Supplement Series*, 247, 33, doi: 10.3847/1538-4365/ab6bcb
- Abdollahi, S., Acero, F., Ackermann, M., et al. 2020, *ApJS*, 247, 33, doi: 10.3847/1538-4365/ab6bcb
- Abdollahi, S., Acero, F., Baldini, L., et al. 2022, *ApJS*, 260, 53, doi: 10.3847/1538-4365/ac6751
- Abeysekara, A. U., Albert, A., Alfaro, R., et al. 2017a, *ApJ*, 841, 100, doi: 10.3847/1538-4357/aa729e
- . 2017b, *ApJ*, 843, 39, doi: 10.3847/1538-4357/aa7555
- . 2017c, *ApJ*, 843, 39, doi: 10.3847/1538-4357/aa7555
- Abeysekara, A. U., Albert, A., Alfaro, R., et al. 2017, *The Astrophysical Journal*, 843, 40, doi: 10.3847/1538-4357/aa7556
- Abramowicz, M. A., Chen, X., Kato, S., Lasota, J.-P., & Regev, O. 1995, *ApJ*, 438, L37, doi: 10.1086/187709
- Abramowicz, M. A., Karas, V., & Lanza, A. 1998, *A&A*, 331, 1143, doi: 10.48550/arXiv.astro-ph/9712245
- Abramowski, A., Acero, F., Aharonian, F., et al. 2013, *Phys. Rev. D*, 88, 102003, doi: 10.1103/PhysRevD.88.102003
- Abramowski, A., Aharonian, F., Ait Benkhali, F., et al. 2015, *ApJ*, 802, 65, doi: 10.1088/0004-637X/802/1/65
- Abu-Mostafa, Y. S., Magdon-Ismail, M., & Lin, H. 2012, *Learning from Data: A Short Course (AML-Book.com)*. <https://books.google.nl/books?id=iZUzMwECAAJ>
- Acciari, V. A., Aliu, E., Beilicke, M., et al. 2008, *ApJ*, 684, L73, doi: 10.1086/592244
- Acciari, V. A., Aliu, E., Arlen, T., et al. 2011, *ApJ*, 738, 169, doi: 10.1088/0004-637X/738/2/169
- Acero, F., Ackermann, M., Ajello, M., et al. 2015, *The Astrophysical Journal Supplement Series*, 218, 23, doi: 10.1088/0067-0049/218/2/23
- Acero, F., Ackermann, M., Ajello, M., et al. 2015, *ApJS*, 218, 23, doi: 10.1088/0067-0049/218/2/23
- Achterberg, A., Gallant, Y. A., Kirk, J. G., & Guthmann, A. W. 2001, *MNRAS*, 328, 393, doi: 10.1046/j.1365-8711.2001.04851.x
- Ackermann, M., Ajello, M., Allafort, A., et al. 2011, *ApJ*, 743, 171, doi: 10.1088/0004-637X/743/2/171

- Ackermann, M., Ajello, M., Atwood, W. B., et al. 2015a, *ApJ*, 810, 14, doi: 10.1088/0004-637X/810/1/14
- Ackermann, M., Albert, A., Anderson, B., et al. 2015b, *Phys. Rev. Lett.*, 115, 231301, doi: 10.1103/PhysRevLett.115.231301
- Ackermann, M., Ajello, M., Albert, A., et al. 2015c, *ApJ*, 813, L41, doi: 10.1088/2041-8205/813/2/L41
- Ackermann, M., Ajello, M., Atwood, W. B., et al. 2015d, *ApJ*, 810, 14, doi: 10.1088/0004-637X/810/1/14
- Adrián-Martínez, S., Ageron, M., Aharonian, F., et al. 2016, *Journal of Physics G Nuclear Physics*, 43, 084001, doi: 10.1088/0954-3899/43/8/084001
- Agarwal, A. 2023, *ApJ*, 946, 109, doi: 10.3847/1538-4357/acbdfa
- Aharonian, F., Akhperjanian, A., Beilicke, M., et al. 2003, *A&A*, 406, L9, doi: 10.1051/0004-6361:20030838
- Aharonian, F., Akhperjanian, A. G., Aye, K. M., et al. 2004, *Astroparticle Physics*, 22, 109, doi: 10.1016/j.astropartphys.2004.06.006
- Aharonian, F., Akhperjanian, A. G., Bazer-Bachi, A. R., et al. 2007a, *ApJ*, 664, L71, doi: 10.1086/520635
- Aharonian, F., Akhperjanian, A. G., Barres de Almeida, U., et al. 2007b, *A&A*, 473, L25, doi: 10.1051/0004-6361:20078412
- . 2007c, *A&A*, 475, L9, doi: 10.1051/0004-6361:20078462
- Aharonian, F., Akhperjanian, A. G., Anton, G., et al. 2009a, *A&A*, 502, 749, doi: 10.1051/0004-6361/200912128
- . 2009b, *ApJ*, 696, L150, doi: 10.1088/0004-637X/696/2/L150
- Aharonian, F. A., Akhperjanian, A. G., Barrio, J. A., et al. 2001, *A&A*, 366, 62, doi: 10.1051/0004-6361:20000098
- Ahmad, J., Farman, H., & Jan, Z. 2019, *Deep Learning Methods and Applications* (Singapore: Springer Singapore), 31–42, doi: 10.1007/978-981-13-3459-7_3
- Ajello, M., Gasparrini, D., Sánchez-Conde, M., et al. 2015, *ApJ*, 800, L27, doi: 10.1088/2041-8205/800/2/L27
- Ajello, M., Gasparrini, D., Sánchez-Conde, M., et al. 2015, *The Astrophysical Journal Letters*, 800, L27, doi: 10.1088/2041-8205/800/2/L27
- Ajello, M., Atwood, W. B., Baldini, L., et al. 2017, *ApJS*, 232, 18, doi: 10.3847/1538-4365/aa8221
- Ajello, M., Angioni, R., Axelsson, M., et al. 2020, *The Astrophysical Journal*, 892, 105, doi: 10.3847/1538-4357/ab791e

- Ajello, M., Angioni, R., Axelsson, M., et al. 2020, *ApJ*, 892, 105, doi: 10.3847/1538-4357/ab791e
- Ajello, M., Baldini, L., Ballet, J., et al. 2022, *ApJS*, 263, 24, doi: 10.3847/1538-4365/ac9523
- Aleksić, J., Antonelli, L. A., Antoranz, P., et al. 2011, *ApJ*, 730, L8, doi: 10.1088/2041-8205/730/1/L8
- Aleksić, J., Ansoldi, S., Antonelli, L. A., et al. 2016, *Astroparticle Physics*, 72, 76, doi: 10.1016/j.astropartphys.2015.02.005
- Aliu, E., Aune, T., Beilicke, M., et al. 2011, *ApJ*, 742, 127, doi: 10.1088/0004-637X/742/2/127
- Aliu, E., Archambault, S., Arlen, T., et al. 2014, *ApJ*, 797, 89, doi: 10.1088/0004-637X/797/2/89
- Aliu, E., Archer, A., Aune, T., et al. 2015, *ApJ*, 799, 7, doi: 10.1088/0004-637X/799/1/7
- Alpaydin, E. 2014, *Introduction to Machine Learning*, 3rd edn., Adaptive Computation and Machine Learning (Cambridge, MA: MIT Press)
- Amenomori, M., Ayabe, S., Cui, S. W., et al. 2003, *ApJ*, 598, 242, doi: 10.1086/378350
- Amorim, R. 2009, *Inf. Process. Manage.*, 45, 490
- An, T. 2019, *Science China Physics, Mechanics & Astronomy*, 62, 1. <https://api.semanticscholar.org/CorpusID:119041065>
- Andreon, S., Gargiulo, G., Longo, G., Tagliaferri, R., & Capuano, N. 2000, *MNRAS*, 319, 700, doi: 10.1046/j.1365-8711.2000.03700.x
- Ansoldi, S., Antonelli, L. A., Arcaro, C., et al. 2018, *ApJ*, 863, L10, doi: 10.3847/2041-8213/aad083
- Antonucci, R. 1993, *ARA&A*, 31, 473, doi: 10.1146/annurev.aa.31.090193.002353
- Arbeiter, C., Pohl, M., & Schlickeiser, R. 2002, *A&A*, 386, 415, doi: 10.1051/0004-6361:20020221
- Archambault, S., Arlen, T., Aune, T., et al. 2013, *ApJ*, 776, 69, doi: 10.1088/0004-637X/776/2/69
- Archambault, S., Aune, T., Behera, B., et al. 2014, *ApJ*, 785, L16, doi: 10.1088/2041-8205/785/1/L16
- Arlen, T., Aune, T., Beilicke, M., et al. 2013, *ApJ*, 762, 92, doi: 10.1088/0004-637X/762/2/92
- Arsioli, B., Chang, Y.-L., & Musiimenta, B. 2020, *Monthly Notices of the Royal Astronomical Society*, 493, 2438–2451, doi: 10.1093/mnras/staa368
- Atoyan, A., & Dermer, C. D. 2001a, *Phys. Rev. Lett.*, 87, 221102, doi: 10.1103/PhysRevLett.87.221102
- . 2001b, *Phys. Rev. Lett.*, 87, 221102, doi: 10.1103/PhysRevLett.87.221102
- Atoyan, A. M., & Dermer, C. D. 2003, *ApJ*, 586, 79, doi: 10.1086/346261

- Atwood, W. B., Abdo, A. A., Ackermann, M., et al. 2009a, *ApJ*, 697, 1071, doi: [10.1088/0004-637X/697/2/1071](https://doi.org/10.1088/0004-637X/697/2/1071)
- . 2009b, *ApJ*, 697, 1071, doi: [10.1088/0004-637X/697/2/1071](https://doi.org/10.1088/0004-637X/697/2/1071)
- Baldi, R. D., Capetti, A., & Giovannini, G. 2015, *A&A*, 576, A38, doi: [10.1051/0004-6361/201425426](https://doi.org/10.1051/0004-6361/201425426)
- Baldwin, J. A., Phillips, M. M., & Terlevich, R. 1981, *PASP*, 93, 5, doi: [10.1086/130766](https://doi.org/10.1086/130766)
- Ball, N. M., Brunner, R. J., Myers, A. D., & Tchong, D. 2006, *ApJ*, 650, 497, doi: [10.1086/507440](https://doi.org/10.1086/507440)
- Ballet, J., Bruel, P., Burnett, T. H., Lott, B., & The Fermi-LAT collaboration. 2023, arXiv e-prints, arXiv:2307.12546, doi: [10.48550/arXiv.2307.12546](https://doi.org/10.48550/arXiv.2307.12546)
- Ballet, J., Burnett, T. H., Digel, S. W., & Lott, B. 2020, arXiv e-prints, arXiv:2005.11208. <https://arxiv.org/abs/2005.11208>
- Baloković, M., Smolčić, V., Ivezić, Ž., et al. 2012, *ApJ*, 759, 30, doi: [10.1088/0004-637X/759/1/30](https://doi.org/10.1088/0004-637X/759/1/30)
- Bambi, C. 2016, *Astrophysics of Black Holes: From Fundamental Aspects to Latest Developments*, Vol. 440, doi: [10.1007/978-3-662-52859-4](https://doi.org/10.1007/978-3-662-52859-4)
- Bartoli, B., Bernardini, P., Bi, X. J., et al. 2011, *ApJ*, 734, 110, doi: [10.1088/0004-637X/734/2/110](https://doi.org/10.1088/0004-637X/734/2/110)
- . 2012, *ApJ*, 758, 2, doi: [10.1088/0004-637X/758/1/2](https://doi.org/10.1088/0004-637X/758/1/2)
- Barvainis, R., Lehár, J., Birkinshaw, M., Falcke, H., & Blundell, K. M. 2005, *ApJ*, 618, 108, doi: [10.1086/425859](https://doi.org/10.1086/425859)
- Bazell, D., & Peng, Y. 1998, *ApJS*, 116, 47, doi: [10.1086/313098](https://doi.org/10.1086/313098)
- Beckmann, V., & Shrader, C. R. 2012, *Active Galactic Nuclei*
- Belli, F., Andreanelli, L., Angelini, F., et al. 2007, *Nuclear Instruments and Methods in Physics Research A*, 570, 276, doi: [10.1016/j.nima.2006.09.062](https://doi.org/10.1016/j.nima.2006.09.062)
- BELLMAN, R. 1961, *Adaptive Control Processes: A Guided Tour* (Princeton University Press). <http://www.jstor.org/stable/j.ctt183ph6v>
- Bellman, R., Corporation, R., & Collection, K. M. R. 1957, *Dynamic Programming*, Rand Corporation research study (Princeton University Press). <https://books.google.nl/books?id=wdtoPwAACAAJ>
- Bengio, Y. 2009
- Bengio, Y., Courville, A., & Vincent, P. 2013a, *IEEE transactions on pattern analysis and machine intelligence*, 35, 1798, doi: [10.1109/TPAMI.2013.50](https://doi.org/10.1109/TPAMI.2013.50)
- . 2013b, *IEEE transactions on pattern analysis and machine intelligence*, 35, 1798—1828, doi: [10.1109/tpami.2013.50](https://doi.org/10.1109/tpami.2013.50)

- Bengio, Y., Lamblin, P., Popovici, D., & Larochelle, H. 2006, in *Advances in Neural Information Processing Systems*, ed. B. Schölkopf, J. Platt, & T. Hoffman, Vol. 19 (MIT Press). https://proceedings.neurips.cc/paper_files/paper/2006/file/5da713a690c067105aeb2fae32403405-Paper.pdf
- Berenji, B., & Siegal-Gaskins, J. M. 2014, in *AAS/High Energy Astrophysics Division*, Vol. 14, AAS/High Energy Astrophysics Division #14, 107.06
- Best, P. N., & Heckman, T. M. 2012, *MNRAS*, 421, 1569, doi: 10.1111/j.1365-2966.2012.20414.x
- Beyer, K., Goldstein, J., Ramakrishnan, R., & Shaft, U. 1999, in *Database Theory — ICDT'99*, ed. C. Beeri & P. Buneman (Berlin, Heidelberg: Springer Berlin Heidelberg), 217–235
- Bian, W., & Zhao, Y. 2003, *ApJ*, 591, 733, doi: 10.1086/375414
- Bishop, C. M. 2007, *Pattern Recognition and Machine Learning (Information Science and Statistics)*, 1st edn. (Springer). <http://www.amazon.com/Pattern-Recognition-Learning-Information-Statistics/dp/0387310738%3FSubscriptionId%3D13CT5CVB80YFWJEPWS02%26tag%3Dws%26linkCode%3Dxm2%26camp%3D2025%26creative%3D165953%26creativeASIN%3D0387310738>
- Biteau, J., & Williams, D. A. 2015, *ApJ*, 812, 60, doi: 10.1088/0004-637X/812/1/60
- Blandford, R. D., & Begelman, M. C. 1999, *MNRAS*, 303, L1, doi: 10.1046/j.1365-8711.1999.02358.x
- Blandford, R. D., & Königl, A. 1979, *Astrophys. Lett.*, 20, 15
- Blandford, R. D., & Königl, A. 1979, *ApJ*, 232, 34, doi: 10.1086/157262
- Blandford, R. D., & Payne, D. G. 1982, *MNRAS*, 199, 883, doi: 10.1093/mnras/199.4.883
- Blandford, R. D., & Znajek, R. L. 1977, *MNRAS*, 179, 433, doi: 10.1093/mnras/179.3.433
- Błażejowski, M., Sikora, M., Moderski, R., & Madejski, G. M. 2000a, *ApJ*, 545, 107, doi: 10.1086/317791
- . 2000b, *ApJ*, 545, 107, doi: 10.1086/317791
- . 2000c, *ApJ*, 545, 107, doi: 10.1086/317791
- Błażejowski, M., Blaylock, G., Bond, I. H., et al. 2005, *ApJ*, 630, 130, doi: 10.1086/431925
- Bloom, S. D., & Marscher, A. P. 1996a, *ApJ*, 461, 657, doi: 10.1086/177092
- . 1996b, *ApJ*, 461, 657, doi: 10.1086/177092
- Boettcher, M., Harris, D. E., & Krawczynski, H. 2012, *Relativistic Jets from Active Galactic Nuclei*
- Borne, K. D. 2008, *Astronomische Nachrichten*, 329, 255, doi: 10.1002/asna.200710946
- Boser, B. E., Guyon, I. M., & Vapnik, V. N. 1992, in *Proceedings of the 5th Annual Workshop on Computational Learning Theory (COLT'92)*, ed. D. Haussler (Pittsburgh, PA, USA: ACM Press), 144–152. <http://doi.acm.org/10.1145/130385.130401>

- Böttcher, M. 2000, in American Institute of Physics Conference Series, Vol. 515, GeV-TeV Gamma Ray Astrophysics Workshop : towards a major atmospheric, ed. B. L. Dingus, M. H. Salamon, & D. B. Kieda, 31–40, doi: [10.1063/1.1291341](https://doi.org/10.1063/1.1291341)
- Böttcher, M., & Chiang, J. 2002, *ApJ*, 581, 127, doi: [10.1086/344155](https://doi.org/10.1086/344155)
- Böttcher, M., & Dermer, C. D. 2002, *ApJ*, 564, 86, doi: [10.1086/324134](https://doi.org/10.1086/324134)
- . 2010, *ApJ*, 711, 445, doi: [10.1088/0004-637X/711/1/445](https://doi.org/10.1088/0004-637X/711/1/445)
- Böttcher, M., Reimer, A., Sweeney, K., & Prakash, A. 2013, *ApJ*, 768, 54, doi: [10.1088/0004-637X/768/1/54](https://doi.org/10.1088/0004-637X/768/1/54)
- Boyd, S., & Vandenberghe, L. 2004a, *Convex Optimization*, doi: [10.1017/CB09780511804441](https://doi.org/10.1017/CB09780511804441)
- . 2004b, *Convex Optimization* (Cambridge University Press). <https://books.google.nl/books?id=IUZdAAAAQBAJ>
- Bozinovski, S. 2020, *Informatica (Slovenia)*, 44
- Bre, F., Gimenez, J., & Fachinotti, V. 2017, *Energy and Buildings*, 158, doi: [10.1016/j.enbuild.2017.11.045](https://doi.org/10.1016/j.enbuild.2017.11.045)
- Breiman, L. 2001a, *Statistical Science*, 16, doi: [10.1214/ss/1009213726](https://doi.org/10.1214/ss/1009213726)
- . 2001b, *Machine Learning*, 45, 5, doi: [10.1023/A:1010933404324](https://doi.org/10.1023/A:1010933404324)
- Bringsjord, S., & Govindarajulu, N. S. 2022, in *The Stanford Encyclopedia of Philosophy*, Fall 2022 edn., ed. E. N. Zalta & U. Nodelman (Metaphysics Research Lab, Stanford University)
- Britzen, S., Vermeulen, R. C., Campbell, R. M., et al. 2008, *A&A*, 484, 119, doi: [10.1051/0004-6361:20077717](https://doi.org/10.1051/0004-6361:20077717)
- Bruel, P., Burnett, T. H., Digel, S. W., et al. 2018, arXiv e-prints, arXiv:1810.11394, doi: [10.48550/arXiv.1810.11394](https://doi.org/10.48550/arXiv.1810.11394)
- Buda, M., Maki, A., & Mazurowski, M. A. 2018, *Neural Networks*, 106, 249, doi: <https://doi.org/10.1016/j.neunet.2018.07.011>
- Buitinck, L., Louppe, G., Blondel, M., et al. 2013, arXiv e-prints, arXiv:1309.0238, doi: [10.48550/arXiv.1309.0238](https://doi.org/10.48550/arXiv.1309.0238)
- Burges, C. J. 1998, *Data Mining and Knowledge Discovery*, 2, 121, doi: [10.1023/A:1009715923555](https://doi.org/10.1023/A:1009715923555)
- Burl, M., Fayyad, U., Perona, P., & Smyth, P. 1994, in *Proceedings of 1st International Conference on Image Processing*, Vol. 3, 236–240 vol.3, doi: [10.1109/ICIP.1994.413852](https://doi.org/10.1109/ICIP.1994.413852)
- Calderone, G., Ghisellini, G., Colpi, M., & Dotti, M. 2013, *MNRAS*, 431, 210, doi: [10.1093/mnras/stt157](https://doi.org/10.1093/mnras/stt157)
- Cao, G., & Wang, J. 2014, *ApJ*, 783, 108, doi: [10.1088/0004-637X/783/2/108](https://doi.org/10.1088/0004-637X/783/2/108)
- Cao, X. 2016, *ApJ*, 833, 30, doi: [10.3847/1538-4357/833/1/30](https://doi.org/10.3847/1538-4357/833/1/30)

- Cao, X., & Jiang, D. R. 1999, *MNRAS*, 307, 802, doi: 10.1046/j.1365-8711.1999.02657.x
- Cao, X., & Lai, D. 2019, *MNRAS*, 485, 1916, doi: 10.1093/mnras/stz580
- Cao, Z., Chen, M.-J., Chen, Song-Zhan Hu, H.-B., et al. 2019, *Chinese Astron. Astrophys.*, 43, 457, doi: 10.1016/j.chinastron.2019.11.001
- Carballo, R., Cofiño, A. S., & González-Serrano, J. I. 2004, *MNRAS*, 353, 211, doi: 10.1111/j.1365-2966.2004.08056.x
- Carballo, R., González-Serrano, J. I., Benn, C. R., & Jiménez-Luján, F. 2008, *MNRAS*, 391, 369, doi: 10.1111/j.1365-2966.2008.13896.x
- Carr, B. J., Kohri, K., Sendouda, Y., & Yokoyama, J. 2010, *Phys. Rev. D*, 81, 104019, doi: 10.1103/PhysRevD.81.104019
- Castelvecchi, D. 2016, *Nature*, 538, 20, doi: 10.1038/538020a
- Cavagnolo, K. W., McNamara, B. R., Nulsen, P. E. J., et al. 2010, *ApJ*, 720, 1066, doi: 10.1088/0004-637X/720/2/1066
- Cavaliere, A., & D'Elia, V. 2002, *ApJ*, 571, 226, doi: 10.1086/339778
- Celotti, A., Padovani, P., & Ghisellini, G. 1997, *MNRAS*, 286, 415, doi: 10.1093/mnras/286.2.415
- Cerruti, M., Benbow, W., Chen, X., et al. 2017, *A&A*, 606, A68, doi: 10.1051/0004-6361/201730799
- Cerruti, M., Zech, A., Boisson, C., et al. 2018, *Monthly Notices of the Royal Astronomical Society: Letters*, 483, L12, doi: 10.1093/mnrasl/sly210
- Cerruti, M., Zech, A., Boisson, C., & Inoue, S. 2015a, *MNRAS*, 448, 910, doi: 10.1093/mnras/stu2691
- . 2015b, *MNRAS*, 448, 910, doi: 10.1093/mnras/stu2691
- Chai, B., Cao, X., & Gu, M. 2012, *ApJ*, 759, 114, doi: 10.1088/0004-637X/759/2/114
- Chandra, P., Yadav, K. K., Rannot, R. C., et al. 2010, *Journal of Physics G Nuclear Physics*, 37, 125201, doi: 10.1088/0954-3899/37/12/125201
- Chandra, P., Rannot, R. C., Yadav, K. K., et al. 2012, *Journal of Physics G Nuclear Physics*, 39, 045201, doi: 10.1088/0954-3899/39/4/045201
- Chang, X.-C., Liu, R.-Y., & Wang, X.-Y. 2015, *ApJ*, 805, 95, doi: 10.1088/0004-637X/805/2/95
- Chang, Y. L., Arsioli, B., Giommi, P., & Padovani, P. 2017, *A&A*, 598, A17, doi: 10.1051/0004-6361/201629487
- Chang, Y.-L., Arsioli, B., Giommi, P., Padovani, P., & Brandt, C. H. 2019, *Astronomy & Astrophysics*, 632, A77, doi: 10.1051/0004-6361/201834526

- Chawla, N., Bowyer, K., Hall, L., & Kegelmeyer, W. 2002, *J. Artif. Intell. Res. (JAIR)*, 16, 321, doi: 10.1613/jair.953
- Chen, T., Chen, H., & Liu, R.-w. 1995, *Neural Networks, IEEE Transactions on*, 6, 25, doi: 10.1109/72.363453
- Chiaro, G., Meyer, M., Di Mauro, M., et al. 2019, *ApJ*, 887, 104, doi: 10.3847/1538-4357/ab46ad
- Chiaro, G., Meyer, M., Mauro, M. D., et al. 2019, *The Astrophysical Journal*, 887, 104, doi: 10.3847/1538-4357/ab46ad
- Chiaro, G., Salvetti, D., La Mura, G., et al. 2016, *MNRAS*, 462, 3180, doi: 10.1093/mnras/stw1830
- Cinelli, L., Marins, M., da Silva, E., & Netto, S. 2021, *Variational Autoencoder*, 111–149, doi: 10.1007/978-3-030-70679-1_5
- Claeskens, J. F., Smette, A., Vandenbulcke, L., & Surdej, J. 2006, *MNRAS*, 367, 879, doi: 10.1111/j.1365-2966.2006.10024.x
- Cohen, M. H., Cannon, W., Purcell, G. H., et al. 1971, *ApJ*, 170, 207, doi: 10.1086/151204
- Cohen, M. H., Lister, M. L., Homan, D. C., et al. 2007, *ApJ*, 658, 232, doi: 10.1086/511063
- Collister, A., Lahav, O., Blake, C., et al. 2007, *MNRAS*, 375, 68, doi: 10.1111/j.1365-2966.2006.11305.x
- Condon, J. J., O'Dell, S. L., Puschell, J. J., & Stein, W. A. 1980, *Nature*, 283, 357, doi: 10.1038/283357a0
- Conselice, C. J. 2003, *ApJS*, 147, 1, doi: 10.1086/375001
- Cormen, T. H., Leiserson, C. E., Rivest, R. L., & Stein, C. 2001, *Introduction to Algorithms*, 2nd edn. (The MIT Press). <http://www.amazon.com/Introduction-Algorithms-Thomas-H-Cormen/dp/0262032937%3FSubscriptionId%3D13CT5CVB80YFWJEPWS02%26tag%3Dws%26linkCode%3Dxm2%26camp%3D2025%26creative%3D165953%26creativeASIN%3D0262032937>
- Costamante, L. 2008, *International Journal of Modern Physics D*, 17, 1449, doi: 10.1142/S0218271808013029
- Costamante, L., Bonnoli, G., Tavecchio, F., et al. 2018, *MNRAS*, 477, 4257, doi: 10.1093/mnras/sty857
- Costamante, L., & Ghisellini, G. 2002, *A&A*, 384, 56, doi: 10.1051/0004-6361:20011749
- Costamante, L., Ghisellini, G., Giommi, P., et al. 2001, *A&A*, 371, 512, doi: 10.1051/0004-6361:20010412
- Cover, T. M., & Thomas, J. A. 2006, *Elements of Information Theory 2nd Edition* (Wiley Series in Telecommunications and Signal Processing) (Wiley-Interscience)
- Crevier, D. 1993, *AI: The Tumultuous History of the Search for Artificial Intelligence* (USA: Basic Books, Inc.)

- Cui, Y., Hada, K., Kawashima, T., et al. 2023, *Nature*, 621, 711, doi: [10.1038/s41586-023-06479-6](https://doi.org/10.1038/s41586-023-06479-6)
- Cybenko, G. 1989, *Mathematics of Control, Signals, and Systems (MCSS)*, 2, 303, doi: [10.1007/BF02551274](https://doi.org/10.1007/BF02551274)
- Dabrowski, Y., Fabian, A. C., Iwasawa, K., Lasenby, A. N., & Reynolds, C. S. 1997, *Monthly Notices of the Royal Astronomical Society*, 288, L11, doi: [10.1093/mnras/288.1.L11](https://doi.org/10.1093/mnras/288.1.L11)
- D’Abrusco, R., Longo, G., & Walton, N. A. 2009, *MNRAS*, 396, 223, doi: [10.1111/j.1365-2966.2009.14754.x](https://doi.org/10.1111/j.1365-2966.2009.14754.x)
- Daniel, M. K., Badran, H. M., Bond, I. H., et al. 2005, *ApJ*, 621, 181, doi: [10.1086/427406](https://doi.org/10.1086/427406)
- Davenport, T. 2006, *Harvard business review*, 84, 98
- de Menezes, R., Buson, S., Garrappa, S., et al. 2022, in *37th International Cosmic Ray Conference*, 955, doi: [10.22323/1.395.0955](https://doi.org/10.22323/1.395.0955)
- de Vaucouleurs, G. 1948, *Annales d’Astrophysique*, 11, 247
- Dempster, A. P., Laird, N. M., & Rubin, D. B. 1977, *Journal of the Royal Statistical Society: Series B (Methodological)*, 39, 1, doi: <https://doi.org/10.1111/j.2517-6161.1977.tb01600.x>
- Dermer, C. D., Cerruti, M., Lott, B., Boisson, C., & Zech, A. 2014, *ApJ*, 782, 82, doi: [10.1088/0004-637X/782/2/82](https://doi.org/10.1088/0004-637X/782/2/82)
- Dermer, C. D., & Giebels, B. 2016, *Comptes Rendus Physique*, 17, 594, doi: [10.1016/j.crhy.2016.04.004](https://doi.org/10.1016/j.crhy.2016.04.004)
- Dermer, C. D., & Schlickeiser, R. 1993a, *ApJ*, 416, 458, doi: [10.1086/173251](https://doi.org/10.1086/173251)
- . 1993b, *ApJ*, 416, 458, doi: [10.1086/173251](https://doi.org/10.1086/173251)
- . 1993c, *ApJ*, 416, 458, doi: [10.1086/173251](https://doi.org/10.1086/173251)
- Dermer, C. D., Sturmer, S. J., & Schlickeiser, R. 1997, *ApJS*, 109, 103, doi: [10.1086/312972](https://doi.org/10.1086/312972)
- Dieleman, S., Willett, K. W., & Dambre, J. 2015, *MNRAS*, 450, 1441, doi: [10.1093/mnras/stv632](https://doi.org/10.1093/mnras/stv632)
- Dimitrakoudis, S., Mastichiadis, A., Protheroe, R. J., & Reimer, A. 2012, *A&A*, 546, A120, doi: [10.1051/0004-6361/201219770](https://doi.org/10.1051/0004-6361/201219770)
- Domínguez, A., Wojtak, R., Finke, J., et al. 2019, *The Astrophysical Journal*, 885, 137, doi: [10.3847/1538-4357/ab4a0e](https://doi.org/10.3847/1538-4357/ab4a0e)
- Done, C., Davis, S. W., Jin, C., Blaes, O., & Ward, M. 2012, *MNRAS*, 420, 1848, doi: [10.1111/j.1365-2966.2011.19779.x](https://doi.org/10.1111/j.1365-2966.2011.19779.x)
- Dougherty, J., Kohavi, R., & Sahami, M. 1995, in *Machine Learning Proceedings 1995*, ed. A. Prieditis & S. Russell (San Francisco (CA): Morgan Kaufmann), 194–202, doi: <https://doi.org/10.1016/B978-1-55860-377-6.50032-3>

- Driver, S. P., Odewahn, S. C., Echevarria, L., et al. 2003, *The Astronomical Journal*, 126, 2662, doi: 10.1086/379560
- Drummond, C., & Holte, R. C. 2005, in *Machine Learning: ECML 2005*, ed. J. Gama, R. Camacho, P. B. Brazdil, A. M. Jorge, & L. Torgo (Berlin, Heidelberg: Springer Berlin Heidelberg), 539–546
- Duan, X. 2023, *Frontiers in Earth Science*, 10, 1046327, doi: 10.3389/feart.2022.1046327
- Duda, R., Hart, P., & Stork, D. 2012, *Pattern Classification* (Wiley). <https://books.google.nl/books?id=Br33IRC3PkQC>
- Dunlop, J. S., McLure, R. J., Kukula, M. J., et al. 2003, *MNRAS*, 340, 1095, doi: 10.1046/j.1365-8711.2003.06333.x
- Efron, B., & Hastie, T. 2016, *Computer age statistical inference: Algorithms, evidence, and data science*, 1–475, doi: 10.1017/CB09781316576533
- Elitzur, M., & Ho, L. C. 2009, *ApJ*, 701, L91, doi: 10.1088/0004-637X/701/2/L91
- Elkan, C. 2001, *Proceedings of the Seventeenth International Conference on Artificial Intelligence: 4-10 August 2001; Seattle*, 1
- Elvis, M., Risaliti, G., & Zamorani, G. 2002, *ApJ*, 565, L75, doi: 10.1086/339197
- Ennis, D. J., Neugebauer, G., & Werner, M. 1982, *ApJ*, 262, 460, doi: 10.1086/160441
- Errando Trias, M. 2009, PhD thesis, Barcelona, IFAE
- Fadely, R., Hogg, D. W., & Willman, B. 2012, *ApJ*, 760, 15, doi: 10.1088/0004-637X/760/1/15
- Falcke, H. 1995, *arXiv: Astrophysics*, 471, 19. <https://api.semanticscholar.org/CorpusID:15749419>
- Falcke, H., & Biermann, P. L. 1995, *A&A*, 293, 665, doi: 10.48550/arXiv.astro-ph/9411096
- Fan, J.-H. 2002, *PASJ*, 54, L55, doi: 10.1093/pasj/54.4.L55
- Fan, J.-H., Bastieri, D., Yang, J.-H., et al. 2014, *Research in Astronomy and Astrophysics*, 14, 1135, doi: 10.1088/1674-4527/14/9/004
- Fan, J.-H., Wang, Y.-J., Yang, J.-H., & Su, C.-Y. 2004, *Chinese J. Astron. Astrophys.*, 4, 533, doi: 10.1088/1009-9271/4/6/533
- Fan, J.-H., Yang, J.-H., Liu, Y., & Zhang, J.-Y. 2013, *Research in Astronomy and Astrophysics*, 13, 259, doi: 10.1088/1674-4527/13/3/002
- Fan, J.-H., Yang, J.-H., Pan, J., & Hua, T.-X. 2011, *Research in Astronomy and Astrophysics*, 11, 1413, doi: 10.1088/1674-4527/11/12/004
- Fan, J. H., Yang, J. H., Liu, Y., et al. 2016, *ApJS*, 226, 20, doi: 10.3847/0067-0049/226/2/20
- Fan, J. H., Kurtanidze, O., Liu, Y., et al. 2017, *ApJ*, 837, 45, doi: 10.3847/1538-4357/aa5def
- Fan, J. H., Kurtanidze, S. O., Liu, Y., et al. 2021a, *ApJS*, 253, 10, doi: 10.3847/1538-4365/abd32d

- . 2021b, *ApJS*, 253, 10, doi: [10.3847/1538-4365/abd32d](https://doi.org/10.3847/1538-4365/abd32d)
- Fan, J.-H., Chen, K.-Y., Xiao, H.-B., et al. 2022, *Universe*, 8, doi: [10.3390/universe8080436](https://doi.org/10.3390/universe8080436)
- Fan, J.-H., Chen, K.-Y., Xiao, H.-B., et al. 2022, *Universe*, 8, 436, doi: [10.3390/universe8080436](https://doi.org/10.3390/universe8080436)
- Fath, E. A. 1909, *Lick Observatory Bulletin*, 149, 71, doi: [10.5479/ADS/bib/1909LicOB.5.71F](https://doi.org/10.5479/ADS/bib/1909LicOB.5.71F)
- Fawcett, T. 2006, *Pattern Recognition Letters*, 27, 861, doi: <https://doi.org/10.1016/j.patrec.2005.10.010>
- Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. 1996, *AI Magazine*, 17, 37, doi: [10.1609/aimag.v17i3.1230](https://doi.org/10.1609/aimag.v17i3.1230)
- Fender, R., & Belloni, T. 2004, *ARA&A*, 42, 317, doi: [10.1146/annurev.astro.42.053102.134031](https://doi.org/10.1146/annurev.astro.42.053102.134031)
- Fermi-LAT Collaboration, Abdollahi, S., Ackermann, M., et al. 2018, *Science*, 362, 1031, doi: [10.1126/science.aat8123](https://doi.org/10.1126/science.aat8123)
- Finke, J. D. 2013, *ApJ*, 763, 134, doi: [10.1088/0004-637X/763/2/134](https://doi.org/10.1088/0004-637X/763/2/134)
- Firmani, C., & Avila-Reese, V. 2003, in *Revista Mexicana de Astronomia y Astrofisica Conference Series*, Vol. 17, *Revista Mexicana de Astronomia y Astrofisica Conference Series*, ed. V. Avila-Reese, C. Firmani, C. S. Frenk, & C. Allen, 107–120, doi: [10.48550/arXiv.astro-ph/0303543](https://doi.org/10.48550/arXiv.astro-ph/0303543)
- FISHER, R. A. 1936, *Annals of Eugenics*, 7, 179, doi: <https://doi.org/10.1111/j.1469-1809.1936.tb02137.x>
- Foffano, L., Prandini, E., Franceschini, A., & Paiano, S. 2019, *MNRAS*, 486, 1741, doi: [10.1093/mnras/stz812](https://doi.org/10.1093/mnras/stz812)
- Fossati, G., Maraschi, L., Celotti, A., Comastri, A., & Ghisellini, G. 1998, *MNRAS*, 299, 433, doi: [10.1046/j.1365-8711.1998.01828.x](https://doi.org/10.1046/j.1365-8711.1998.01828.x)
- Fraley, C., & Raftery, A. E. 2002, *Journal of the American Statistical Association*, 97, 611, doi: [10.1198/016214502760047131](https://doi.org/10.1198/016214502760047131)
- Francis, P. J., Hewett, P. C., Foltz, C. B., et al. 1991, *ApJ*, 373, 465, doi: [10.1086/170066](https://doi.org/10.1086/170066)
- Frank, J., King, A., & Raine, D. J. 2002, *Accretion Power in Astrophysics: Third Edition*
- Freeman, K. C. 1970, *ApJ*, 160, 811, doi: [10.1086/150474](https://doi.org/10.1086/150474)
- Freund, Y., & Schapire, R. E. 1997, *Journal of Computer and System Sciences*, 55, 119, doi: <https://doi.org/10.1006/jcss.1997.1504>
- Gaia Collaboration, Prusti, T., de Bruijne, J. H. J., et al. 2016, *A&A*, 595, A1, doi: [10.1051/0004-6361/201629272](https://doi.org/10.1051/0004-6361/201629272)
- Galván, A., Fraija, N., Aguilar-Ruiz, E., et al. 2022, in *37th International Cosmic Ray Conference*, 1009, doi: [10.22323/1.395.01009](https://doi.org/10.22323/1.395.01009)

- Gao, D., Zhang, Y.-X., & Zhao, Y.-H. 2008, *Monthly Notices of the Royal Astronomical Society*, 386, 1417, doi: 10.1111/j.1365-2966.2008.13070.x
- Gao, S., Fedynitch, A., Winter, W., & Pohl, M. 2019, *Nature Astronomy*, 3, 88, doi: 10.1038/s41550-018-0610-1
- Gao, S., Pohl, M., & Winter, W. 2017, *ApJ*, 843, 109, doi: 10.3847/1538-4357/aa7754
- Garrappa, S., Buson, S., Franckowiak, A., et al. 2022, in 37th International Cosmic Ray Conference, 956, doi: 10.22323/1.395.0956
- Garrappa, S., Buson, S., Sinapius, J., Kadler, M., & Fermi-LAT Collaboration. 2021, *GRB Coordinates Network*, 31194, 1
- Garrappa, S., Buson, S., Franckowiak, A., et al. 2019, *ApJ*, 880, 103, doi: 10.3847/1538-4357/ab2ada
- Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. 2004, *Bayesian Data Analysis*, 2nd edn. (Chapman and Hall/CRC)
- Georganopoulos, M., Aharonian, F. A., & Kirk, J. G. 2002, *A&A*, 388, L25, doi: 10.1051/0004-6361:20020567
- Georganopoulos, M., & Kazanas, D. 2003, *ApJ*, 594, L27, doi: 10.1086/378557
- Ghisellini, G. 2013, *Radiative Processes in High Energy Astrophysics*, Vol. 873, doi: 10.1007/978-3-319-00612-3
- Ghisellini, G., Celotti, A., Fossati, G., Maraschi, L., & Comastri, A. 1998, *MNRAS*, 301, 451, doi: 10.1046/j.1365-8711.1998.02032.x
- Ghisellini, G., Haardt, F., & Matt, G. 2004, *A&A*, 413, 535, doi: 10.1051/0004-6361:20031562
- Ghisellini, G., Maraschi, L., & Tavecchio, F. 2009, *MNRAS*, 396, L105, doi: 10.1111/j.1745-3933.2009.00673.x
- Ghisellini, G., Maraschi, L., & Treves, A. 1985, *A&A*, 146, 204
- Ghisellini, G., Padovani, P., Celotti, A., & Maraschi, L. 1993, *ApJ*, 407, 65, doi: 10.1086/172493
- Ghisellini, G., & Tavecchio, F. 2008, *MNRAS*, 387, 1669, doi: 10.1111/j.1365-2966.2008.13360.x
- . 2010, *MNRAS*, 409, L79, doi: 10.1111/j.1745-3933.2010.00952.x
- Ghisellini, G., Tavecchio, F., & Chiaberge, M. 2005, *A&A*, 432, 401, doi: 10.1051/0004-6361:20041404
- Ghisellini, G., Tavecchio, F., Foschini, L., & Ghirlanda, G. 2011, *MNRAS*, 414, 2674, doi: 10.1111/j.1365-2966.2011.18578.x
- Ghisellini, G., Tavecchio, F., Foschini, L., et al. 2010, *MNRAS*, 402, 497, doi: 10.1111/j.1365-2966.2009.15898.x

- Ghisellini, G., Tavecchio, F., Maraschi, L., Celotti, A., & Sbarrato, T. 2014, *Nature*, 515, 376, doi: 10.1038/nature13856
- Gilmore, R. C., Somerville, R. S., Primack, J. R., & Domínguez, A. 2012, *MNRAS*, 422, 3189, doi: 10.1111/j.1365-2966.2012.20841.x
- Giommi, P., Ansari, S. G., & Micol, A. 1995, *A&AS*, 109, 267
- Giommi, P., Capalbi, M., Fiocchi, M., et al. 2002, in *Blazar Astrophysics with BeppoSAX and Other Observatories*, ed. P. Giommi, E. Massaro, & G. Palumbo, 63. <https://arxiv.org/abs/astro-ph/0209596>
- Giommi, P., Padovani, P., Polenta, G., et al. 2012, *MNRAS*, 420, 2899, doi: 10.1111/j.1365-2966.2011.20044.x
- Glorot, X., & Bengio, Y. 2010, in *JMLR Proceedings*, Vol. 9, AISTATS, ed. Y. W. Teh & D. M. Titterton (JMLR.org), 249–256. <http://dblp.uni-trier.de/db/journals/jmlr/jmlrp9.html#GlorotB10>
- Godambe, S. V., Rannot, R. C., Chandra, P., et al. 2008, *Journal of Physics G Nuclear Physics*, 35, 065202, doi: 10.1088/0954-3899/35/6/065202
- Goodfellow, I., Bengio, Y., & Courville, A. 2016, *Deep Learning* (MIT Press)
- Goodfellow, I., Pouget-Abadie, J., Mirza, M., et al. 2014, in *Advances in Neural Information Processing Systems*, ed. Z. Ghahramani, M. Welling, C. Cortes, N. Lawrence, & K. Weinberger, Vol. 27 (Curran Associates, Inc.). https://proceedings.neurips.cc/paper_files/paper/2014/file/5ca3e9b122f61f8f06494c97b1afccf3-Paper.pdf
- Gopal-Krishna, Mangalam, A., & Wiita, P. J. 2008, *ApJ*, 680, L13, doi: 10.1086/589739
- Graham, A. W., & Driver, S. P. 2005, *Publ. Astron. Soc. Australia*, 22, 118, doi: 10.1071/AS05001
- Greenstein, J. L., & Schmidt, M. 1964, *ApJ*, 140, 1, doi: 10.1086/147889
- Groves, B. A., Heckman, T. M., & Kauffmann, G. 2006, *MNRAS*, 371, 1559, doi: 10.1111/j.1365-2966.2006.10812.x
- Gültekin, K., Richstone, D. O., Gebhardt, K., et al. 2009, *ApJ*, 698, 198, doi: 10.1088/0004-637X/698/1/198
- Gupta, A. C., Kalita, N., Gaur, H., & Duorah, K. 2016, *Monthly Notices of the Royal Astronomical Society*, 462, 1508
- Gupta, A. C., Agarwal, A., Bhagwan, J., et al. 2016, *MNRAS*, 458, 1127, doi: 10.1093/mnras/stw377
- Guyon, I., & Elisseff, A. 2003, *J. Mach. Learn. Res.*, 3, 1157. <http://portal.acm.org/citation.cfm?id=944919.944968>
- H. E. S. S. Collaboration, Abramowski, A., Acero, F., et al. 2010, *A&A*, 520, A83, doi: 10.1051/0004-6361/201014484

- Haack, C., Wiebusch, C., & IceCube Collaboration. 2017a, in International Cosmic Ray Conference, Vol. 301, 35th International Cosmic Ray Conference (ICRC2017), 1005, doi: [10.22323/1.301.01005](https://doi.org/10.22323/1.301.01005)
- Haack, C., Wiebusch, C., & IceCube Collaboration. 2017b, in International Cosmic Ray Conference, Vol. 301, 35th International Cosmic Ray Conference (ICRC2017), 1005, doi: [10.22323/1.301.01005](https://doi.org/10.22323/1.301.01005)
- Hachimi, M., Kaddoum, G., Gagnon, G., & Illy, P. 2020, in 2020 International Symposium on Networks, Computers and Communications (ISNCC), 1–5, doi: [10.1109/ISNCC49221.2020.9297290](https://doi.org/10.1109/ISNCC49221.2020.9297290)
- Hada, K. 2017, *Galaxies*, 5, doi: [10.3390/galaxies5010002](https://doi.org/10.3390/galaxies5010002)
- Halzen, F. 2016, *Nature Physics*, 13, doi: [10.1038/nphys3816](https://doi.org/10.1038/nphys3816)
- Halzen, F. 2017, *Nature Physics*, 13, 232, doi: [10.1038/nphys3816](https://doi.org/10.1038/nphys3816)
- Halzen, F., & Zas, E. 1997, *ApJ*, 488, 669, doi: [10.1086/304741](https://doi.org/10.1086/304741)
- Han, H., Wang, W., & Mao, B. 2005, in International Conference on Intelligent Computing. <https://api.semanticscholar.org/CorpusID:12126950>
- Han, J., Kamber, M., & Pei, J. 2012, *Data mining concepts and techniques*, third edition, Waltham, Mass.: Morgan Kaufmann Publishers. http://www.amazon.de/Data-Mining-Concepts-Techniques-Management/dp/0123814790/ref=tmm_hrd_title_0?ie=UTF8&qid=1366039033&sr=1-1
- Hardcastle, M. J., Birkinshaw, M., & Worrall, D. M. 2001, *MNRAS*, 326, 1499, doi: [10.1111/j.1365-2966.2001.04699.x](https://doi.org/10.1111/j.1365-2966.2001.04699.x)
- Hastie, T., Tibshirani, R., & Friedman, J. 2009, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, Springer series in statistics (Springer). <https://books.google.nl/books?id=eBSgoAEACAAJ>
- Havaei, M., Davy, A., Warde-Farley, D., et al. 2017, *Medical Image Analysis*, 35, 18, doi: <https://doi.org/10.1016/j.media.2016.05.004>
- He, H., Bai, Y., Garcia, E. A., & Li, S. 2008, in 2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence), 1322–1328, doi: [10.1109/IJCNN.2008.4633969](https://doi.org/10.1109/IJCNN.2008.4633969)
- He, H., & Garcia, E. A. 2009, *IEEE Transactions on Knowledge and Data Engineering*, 21, 1263, doi: [10.1109/TKDE.2008.239](https://doi.org/10.1109/TKDE.2008.239)
- He, H., & Shen, X. 2007, in *Proceedings of the 2007 International Conference on Artificial Intelligence, ICAI 2007, Volume I, June 25-28, 2007, Las Vegas, Nevada, USA*, ed. H. R. Arabnia, M. Q. Yang, & J. Y. Yang (CSREA Press), 358–364
- He, H.-N., Wang, T., Fan, Y.-Z., Liu, S.-M., & Wei, D.-M. 2013, *Phys. Rev. D*, 87, 063011, doi: [10.1103/PhysRevD.87.063011](https://doi.org/10.1103/PhysRevD.87.063011)

- He, K., Zhang, X., Ren, S., & Sun, J. 2015, IEEE International Conference on Computer Vision (ICCV 2015), 1502, doi: [10.1109/ICCV.2015.123](https://doi.org/10.1109/ICCV.2015.123)
- He, Z., Wang, T., Liu, G., et al. 2019, Nature Astronomy, 3, 265, doi: [10.1038/s41550-018-0669-8](https://doi.org/10.1038/s41550-018-0669-8)
- He, Z., Liu, G., Wang, T., et al. 2022, Science Advances, 8, eabk3291, doi: [10.1126/sciadv.abk3291](https://doi.org/10.1126/sciadv.abk3291)
- HESS Collaboration, Abramowski, A., Acero, F., et al. 2013, MNRAS, 434, 1889, doi: [10.1093/mnras/stt1081](https://doi.org/10.1093/mnras/stt1081)
- Hey, T., Tansley, S., Tolle, K., & Gray, J. 2009, The Fourth Paradigm: Data-Intensive Scientific Discovery (Microsoft Research). <https://www.microsoft.com/en-us/research/publication/fourth-paradigm-data-intensive-scientific-discovery/>
- Hickox, R. C., & Alexander, D. M. 2018, ARA&A, 56, 625, doi: [10.1146/annurev-astro-081817-051803](https://doi.org/10.1146/annurev-astro-081817-051803)
- Hinton, G., Osindero, S., & Teh, Y.-W. 2006, Neural computation, 18, 1527, doi: [10.1162/neco.2006.18.7.1527](https://doi.org/10.1162/neco.2006.18.7.1527)
- Hinton, G., & Sejnowski, T. 1999, Unsupervised Learning: Foundations of Neural Computation, Computational Neuroscience Series (MIT Press). <https://books.google.nl/books?id=yj04Y0lje4cC>
- Hirata, K., Kajita, T., Koshiya, M., et al. 1987, Phys. Rev. Lett., 58, 1490, doi: [10.1103/PhysRevLett.58.1490](https://doi.org/10.1103/PhysRevLett.58.1490)
- Hochreiter, S. 1991, Untersuchungen zu dynamischen neuronalen Netzen. Diploma thesis, Institut für Informatik, Lehrstuhl Prof. Brauer, Technische Universität München
- Hochreiter, S., & Schmidhuber, J. 1997, Neural Computation, 9, 1735, doi: [10.1162/neco.1997.9.8.1735](https://doi.org/10.1162/neco.1997.9.8.1735)
- Hooper, D., Linden, T., & Vieregge, A. 2019, J. Cosmology Astropart. Phys., 2019, 012, doi: [10.1088/1475-7516/2019/02/012](https://doi.org/10.1088/1475-7516/2019/02/012)
- Hsieh, K.-K., & Wang, L. 2018, ArXiv, abs/1803.08625. <https://api.semanticscholar.org/CorpusID:4342149>
- Huang, W., Pan, Y., & Wu, J. 2014, Procedia Computer Science, 30, 75, doi: <https://doi.org/10.1016/j.procs.2014.05.383>
- Hubble, E. P. 1926, ApJ, 64, 321, doi: [10.1086/143018](https://doi.org/10.1086/143018)
- Hubel, D. H., & Wiesel, T. N. 1959, The Journal of Physiology, 148, 574, doi: <https://doi.org/10.1113/jphysiol.1959.sp006308>
- Huertas-Company, M., Rouan, D., Tasca, L., Soucail, G., & Le Fèvre, O. 2008, A&A, 478, 971, doi: [10.1051/0004-6361:20078625](https://doi.org/10.1051/0004-6361:20078625)

- Huertas-Company, M., Tasca, L., Rouan, D., et al. 2009, *A&A*, 497, 743, doi: [10.1051/0004-6361/200811255](https://doi.org/10.1051/0004-6361/200811255)
- IceCube Collaboration. 2005, *Nuclear Physics B Proceedings Supplements*, 138, 179, doi: [10.1016/j.nuclphysbps.2004.11.042](https://doi.org/10.1016/j.nuclphysbps.2004.11.042)
- . 2013a, *Science*, 342, 1242856, doi: [10.1126/science.1242856](https://doi.org/10.1126/science.1242856)
- . 2013b, *Science*, 342, 1242856, doi: [10.1126/science.1242856](https://doi.org/10.1126/science.1242856)
- IceCube Collaboration, Aartsen, M. G., Ackermann, M., et al. 2018a, *Science*, 361, eaat1378, doi: [10.1126/science.aat1378](https://doi.org/10.1126/science.aat1378)
- . 2018b, *Science*, 361, 147, doi: [10.1126/science.aat2890](https://doi.org/10.1126/science.aat2890)
- . 2018c, *Science*, 361, eaat1378, doi: [10.1126/science.aat1378](https://doi.org/10.1126/science.aat1378)
- Ioffe, S., & Szegedy, C. 2015
- Ivezić, Ž., Becker, R. H., Blanton, M., et al. 2002, in *Astronomical Society of the Pacific Conference Series*, Vol. 284, IAU Colloq. 184: AGN Surveys, ed. R. F. Green, E. Y. Khachikian, & D. B. Sanders, 137, doi: [10.48550/arXiv.astro-ph/0111024](https://doi.org/10.48550/arXiv.astro-ph/0111024)
- James, G., Witten, D., Hastie, T., & Tibshirani, R. 2013, *An Introduction to Statistical Learning: with Applications in R* (Springer). <https://faculty.marshall.usc.edu/gareth-james/ISL/>
- Japkowicz, N., & Stephen, S. 2002, *Intell. Data Anal.*, 6, 429–449
- Jolliffe, I. 2014, *Principal Component Analysis* (John Wiley Sons, Ltd), doi: <https://doi.org/10.1002/9781118445112.stat06472>
- Jolliffe, I. T. 2002, *Principal Component Analysis*, Springer Series in Statistics (New York: Springer-Verlag), doi: [10.1007/b98835](https://doi.org/10.1007/b98835)
- Jordan, M. I., & Mitchell, T. M. 2015, *Science*, 349, 255, doi: [10.1126/science.aaa8415](https://doi.org/10.1126/science.aaa8415)
- Jorstad, S., & Marscher, A. 2016, *Galaxies*, 4, 47, doi: [10.3390/galaxies4040047](https://doi.org/10.3390/galaxies4040047)
- Jorstad, S. G., Marscher, A. P., Mattox, J. R., et al. 2001, *ApJ*, 556, 738, doi: [10.1086/321605](https://doi.org/10.1086/321605)
- Jorstad, S. G., Marscher, A. P., Lister, M. L., et al. 2005, *AJ*, 130, 1418, doi: [10.1086/444593](https://doi.org/10.1086/444593)
- Jorstad, S. G., Marscher, A. P., Smith, P. S., et al. 2013, *ApJ*, 773, 147, doi: [10.1088/0004-637X/773/2/147](https://doi.org/10.1088/0004-637X/773/2/147)
- Jorstad, S. G., Marscher, A. P., Morozova, D. A., et al. 2017, *ApJ*, 846, 98, doi: [10.3847/1538-4357/aa8407](https://doi.org/10.3847/1538-4357/aa8407)
- Joshi, M., & Böttcher, M. 2011, *ApJ*, 727, 21, doi: [10.1088/0004-637X/727/1/21](https://doi.org/10.1088/0004-637X/727/1/21)
- Kadler, M., Krauß, F., Mannheim, K., et al. 2016, *Nature Physics*, 12, 807, doi: [10.1038/nphys3715](https://doi.org/10.1038/nphys3715)
- Kang, S.-J., Li, E., Ou, W., et al. 2019, *ApJ*, 887, 134, doi: [10.3847/1538-4357/ab558b](https://doi.org/10.3847/1538-4357/ab558b)

- Kapanadze, B., Dorner, D., Vercellone, S., et al. 2018, *MNRAS*, 473, 2542, doi: [10.1093/mnras/stx2492](https://doi.org/10.1093/mnras/stx2492)
- Karimpanal, T. G., & Bouffanais, R. 2019, *Adapt. Behav.*, 27, doi: [10.1177/1059712318818568](https://doi.org/10.1177/1059712318818568)
- Kaspi, S., Maoz, D., Netzer, H., et al. 2005, *ApJ*, 629, 61, doi: [10.1086/431275](https://doi.org/10.1086/431275)
- Kaspi, S., Smith, P. S., Netzer, H., et al. 2000a, *ApJ*, 533, 631, doi: [10.1086/308704](https://doi.org/10.1086/308704)
- . 2000b, *ApJ*, 533, 631, doi: [10.1086/308704](https://doi.org/10.1086/308704)
- Kass, R. E., & Raftery, A. E. 1995, *Journal of the American Statistical Association*, 90, 773
- Kataoka, J., Takahashi, T., Makino, F., et al. 2000, *The Astrophysical Journal*, 528, 243, doi: [10.1086/308154](https://doi.org/10.1086/308154)
- Keivani, A., Murase, K., Petropoulou, M., et al. 2018, *ApJ*, 864, 84, doi: [10.3847/1538-4357/aad59a](https://doi.org/10.3847/1538-4357/aad59a)
- Kellermann, K. I. 2003, *Astronomical Society of the Pacific Conference Series*, 300, 185. <https://arxiv.org/abs/astro-ph/0304165>
- Kellermann, K. I., Condon, J. J., Kimball, A. E., Perley, R. A., & Ivezić, Ž. 2016, *ApJ*, 831, 168, doi: [10.3847/0004-637X/831/2/168](https://doi.org/10.3847/0004-637X/831/2/168)
- Kellermann, K. I., & Pauliny-Toth, I. I. K. 1969, *ApJ*, 155, L71, doi: [10.1086/180305](https://doi.org/10.1086/180305)
- Kellermann, K. I., Sramek, R., Schmidt, M., Shaffer, D. B., & Green, R. 1989, *AJ*, 98, 1195, doi: [10.1086/115207](https://doi.org/10.1086/115207)
- Kellermann, K. I., Lister, M. L., Homan, D. C., et al. 2004, *ApJ*, 609, 539, doi: [10.1086/421289](https://doi.org/10.1086/421289)
- Kellermann, K. I., Kovalev, Y. Y., Lister, M. L., et al. 2007, *Ap&SS*, 311, 231, doi: [10.1007/s10509-007-9622-5](https://doi.org/10.1007/s10509-007-9622-5)
- Kennamer, N., Kirkby, D., Ihler, A., & Sanchez-Lopez, F. J. 2018, in *Proceedings of Machine Learning Research*, Vol. 80, *Proceedings of the 35th International Conference on Machine Learning*, ed. J. Dy & A. Krause (PMLR), 2582–2590. <https://proceedings.mlr.press/v80/kennamer18a.html>
- Ketkar, N. 2017, *Deep Learning with Python: A Hands-on Introduction (Apress)*. <https://books.google.nl/books?id=19CyDgAAQBAJ>
- Kim, E. J., & Brunner, R. J. 2016, *Monthly Notices of the Royal Astronomical Society*, 464, 4463, doi: [10.1093/mnras/stw2672](https://doi.org/10.1093/mnras/stw2672)
- Kim, J., Im, M., Choi, C., & Hwang, S. 2019, *ApJ*, 884, 103, doi: [10.3847/1538-4357/ab40cd](https://doi.org/10.3847/1538-4357/ab40cd)
- Kirk, J. G., Rieger, F. M., & Mastichiadis, A. 1998, *A&A*, 333, 452, doi: [10.48550/arXiv.astro-ph/9801265](https://doi.org/10.48550/arXiv.astro-ph/9801265)
- Kishimoto, M., Anderson, M., ten Brummelaar, T., et al. 2022, *ApJ*, 940, 28, doi: [10.3847/1538-4357/ac91c4](https://doi.org/10.3847/1538-4357/ac91c4)

- Kneiske, T. M., Bretz, T., Mannheim, K., & Hartmann, D. H. 2004, *A&A*, 413, 807, doi: [10.1051/0004-6361:20031542](https://doi.org/10.1051/0004-6361:20031542)
- Kohavi, R., & John, G. H. 1997, *Artificial Intelligence*, 97, 273, doi: [https://doi.org/10.1016/S0004-3702\(97\)00043-X](https://doi.org/10.1016/S0004-3702(97)00043-X)
- Koller, D., & Sahami, M. 1996, in *Proceedings of the Thirteenth International Conference on International Conference on Machine Learning, ICML'96* (San Francisco, CA, USA: Morgan Kaufmann Publishers Inc.), 284–292
- Komatsu, E., Smith, K. M., Dunkley, J., et al. 2011, *ApJS*, 192, 18, doi: [10.1088/0067-0049/192/2/18](https://doi.org/10.1088/0067-0049/192/2/18)
- Kotsiantis, S., & Kanellopoulos, D. 2005, *GESTS International Transactions on Computer Science and Engineering*, 32, 47
- Kotsiantis, S. B., & Pintelas, P. 2004. <https://api.semanticscholar.org/CorpusID:18540362>
- Krawczynski, H., Hughes, S. B., Horan, D., et al. 2004, *ApJ*, 601, 151, doi: [10.1086/380393](https://doi.org/10.1086/380393)
- Krolik, J. H. 1998, *Active Galactic Nuclei: From the Central Black Hole to the Galactic Environment*
- Kubat, M., Holte, R. C., & Matwin, S. 1998, *Mach. Learn.*, 30, 195–215, doi: [10.1023/A:1007452223027](https://doi.org/10.1023/A:1007452223027)
- Laboratories, S. U. S. E., Widrow, B., Laboratory, S. U. S. E. L. S. S. E., et al. 1960, Adaptive "adaline" Neuron Using Chemical "memistors.". <https://books.google.nl/books?id=Yc4EAAAAIAAJ>
- Lacy, M., Laurent-Muehleisen, S. A., Ridgway, S. E., Becker, R. H., & White, R. L. 2001, *ApJ*, 551, L17, doi: [10.1086/319836](https://doi.org/10.1086/319836)
- LaMassa, S. M., Cales, S., Moran, E. C., et al. 2015, *ApJ*, 800, 144, doi: [10.1088/0004-637X/800/2/144](https://doi.org/10.1088/0004-637X/800/2/144)
- Langley, P. 2011, *Machine Learning*, 82, 275, doi: [10.1007/s10994-011-5242-y](https://doi.org/10.1007/s10994-011-5242-y)
- Laor, A., & Behar, E. 2008, *MNRAS*, 390, 847, doi: [10.1111/j.1365-2966.2008.13806.x](https://doi.org/10.1111/j.1365-2966.2008.13806.x)
- LeCun, Y., Bengio, Y., & Hinton, G. 2015, *Nature*, 521, 436, doi: [10.1038/nature14539](https://doi.org/10.1038/nature14539)
- LeCun, Y., Bengio, Y., & Hinton, G. 2015, *Nature*, 521, 436, doi: [10.1038/nature14539](https://doi.org/10.1038/nature14539)
- Li, H., & Kusunose, M. 2000, *ApJ*, 536, 729, doi: [10.1086/308960](https://doi.org/10.1086/308960)
- Li, R.-L., Zhu, B.-Y., & Liang, Y.-F. 2022a, *Phys. Rev. D*, 106, 083024, doi: [10.1103/PhysRevD.106.083024](https://doi.org/10.1103/PhysRevD.106.083024)
- . 2022b, *Phys. Rev. D*, 106, 083024, doi: [10.1103/PhysRevD.106.083024](https://doi.org/10.1103/PhysRevD.106.083024)
- Liao, N.-H., Sheng, Z.-F., Jiang, N., et al. 2022, *ApJ*, 932, L25, doi: [10.3847/2041-8213/ac756f](https://doi.org/10.3847/2041-8213/ac756f)

- Lin, C., & Fan, J.-H. 2016, *Research in Astronomy and Astrophysics*, 16, 103, doi: [10.1088/1674-4527/16/7/103](https://doi.org/10.1088/1674-4527/16/7/103)
- Liodakis, I., Hovatta, T., Huppenkothen, D., et al. 2018, *ApJ*, 866, 137, doi: [10.3847/1538-4357/aae2b7](https://doi.org/10.3847/1538-4357/aae2b7)
- Lister, M. L., Aller, M. F., Aller, H. D., et al. 2018, *ApJS*, 234, 12, doi: [10.3847/1538-4365/aa9c44](https://doi.org/10.3847/1538-4365/aa9c44)
- . 2015, *ApJ*, 810, L9, doi: [10.1088/2041-8205/810/1/L9](https://doi.org/10.1088/2041-8205/810/1/L9)
- Lister, M. L., & Marscher, A. P. 1999, *Astroparticle Physics*, 11, 65, doi: [10.1016/S0927-6505\(99\)00027-4](https://doi.org/10.1016/S0927-6505(99)00027-4)
- Lister, M. L., Cohen, M. H., Homan, D. C., et al. 2009, *AJ*, 138, 1874, doi: [10.1088/0004-6256/138/6/1874](https://doi.org/10.1088/0004-6256/138/6/1874)
- Lister, M. L., Homan, D. C., Hovatta, T., et al. 2019, *ApJ*, 874, 43, doi: [10.3847/1538-4357/ab08ee](https://doi.org/10.3847/1538-4357/ab08ee)
- Liu, R.-Y., Wang, K., Xue, R., et al. 2019, *Phys. Rev. D*, 99, 063008, doi: [10.1103/PhysRevD.99.063008](https://doi.org/10.1103/PhysRevD.99.063008)
- Liu, R.-Y., Wang, X.-Y., Inoue, S., Crocker, R., & Aharonian, F. 2014, *Phys. Rev. D*, 89, 083004, doi: [10.1103/PhysRevD.89.083004](https://doi.org/10.1103/PhysRevD.89.083004)
- Lott, B., Gasparri, D., & Ciprini, S. 2020, The Fourth Catalog of Active Galactic Nuclei Detected by the Fermi Large Area Telescope – Data Release 2. <https://arxiv.org/abs/2010.08406>
- Lu, Z., Pu, H., Wang, F., Hu, Z., & Wang, L. 2017, in *Neural Information Processing Systems*. <https://api.semanticscholar.org/CorpusID:3235741>
- Luenberger, D. G., & Ye, Y. 2015, *Linear and Nonlinear Programming* (Springer Publishing Company, Incorporated)
- Lunardini, C., & Winter, W. 2017, *Phys. Rev. D*, 95, 123001, doi: [10.1103/PhysRevD.95.123001](https://doi.org/10.1103/PhysRevD.95.123001)
- Lynden-Bell, D. 1969, *Nature*, 223, 690, doi: [10.1038/223690a0](https://doi.org/10.1038/223690a0)
- . 2003, *MNRAS*, 341, 1360, doi: [10.1046/j.1365-8711.2003.06506.x](https://doi.org/10.1046/j.1365-8711.2003.06506.x)
- MacDonald, G. H., Kenderdine, S., & Neville, A. C. 1968, *MNRAS*, 138, 259, doi: [10.1093/mnras/138.3.259](https://doi.org/10.1093/mnras/138.3.259)
- MacKay, D. 2003, *Information Theory, Inference and Learning Algorithms* (Cambridge University Press). <https://books.google.nl/books?id=AKuMj4PN EMC>
- MAGIC Collaboration. 2000, in *American Institute of Physics Conference Series*, Vol. 515, *GeV-TeV Gamma Ray Astrophysics Workshop : towards a major atmospheric*, ed. B. L. Dingus, M. H. Salamon, & D. B. Kieda, 510–514, doi: [10.1063/1.1291417](https://doi.org/10.1063/1.1291417)
- Majumder, A., Mitra, K., Chatterjee, R., et al. 2019, *MNRAS*, 490, 124, doi: [10.1093/mnras/stz2557](https://doi.org/10.1093/mnras/stz2557)

- Mannheim, K., Stanev, T., & Biermann, P. L. 1992, *A&A*, 260, L1
- Manyika, J. 2011. <https://api.semanticscholar.org/CorpusID:166449414>
- Maraschi, L., Ghisellini, G., & Celotti, A. 1992a, *ApJ*, 397, L5, doi: 10.1086/186531
- . 1992b, *ApJ*, 397, L5, doi: 10.1086/186531
- Marchesini, D., Celotti, A., & Ferrarese, L. 2004, *MNRAS*, 351, 733, doi: 10.1111/j.1365-2966.2004.07822.x
- Marin, F. 2016, *MNRAS*, 460, 3679, doi: 10.1093/mnras/stw1131
- Marscher, A. P. 2014, *ApJ*, 780, 87, doi: 10.1088/0004-637X/780/1/87
- Marscher, A. P., & Gear, W. K. 1985, *ApJ*, 298, 114, doi: 10.1086/163592
- Marscher, A. P., Jorstad, S. G., Larionov, V. M., et al. 2010, *ApJ*, 710, L126, doi: 10.1088/2041-8205/710/2/L126
- Marshall, H. L., Harris, D. E., Grimes, J. P., et al. 2001, *ApJ*, 549, L167, doi: 10.1086/319161
- Massaro, F., Paggi, A., Errando, M., et al. 2013, *ApJS*, 207, 16, doi: 10.1088/0067-0049/207/1/16
- Massey, F. J. 1951, *Journal of the American Statistical Association*, 46, 68. <http://www.jstor.org/stable/2280095>
- Masters, K. L., Lintott, C. J., Hart, R. E., et al. 2019, *Monthly Notices of the Royal Astronomical Society*, 487, 1808, doi: 10.1093/mnras/stz1153
- Mastichiadis, A. 2016, *Galaxies*, 4, doi: 10.3390/galaxies4040059
- Mastichiadis, A., & Kirk, J. G. 1997, *A&A*, 320, 19, doi: 10.48550/arXiv.astro-ph/9610058
- Matthews, T. A., Morgan, W. W., & Schmidt, M. 1964, *ApJ*, 140, 35, doi: 10.1086/147890
- Matthews, T. A., & Sandage, A. R. 1963, *ApJ*, 138, 30, doi: 10.1086/147615
- Mazurowski, M. A., Habas, P. A., Zurada, J. M., et al. 2008, *Neural Netw.*, 21, 427–436
- McCarthy, J. 2004
- McCullagh, P., & Nelder, J. A. 1989, *Generalized Linear Models* (London: Chapman Hall / CRC)
- Mcculloch, W., & Pitts, W. 1943, *Bulletin of Mathematical Biophysics*, 5, 127
- McLure, R. J., & Dunlop, J. S. 2001, *MNRAS*, 327, 199, doi: 10.1046/j.1365-8711.2001.04709.x
- McLure, R. J., & Jarvis, M. J. 2004, *MNRAS*, 353, L45, doi: 10.1111/j.1365-2966.2004.08305.x
- Meegan, C., Lichti, G., Bhat, P. N., et al. 2009, *ApJ*, 702, 791, doi: 10.1088/0004-637X/702/1/791
- Meyer, E. T., Fossati, G., Georganopoulos, M., & Lister, M. L. 2011, *ApJ*, 740, 98, doi: 10.1088/0004-637X/740/2/98

- . 2012, *ApJ*, 752, L4, doi: 10.1088/2041-8205/752/1/L4
- Michelson, P. F., Atwood, W. B., & Ritz, S. 2010, *Reports on Progress in Physics*, 73, 074901, doi: 10.1088/0034-4885/73/7/074901
- Miebach, M. P. 2002, in *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, Vol. 4844, *Observatory Operations to Optimize Scientific Return III*, ed. P. J. Quinn, 408–416, doi: 10.1117/12.460637
- Mikolov, T., Karafiát, M., Burget, L., Cernocký, J., & Khudanpur, S. 2010, in , 1045–1048
- Miller, L., Peacock, J. A., & Mead, A. R. G. 1990, *MNRAS*, 244, 207
- Minsky, M., & Papert, S. 1969, *Perceptrons* (Cambridge, MA: MIT Press)
- Mitchell, T. 1997, *Machine Learning*, McGraw-Hill International Editions (McGraw-Hill). <https://books.google.nl/books?id=EoYBngEACAAJ>
- Miyoshi, M., Moran, J., Herrnstein, J., et al. 1995, *Nature*, 373, 127, doi: 10.1038/373127a0
- Mohri, M., Rostamizadeh, A., & Talwalkar, A. 2018, *Foundations of Machine Learning*, 2nd edn., *Adaptive Computation and Machine Learning* (Cambridge, MA: MIT Press)
- Moiseev, A. A., Hartman, R. C., Ormes, J. F., et al. 2007, *Astroparticle Physics*, 27, 339, doi: 10.1016/j.astropartphys.2006.12.003
- Morgan, W. W. 1958, *PASP*, 70, 364, doi: 10.1086/127243
- . 1959, *PASP*, 71, 394, doi: 10.1086/127415
- Moskalenko, I. V., Strong, A. W., Digel, S. W., & Porter, T. A. 2007, in *American Institute of Physics Conference Series*, Vol. 921, *The First GLAST Symposium*, ed. S. Ritz, P. Michelson, & C. A. Meegan, 490–491, doi: 10.1063/1.2757416
- Mukherjee, S., Niyogi, P., Poggio, T., & Rifkin, R. 2006, *Adv. Comput. Math.*, 25, 161, doi: 10.1007/s10444-004-7634-z
- Müller-Sánchez, F., Prieto, M. A., Hicks, E. K. S., et al. 2011, *ApJ*, 739, 69, doi: 10.1088/0004-637X/739/2/69
- Murase, K., Inoue, Y., & Dermer, C. D. 2014, *Phys. Rev. D*, 90, 023007, doi: 10.1103/PhysRevD.90.023007
- Murase, K., Kimura, S. S., Zhang, B. T., Oikonomou, F., & Petropoulou, M. 2020, *ApJ*, 902, 108, doi: 10.3847/1538-4357/abb3c0
- Murase, K., Oikonomou, F., & Petropoulou, M. 2018, *ApJ*, 865, 124, doi: 10.3847/1538-4357/aada00
- Murray, N., Chiang, J., Grossman, S. A., & Voit, G. M. 1995, *ApJ*, 451, 498, doi: 10.1086/176238
- Nair, V., & Hinton, G. E. 2010, in *Proceedings of the 27th International Conference on Machine Learning (ICML-10)*, ed. J. Fürnkranz & T. Joachims, 807–814

- Narayan, R., McClintock, J. E., & Tchekhovskoy, A. 2014, in *General Relativity, Cosmology and Astrophysics*, ed. J. Bičák & T. Ledvinka, Vol. 177, 523, doi: 10.1007/978-3-319-06349-2_25
- Narayan, R., & Yi, I. 1994, *ApJ*, 428, L13, doi: 10.1086/187381
- . 1995, *ApJ*, 452, 710, doi: 10.1086/176343
- Nesterov, I. 2004, *Introductory Lectures on Convex Optimization: A Basic Course*, Mathematics and its applications (Kluwer Academic Publishers). <https://books.google.nl/books?id=klv2PwAACAAJ>
- Netzer, H. 2013, *The Physics and Evolution of Active Galactic Nuclei*
- . 2015, *ARA&A*, 53, 365, doi: 10.1146/annurev-astro-082214-122302
- Newell, A., & Simon, H. A. 1976, *Communications of the ACM*, 19, 113, doi: 10.1145/1283920.1283930
- Nieppola, E., Valtaoja, E., Tornikoski, M., Hovatta, T., & Kotiranta, M. 2008, *A&A*, 488, 867, doi: 10.1051/0004-6361:200809716
- Nievas Rosillo, M., Domínguez, A., Chiaro, G., et al. 2022, *MNRAS*, 512, 137, doi: 10.1093/mnras/stac491
- Nocedal, J., & Wright, S. 2006, *Numerical Optimization*, Springer Series in Operations Research and Financial Engineering (Springer New York). <https://books.google.nl/books?id=VbHYoSyeIFcC>
- Nokhrina, E. E., Pashchenko, I. N., & Kutkin, A. M. 2022, *MNRAS*, 509, 1899, doi: 10.1093/mnras/stab3094
- Nolan, P. L., Abdo, A. A., Ackermann, M., et al. 2012a, *ApJS*, 199, 31, doi: 10.1088/0067-0049/199/2/31
- . 2012b, *ApJS*, 199, 31, doi: 10.1088/0067-0049/199/2/31
- Odewahn, S. C., Cohen, S. H., Windhorst, R. A., & Philip, N. S. 2002, *ApJ*, 568, 539, doi: 10.1086/339036
- Odewahn, S. C., Humphreys, R. M., Aldering, G., & Thurmes, P. 1993, *PASP*, 105, 1354, doi: 10.1086/133317
- Odewahn, S. C., Stockwell, E. B., Pennington, R. L., Humphreys, R. M., & Zumach, W. A. 1992, *AJ*, 103, 318, doi: 10.1086/116063
- Odewahn, S. C., Windhorst, R. A., Driver, S. P., & Keel, W. C. 1996, *ApJ*, 472, L13, doi: 10.1086/310345
- Odewahn, S. C., de Carvalho, R. R., Gal, R. R., et al. 2004, *The Astronomical Journal*, 128, 3092, doi: 10.1086/425525

- Ormes, J., & for High Energy Astrophysics (Goddard Space Flight Center), L. 1996, *Current Perspectives in High Energy Astrophysics*, NASA reference publication (National Aeronautics and Space Administration, Goddard Space Flight Center). <https://books.google.de/books?id=3pI6AQAAIAAJ>
- Ortigosa-Hernández, J., Inza, I., & Lozano, J. A. 2017, *Pattern Recognition Letters*, 98, 32, doi: 10.1016/j.patrec.2017.08.002
- Osorio, M., Sacahui, R., González, M. M., Fraija, N., & García-González, J. A. 2019, in *International Cosmic Ray Conference*, Vol. 36, 36th International Cosmic Ray Conference (ICRC2019), 686
- Osterbrock, D. E. 1981, *ApJ*, 249, 462, doi: 10.1086/159306
- Ostrowski, M., & Bednarz, J. 2002, *A&A*, 394, 1141, doi: 10.1051/0004-6361:20021173
- Padovani, P. 2007, *Ap&SS*, 309, 63, doi: 10.1007/s10509-007-9455-2
- Padovani, P. 2016, in *Active Galactic Nuclei 12: A Multi-Messenger Perspective (AGN12)*, 14, doi: 10.5281/zenodo.163592
- Padovani, P., Boccardi, B., Falomo, R., & Giommi, P. 2022, *MNRAS*, 511, 4697, doi: 10.1093/mnras/stac376
- Padovani, P., Costamante, L., Ghisellini, G., Giommi, P., & Perlmán, E. 2002, *ApJ*, 581, 895, doi: 10.1086/344406
- Padovani, P., Costamante, L., Giommi, P., et al. 2004, *MNRAS*, 347, 1282, doi: 10.1111/j.1365-2966.2004.07313.x
- Padovani, P., & Giommi, P. 1995, *ApJ*, 444, 567, doi: 10.1086/175631
- Padovani, P., Giommi, P., & Rau, A. 2012, *MNRAS*, 422, L48, doi: 10.1111/j.1745-3933.2012.01234.x
- Padovani, P., Oikonomou, F., Petropoulou, M., Giommi, P., & Resconi, E. 2019, *MNRAS*, 484, L104, doi: 10.1093/mnrasl/slz011
- Padovani, P., Perlmán, E. S., Landt, H., Giommi, P., & Perri, M. 2003, *ApJ*, 588, 128, doi: 10.1086/373899
- Padovani, P., Petropoulou, M., Giommi, P., & Resconi, E. 2015, *Monthly Notices of the Royal Astronomical Society*, 452, 1877, doi: 10.1093/mnras/stv1467
- Padovani, P., & Resconi, E. 2014, *MNRAS*, 443, 474, doi: 10.1093/mnras/stu1166
- Padovani, P., Resconi, E., Giommi, P., Arsioli, B., & Chang, Y. L. 2016, *MNRAS*, 457, 3582, doi: 10.1093/mnras/stw228
- Padovani, P., Resconi, E., Giommi, P., Arsioli, B., & Chang, Y. L. 2016a, *Monthly Notices of the Royal Astronomical Society*, 457, 3582, doi: 10.1093/mnras/stw228
- Padovani, P., Alexander, D. M., Assef, R. J., et al. 2016b, *The Astronomy and Astrophysics Review*, 25, 1. <https://api.semanticscholar.org/CorpusID:119528797>

- Paiano, S., Landoni, M., Falomo, R., et al. 2017, *ApJ*, 837, 144, doi: 10.3847/1538-4357/837/2/144
- Paliya, V. S., Domínguez, A., Ajello, M., Olmo-García, A., & Hartmann, D. 2021, *ApJS*, 253, 46, doi: 10.3847/1538-4365/abe135
- Paliya, V. S., Sahayanathan, S., & Stalin, C. S. 2015, *ApJ*, 803, 15, doi: 10.1088/0004-637X/803/1/15
- Palladino, A., Rodrigues, X., Gao, S., & Winter, W. 2019a, *ApJ*, 871, 41, doi: 10.3847/1538-4357/aaf507
- . 2019b, *ApJ*, 871, 41, doi: 10.3847/1538-4357/aaf507
- Paltani, S., Courvoisier, T. J. L., & Walter, R. 1998, *A&A*, 340, 47, doi: 10.48550/arXiv.astro-ph/9809113
- Pandey, A., & Stalin, C. S. 2022, *A&A*, 668, A152, doi: 10.1051/0004-6361/202244648
- Panessa, F., Baldi, R. D., Laor, A., et al. 2019, *Nature Astronomy*, 3, 387, doi: 10.1038/s41550-019-0765-4
- Pâris, I., Petitjean, P., Aubourg, É., et al. 2018, *A&A*, 613, A51, doi: 10.1051/0004-6361/201732445
- Pascanu, R., Mikolov, T., & Bengio, Y. 2012a, *CoRR*, abs/1211.5063. <http://dblp.uni-trier.de/db/journals/corr/corr1211.html#abs-1211-5063>
- . 2012b, 30th International Conference on Machine Learning, ICML 2013
- Patil, B. 2023, *International Journal of Advanced Research in Science, Communication and Technology*, 476, doi: 10.48175/IJAR SCT-12167
- Pearson, K. 1901, *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 2, 559, doi: 10.1080/14786440109462720
- Pedregosa, F., Varoquaux, G., Gramfort, A., et al. 2011a, *Journal of Machine Learning Research*, 12, 2825
- . 2011b, *Journal of Machine Learning Research*, 12, 2825
- Pei, Z., Fan, J., Bastieri, D., Yang, J., & Xiao, H. 2020, *Science China Physics, Mechanics, and Astronomy*, 63, 259511, doi: 10.1007/s11433-019-1454-6
- Pei, Z.-Y., Fan, J.-H., Bastieri, D., Sawangwit, U., & Yang, J.-H. 2019, *Research in Astronomy and Astrophysics*, 19, 070, doi: 10.1088/1674-4527/19/5/70
- Pei, Z.-Y., Fan, J.-H., Liu, Y., et al. 2016, *Ap&SS*, 361, 237, doi: 10.1007/s10509-016-2822-0
- Peng, J., Jury, E., Dönnès, P., & Ciurtin, C. 2021, *Frontiers in Pharmacology*, 12, doi: 10.3389/fphar.2021.720694
- Perrone, L. 2023, in *European Physical Journal Web of Conferences*, Vol. 283, *European Physical Journal Web of Conferences*, 04004, doi: 10.1051/epjconf/202328304004

- Peterson, B. M. 1993, *PASP*, 105, 247, doi: 10.1086/133140
- . 1997, *An Introduction to Active Galactic Nuclei*
- Peterson, B. M., Barth, A. J., Berlind, P., et al. 1999, *ApJ*, 510, 659, doi: 10.1086/306604
- Peterson, B. M., McHardy, I. M., Wilkes, B. J., et al. 2000, *ApJ*, 542, 161, doi: 10.1086/309518
- Petropoulou, M., Dimitrakoudis, S., Padovani, P., Mastichiadis, A., & Resconi, E. 2015, *MNRAS*, 448, 2412, doi: 10.1093/mnras/stv179
- Petropoulou, M., & Mastichiadis, A. 2014, *Monthly Notices of the Royal Astronomical Society*, 447, 36, doi: 10.1093/mnras/stu2364
- Philip, N. S., Wadadekar, Y., Kembhavi, A., & Joseph, K. B. 2002, *A&A*, 385, 1119, doi: 10.1051/0004-6361:20020219
- Piner, B. G., & Edwards, P. G. 2014, *ApJ*, 797, 25, doi: 10.1088/0004-637X/797/1/25
- . 2018, *ApJ*, 853, 68, doi: 10.3847/1538-4357/aaa425
- Piner, B. G., Mahmud, M., Fey, A. L., & Gospodinova, K. 2007, *AJ*, 133, 2357, doi: 10.1086/514812
- Piner, B. G., Pant, N., & Edwards, P. G. 2010, *ApJ*, 723, 1150, doi: 10.1088/0004-637X/723/2/1150
- Piner, B. G., Pushkarev, A. B., Kovalev, Y. Y., et al. 2012, *ApJ*, 758, 84, doi: 10.1088/0004-637X/758/2/84
- Pogge, R. W., & Peterson, B. M. 1992, *AJ*, 103, 1084, doi: 10.1086/116127
- Poole, D., Mackworth, A., & Goebel, R. 1998, *Computational Intelligence: A Logical Approach*
- Primack, J. R., Bullock, J. S., Somerville, R. S., & MacMinn, D. 1999, *Astroparticle Physics*, 11, 93, doi: 10.1016/S0927-6505(99)00031-6
- Pudil, P., Novovičová, J., & Kittler, J. 1994, *Pattern Recognition Letters*, 15, 1119, doi: 10.1016/0167-8655(94)90127-9
- Punch, M., Akerlof, C. W., Cawley, M. F., et al. 1992, *Nature*, 358, 477, doi: 10.1038/358477a0
- Qin, D.-M., Guo, P., Hu, Z.-Y., & Zhao, Y.-H. 2003, *Chinese J. Astron. Astrophys.*, 3, 277, doi: 10.1088/1009-9271/3/3/277
- Quinlan, J. R. 1986, *Machine Learning*, 1, 81
- Rachen, J. P. 2000, in *American Institute of Physics Conference Series*, Vol. 515, *GeV-TeV Gamma Ray Astrophysics Workshop : towards a major atmospheric*, ed. B. L. Dingus, M. H. Salamon, & D. B. Kieda, 41–52, doi: 10.1063/1.1291342
- Radcliffe, J. F., Barthel, P. D., Garrett, M. A., et al. 2021, *A&A*, 649, L9, doi: 10.1051/0004-6361/202140791

- Raschka, S. 2015a, Python Machine Learning (Packt Publishing - ebooks Account). <http://www.amazon.com/exec/obidos/redirect?tag=citeulike07-20&path=ASIN/1783555130>
- . 2015b, Python Machine Learning (Packt Publishing)
- Rashidi, H., Tran, N., Betts, E., Howell, L., & Green, R. 2019, *Academic Pathology*, 6, 237428951987308, doi: 10.1177/2374289519873088
- Rawlings, S., & Saunders, R. 1991, *Nature*, 349, 138, doi: 10.1038/349138a0
- Readhead, A. C. S. 1994, *ApJ*, 426, 51, doi: 10.1086/174038
- Rees, M. J. 1966, *Nature*, 211, 468, doi: 10.1038/211468a0
- . 1984, *ARA&A*, 22, 471, doi: 10.1146/annurev.aa.22.090184.002351
- Reichard, T. A., Richards, G. T., Hall, P. B., et al. 2003, *AJ*, 126, 2594, doi: 10.1086/379293
- Richards, G. T., Myers, A. D., Gray, A. G., et al. 2009, *ApJS*, 180, 67, doi: 10.1088/0067-0049/180/1/67
- Rieger, F. M., & Duffy, P. 2004, *ApJ*, 617, 155, doi: 10.1086/425167
- Righi, C., Tavecchio, F., & Guetta, D. 2017, *A&A*, 598, A36, doi: 10.1051/0004-6361/201629412
- Roberts, M. S., & Haynes, M. P. 1994, *ARA&A*, 32, 115, doi: 10.1146/annurev.aa.32.090194.000555
- Robson, E. I., Gear, W. K., Smith, M. G., Ade, P. A. R., & Nolt, I. G. 1985, *MNRAS*, 213, 355, doi: 10.1093/mnras/213.2.355
- Rodrigues, X., Fedynitch, A., Gao, S., Boncioli, D., & Winter, W. 2018, *ApJ*, 854, 54, doi: 10.3847/1538-4357/aaa7ee
- Rodriguez, J.-V., Rodríguez-Rodríguez, I., & Woo, W. L. 2022, *WIREs Data Mining and Knowledge Discovery*, 12, e1476, doi: 10.1002/widm.1476
- Ruder, S. 2016, An overview of gradient descent optimization algorithms. <http://arxiv.org/abs/1609.04747>
- Rumelhart, D. E., Hinton, G. E., & Williams, R. J. 1986, *Nature*, 323, 533. <https://api.semanticscholar.org/CorpusID:205001834>
- Russell, S., & Norvig, P. 2010, *Artificial Intelligence: A Modern Approach*, 3rd edn. (Prentice Hall)
- . 2020, *Artificial Intelligence: A Modern Approach (4th Edition)* (Pearson). <http://aima.cs.berkeley.edu/>
- Russell, S. J., & Norvig, P. 2002, *Artificial Intelligence: A Modern Approach (2nd Edition)* (Prentice Hall). <http://www.amazon.ca/exec/obidos/redirect?tag=citeulike09-20&path=ASIN/0137903952>
- Rybicki, G. B., & Lightman, A. P. 1986, *Radiative Processes in Astrophysics*

- Saeys, Y., Inza, I., & Larranaga, P. 2007, *Bioinformatics* (Oxford, England), 23, 2507, doi: [10.1093/bioinformatics/btm344](https://doi.org/10.1093/bioinformatics/btm344)
- Sahakyan, N. 2018, *ApJ*, 866, 109, doi: [10.3847/1538-4357/aadade](https://doi.org/10.3847/1538-4357/aadade)
- Sahu, S., López Fortín, C. E., & Nagataki, S. 2019, *ApJ*, 884, L17, doi: [10.3847/2041-8213/ab43c7](https://doi.org/10.3847/2041-8213/ab43c7)
- Saldana-Lopez, A., Domínguez, A., Pérez-González, P. G., et al. 2021, *MNRAS*, 507, 5144, doi: [10.1093/mnras/stab2393](https://doi.org/10.1093/mnras/stab2393)
- Salpeter, E. E. 1964, *ApJ*, 140, 796, doi: [10.1086/147973](https://doi.org/10.1086/147973)
- Sambruna, R. M., Maraschi, L., & Urry, C. M. 1996, *ApJ*, 463, 444, doi: [10.1086/177260](https://doi.org/10.1086/177260)
- Sambruna, R. M., Aharonian, F. A., Krawczynski, H., et al. 2000, *ApJ*, 538, 127, doi: [10.1086/309133](https://doi.org/10.1086/309133)
- Sammut, C., & Webb, G. I., eds. 2017, *Encyclopedia of Machine Learning and Data Mining*, 2nd edn., Springer Reference (New York: Springer), doi: [10.1007/978-1-4899-7687-1](https://doi.org/10.1007/978-1-4899-7687-1)
- Samuel, A. L. 1959, *IBM Journal of Research and Development*, 3, 210, doi: [10.1147/rd.33.0210](https://doi.org/10.1147/rd.33.0210)
- Sanchez Saez, P. 2022, in *SciOps 2022: Artificial Intelligence for Science and Operations in Astronomy (SCIOPS)*. Proceedings of the ESA/ESO SCOPS Workshop held 16-20 May, 17, doi: [10.5281/zenodo.6563034](https://doi.org/10.5281/zenodo.6563034)
- Sandage, A. 1961, *The Hubble Atlas of Galaxies*
- Sanders, D. B., Phinney, E. S., Neugebauer, G., Soifer, B. T., & Matthews, K. 1989, *ApJ*, 347, 29, doi: [10.1086/168094](https://doi.org/10.1086/168094)
- Santosh, K., Das, N., & Ghosh, S. 2022, *Deep learning models*, 65–97, doi: [10.1016/B978-0-12-823504-1.00013-1](https://doi.org/10.1016/B978-0-12-823504-1.00013-1)
- Sarangi, S., Sahidullah, M., & Saha, G. 2020, *Digital Signal Processing*, 104, 102795, doi: <https://doi.org/10.1016/j.dsp.2020.102795>
- Scheuer, P. A. G., & Readhead, A. C. S. 1979, *Nature*, 277, 182, doi: [10.1038/277182a0](https://doi.org/10.1038/277182a0)
- Schlegel, D. J., Finkbeiner, D. P., & Davis, M. 1998, *ApJ*, 500, 525, doi: [10.1086/305772](https://doi.org/10.1086/305772)
- Schmidt, M. 1963, *Nature*, 197, 1040, doi: [10.1038/1971040a0](https://doi.org/10.1038/1971040a0)
- Schmidt, M., & Green, R. F. 1983, *ApJ*, 269, 352, doi: [10.1086/161048](https://doi.org/10.1086/161048)
- Schölkopf, B., Smola, A., & Müller, K.-R. 1997, in *Artificial Neural Networks — ICANN'97*, ed. W. Gerstner, A. Germond, M. Hasler, & J.-D. Nicoud (Berlin, Heidelberg: Springer Berlin Heidelberg), 583–588
- Schroedter, M., Badran, H. M., Buckley, J. H., et al. 2005, *ApJ*, 634, 947, doi: [10.1086/496968](https://doi.org/10.1086/496968)
- Schroff, F., Kalenichenko, D., & Philbin, J. 2015, in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 815–823, doi: [10.1109/CVPR.2015.7298682](https://doi.org/10.1109/CVPR.2015.7298682)

- Schwartz, D. A., Marshall, H. L., Lovell, J. E. J., et al. 2000, *ApJ*, 540, 69, doi: 10.1086/312875
- Schölkopf, B., Smola, A., & Müller, K.-R. 1998, *Neural Computation*, 10, 1299, doi: 10.1162/089976698300017467
- Searle, J. R. 1980, *Behavioral and Brain Sciences*, 3, 417–424, doi: 10.1017/S0140525X00005756
- Senno, N., Murase, K., & Mészáros, P. 2017, *ApJ*, 838, 3, doi: 10.3847/1538-4357/aa6344
- Sersic, J. L. 1968, *Atlas de Galaxias Australes*
- Seyfert, C. K. 1943, *ApJ*, 97, 28, doi: 10.1086/144488
- Shakura, N. I., & Sunyaev, R. A. 1973, *A&A*, 24, 337
- Shapiro, S. L., Lightman, A. P., & Eardley, D. M. 1976, *ApJ*, 204, 187, doi: 10.1086/154162
- Sharma, M., Nayak, J., Koul, M. K., et al. 2015, *Nuclear Instruments and Methods in Physics Research A*, 770, 42, doi: 10.1016/j.nima.2014.10.012
- Shastri, P., Wilkes, B. J., Elvis, M., & McDowell, J. 1993, *ApJ*, 410, 29, doi: 10.1086/172721
- Shaw, M. S., Romani, R. W., Cotter, G., et al. 2012, *ApJ*, 748, 49, doi: 10.1088/0004-637X/748/1/49
- . 2013, *ApJ*, 764, 135, doi: 10.1088/0004-637X/764/2/135
- Sikora, M., Begelman, M. C., Madejski, G. M., & Lasota, J.-P. 2005, *ApJ*, 625, 72, doi: 10.1086/429314
- Sikora, M., Begelman, M. C., & Rees, M. J. 1994a, *ApJ*, 421, 153, doi: 10.1086/173633
- . 1994b, *ApJ*, 421, 153, doi: 10.1086/173633
- . 1994c, *ApJ*, 421, 153, doi: 10.1086/173633
- . 1994d, *ApJ*, 421, 153, doi: 10.1086/173633
- Sikora, M., Błażejowski, M., Begelman, M. C., & Moderski, R. 2001, *ApJ*, 554, 1, doi: 10.1086/321329
- Sikora, M., Stawarz, Ł., & Lasota, J.-P. 2007, *ApJ*, 658, 815, doi: 10.1086/511972
- Singh, K. K., Bhatt, H., Bhattacharyya, S., et al. 2019, *Advances in Space Research*, 63, 766, doi: 10.1016/j.asr.2018.08.013
- Smith, D., Hooper, D., & Viereg, A. 2021, *J. Cosmology Astropart. Phys.*, 2021, 031, doi: 10.1088/1475-7516/2021/03/031
- Sober, E. 2015, *Ockham's Razors: A User's Manual* (Cambridge: Cambridge University Press)
- Sokolov, A., & Marscher, A. P. 2005, *ApJ*, 629, 52, doi: 10.1086/431321
- Sokolov, A., Marscher, A. P., & McHardy, I. M. 2004, *ApJ*, 613, 725, doi: 10.1086/423165

- Spada, M., Ghisellini, G., Lazzati, D., & Celotti, A. 2001, MNRAS, 325, 1559, doi: 10.1046/j.1365-8711.2001.04557.x
- Srivastava, N., Hinton, G. E., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. 2014, Journal of Machine Learning Research, 15, 1929
- Stecker, F. W. 2013, Phys. Rev. D, 88, 047301, doi: 10.1103/PhysRevD.88.047301
- Stecker, F. W., Done, C., Salamon, M. H., & Sommers, P. 1991, Phys. Rev. Lett., 66, 2697, doi: 10.1103/PhysRevLett.66.2697
- Steele, N. C., Reeves, C. R., & Gaura, E. I. 2001, in Artificial Neural Nets and Genetic Algorithms, ed. V. Kůrková, R. Neruda, M. Kárný, & N. C. Steele (Vienna: Springer Vienna), 27–30
- Stickel, M., Padovani, P., Urry, C. M., Fried, J. W., & Kuehr, H. 1991, ApJ, 374, 431, doi: 10.1086/170133
- Stoche, J. T., Morris, S. L., Weymann, R. J., & Foltz, C. B. 1992, ApJ, 396, 487, doi: 10.1086/171735
- Stone, E. M., Fingert, J. H., Alward, W. L., et al. 1997, Science (American Association for the Advancement of Science), 275, 668
- Stoughton, C., Lupton, R. H., Bernardi, M., et al. 2002, AJ, 123, 485, doi: 10.1086/324741
- Strittmatter, P. A., Hill, P., Pauliny-Toth, I. I. K., Steppe, H., & Witzel, A. 1980, A&A, 88, L12
- Suchkov, A. A., Hanisch, R. J., & Margon, B. 2005, The Astronomical Journal, 130, 2439, doi: 10.1086/497363
- Tadhunter, C. 2008, New Astron. Rev., 52, 227, doi: 10.1016/j.newar.2008.06.004
- Tamborra, I., Ando, S., & Murase, K. 2014, J. Cosmology Astropart. Phys., 2014, 043, doi: 10.1088/1475-7516/2014/09/043
- Tan, C., Xue, R., Du, L.-M., et al. 2020, ApJS, 248, 27, doi: 10.3847/1538-4365/ab8cc6
- Tanaka, Y. T., Doi, A., Inoue, Y., et al. 2015, ApJ, 799, L18, doi: 10.1088/2041-8205/799/2/L18
- Tavecchio, F. 2014, Monthly Notices of the Royal Astronomical Society, 438, 3255, doi: 10.1093/mnras/stt2437
- Taylor, C. 2019, Applications Of Dynamic Programming To Agricultural Decision Problems (CRC Press). <https://books.google.it/books?id=71SdWAAQBAJ>
- Taylor, G. B., Vermeulen, R. C., Readhead, A. C. S., et al. 1996, ApJS, 107, 37, doi: 10.1086/192354
- Tchekhovskoy, A., Narayan, R., & McKinney, J. C. 2011, MNRAS, 418, L79, doi: 10.1111/j.1745-3933.2011.01147.x
- The CTA Consortium, Actis, M., Agnetta, G., Aharonian, F., et al. 2011, Experimental Astronomy, 32, 193, doi: 10.1007/s10686-011-9247-0

- Tibshirani, R. 1996, *Journal of the Royal Statistical Society: Series B (Methodological)*, 58, 267, doi: <https://doi.org/10.1111/j.2517-6161.1996.tb02080.x>
- Tin Kam Ho. 1995, in *Proceedings of 3rd International Conference on Document Analysis and Recognition*, Vol. 1, 278–282 vol.1, doi: [10.1109/ICDAR.1995.598994](https://doi.org/10.1109/ICDAR.1995.598994)
- Tolamatti, A., Singh, K. K., & Yadav, K. K. 2023, *MNRAS*, 523, 5341, doi: [10.1093/mnras/stad1826](https://doi.org/10.1093/mnras/stad1826)
- Tomek, I. 1976, *IEEE Transactions on Systems, Man, and Cybernetics*, SMC-6, 769, doi: [10.1109/TSMC.1976.4309452](https://doi.org/10.1109/TSMC.1976.4309452)
- Tosti, G., Ballet, J., Digel, S. W., Burnett, T. H., & Fermi LAT Collaboration. 2011, in *AAS/High Energy Astrophysics Division*, Vol. 12, AAS/High Energy Astrophysics Division #12, 39.19
- Tukey, J. W. 1977, *Exploratory Data Analysis* (Addison-Wesley)
- Urry, C. M., & Padovani, P. 1995, *PASP*, 107, 803, doi: [10.1086/133630](https://doi.org/10.1086/133630)
- Urry, M. 2003, in *Astronomical Society of the Pacific Conference Series*, Vol. 290, Active Galactic Nuclei: From Central Engine to Host Galaxy, ed. S. Collin, F. Combes, & I. Shlosman, 3, doi: [10.48550/arXiv.astro-ph/0301309](https://doi.org/10.48550/arXiv.astro-ph/0301309)
- van den Bergh, S. 1998, *Galaxy Morphology and Classification*
- van Eldik, C., Holler, M., Berge, D., et al. 2015, in *International Cosmic Ray Conference*, Vol. 34, 34th International Cosmic Ray Conference (ICRC2015), 847
- Vanden Berk, D. E., Richards, G. T., Bauer, A., et al. 2001, *AJ*, 122, 549, doi: [10.1086/321167](https://doi.org/10.1086/321167)
- Vapnik, V. 2000, *The Nature of Statistical Learning Theory*, Vol. 8, 1–15, doi: [10.1007/978-1-4757-3264-1_1](https://doi.org/10.1007/978-1-4757-3264-1_1)
- Vapnik, V. N. 1998, *Statistical Learning Theory* (Wiley-Interscience)
- Varma, S., & Simon, R. 2006, *BMC bioinformatics*, 7, 91, doi: [10.1186/1471-2105-7-91](https://doi.org/10.1186/1471-2105-7-91)
- Vasudevan, R. V., & Fabian, A. C. 2009, *MNRAS*, 392, 1124, doi: [10.1111/j.1365-2966.2008.14108.x](https://doi.org/10.1111/j.1365-2966.2008.14108.x)
- VERITAS Collaboration, Badran, H. M., Blaylock, G., et al. 2005, in *American Institute of Physics Conference Series*, Vol. 745, High Energy Gamma-Ray Astronomy, ed. F. A. Aharonian, H. J. Völk, & D. Horns, 633–638, doi: [10.1063/1.1878475](https://doi.org/10.1063/1.1878475)
- Vermeulen, R. C., Readhead, A. C. S., & Backer, D. C. 1994, *ApJ*, 430, L41, doi: [10.1086/187433](https://doi.org/10.1086/187433)
- Véron-Cetty, M. P., & Véron, P. 2010, *A&A*, 518, A10, doi: [10.1051/0004-6361/201014188](https://doi.org/10.1051/0004-6361/201014188)
- Vestergaard, M. 2002, *ApJ*, 571, 733, doi: [10.1086/340045](https://doi.org/10.1086/340045)
- Vestergaard, M., & Peterson, B. M. 2006, *ApJ*, 641, 689, doi: [10.1086/500572](https://doi.org/10.1086/500572)
- Villata, M., Raiteri, C. M., Balonek, T. J., et al. 2006, *A&A*, 453, 817, doi: [10.1051/0004-6361:20064817](https://doi.org/10.1051/0004-6361:20064817)

- Virtanen, J. J. P., & Vainio, R. 2005, *ApJ*, 621, 313, doi: 10.1086/427324
- Wakely, S. P., & Horan, D. 2008, in *International Cosmic Ray Conference*, Vol. 3, *International Cosmic Ray Conference*, 1341–1344
- Wan, L., Zeiler, M., Zhang, S., Le Cun, Y., & Fergus, R. 2013in (Atlanta, Georgia, USA: PMLR), 1058–1066. <https://proceedings.mlr.press/v28/wan13.html>
- Wandel, A., Peterson, B. M., & Malkan, M. A. 1999, *ApJ*, 526, 579, doi: 10.1086/308017
- Wang, J.-M., Du, P., Valls-Gabaud, D., Hu, C., & Netzer, H. 2013, *Phys. Rev. Lett.*, 110, 081301, doi: 10.1103/PhysRevLett.110.081301
- Wang, J.-M., & Zhou, Y.-Y. 1999, *ApJ*, 516, 420, doi: 10.1086/307080
- Wang, J.-M., Du, P., Hu, C., et al. 2014, *ApJ*, 793, 108, doi: 10.1088/0004-637X/793/2/108
- Wang, K., Guo, P., Yu, F., et al. 2018, *International Journal of Computational Intelligence Systems*, 11, 575, doi: 10.2991/ijcis..11.1.43
- Wang, S., Liu, W., Wu, J., et al. 2016, in *2016 International Joint Conference on Neural Networks (IJCNN)*, 4368–4374, doi: 10.1109/IJCNN.2016.7727770
- Weaver, Z. R., Jorstad, S. G., Marscher, A. P., et al. 2022, *ApJS*, 260, 12, doi: 10.3847/1538-4365/ac589c
- Webb, G. I. 2010, *Occam’s Razor*, ed. C. Sammut & G. I. Webb (Boston, MA: Springer US), 735–735, doi: 10.1007/978-0-387-30164-8_609
- Weekes, T. C., Cawley, M. F., Fegan, D. J., et al. 1989, *ApJ*, 342, 379, doi: 10.1086/167599
- Weir, N., Fayyad, U. M., & Djorgovski, S. 1995, *AJ*, 109, 2401, doi: 10.1086/117459
- Weiss, G. M. 2004, *SIGKDD Explor. Newsl.*, 6, 7–19, doi: 10.1145/1007730.1007734
- . 2005, *Mining with Rare Cases*, ed. O. Maimon & L. Rokach (Boston, MA: Springer US), 765–776, doi: 10.1007/0-387-25465-X_35
- West, J., Ventura, D., & Warnick, S. 2007, *Journal of Software Engineering and Applications*, 12, 11
- White, R. L., Becker, R. H., Gregg, M. D., et al. 2000, *ApJS*, 126, 133, doi: 10.1086/313300
- Willott, C. J., Rawlings, S., Blundell, K. M., & Lacy, M. 1999, *MNRAS*, 309, 1017, doi: 10.1046/j.1365-8711.1999.02907.x
- Wills, B. J., Wills, D., Breger, M., Antonucci, R. R. J., & Barvainis, R. 1992, *ApJ*, 398, 454, doi: 10.1086/171869
- Wilms, J., Reynolds, C. S., Begelman, M. C., et al. 2001, *MNRAS*, 328, L27, doi: 10.1046/j.1365-8711.2001.05066.x
- Wilson, A. S., & Colbert, E. J. M. 1995, *ApJ*, 438, 62, doi: 10.1086/175054
- Wilson, A. S., Young, A. J., & Shopbell, P. L. 2001, *ApJ*, 547, 740, doi: 10.1086/318412

- Wilson, D. L. 1972, *IEEE Transactions on Systems, Man, and Cybernetics*, SMC-2, 408, doi: [10.1109/TSMC.1972.4309137](https://doi.org/10.1109/TSMC.1972.4309137)
- Windhorst, R., Odewahn, S., Burg, C., Cohen, S., & Waddington, I. 1999, *Astrophysics and Space Science*, 269-270, 243
- Woltjer, L. 1959, *ApJ*, 130, 38, doi: [10.1086/146694](https://doi.org/10.1086/146694)
- Xiao, H., Ouyang, Z., Zhang, L., et al. 2022a, *ApJ*, 925, 40, doi: [10.3847/1538-4357/ac36da](https://doi.org/10.3847/1538-4357/ac36da)
- Xiao, H., Zhu, J., Fu, L., Zhang, S., & Fan, J. 2022b, *PASJ*, 74, 239, doi: [10.1093/pasj/psab121](https://doi.org/10.1093/pasj/psab121)
- . 2022c, *PASJ*, 74, 239, doi: [10.1093/pasj/psab121](https://doi.org/10.1093/pasj/psab121)
- Xiao, H., Fan, J., Yang, J., et al. 2019, *Science China Physics, Mechanics, and Astronomy*, 62, 129811, doi: [10.1007/s11433-018-9371-x](https://doi.org/10.1007/s11433-018-9371-x)
- Xiao, H. B., Fan, J. H., Rando, R., Zhu, J. T., & Hu, L. J. 2020, *Astronomische Nachrichten*, 341, 462, doi: <https://doi.org/10.1002/asna.202013733>
- Xiao, H. B., Pei, Z. Y., Xie, H. J., et al. 2015, *Ap&SS*, 359, 39, doi: [10.1007/s10509-015-2433-1](https://doi.org/10.1007/s10509-015-2433-1)
- Xiao, H. B., Zhu, J. T., Fan, J. H., et al. 2022d, *MNRAS*, 517, 4202, doi: [10.1093/mnras/stac2959](https://doi.org/10.1093/mnras/stac2959)
- Xie, G. Z., Liu, F. K., Liu, B. F., et al. 1991, *A&A*, 249, 65
- Xie, L., Li, Z., Zhou, Y., He, Y., & Zhu, J. 2020, *Sensors*
- Xiong, D. R., & Zhang, X. 2014, *MNRAS*, 441, 3375, doi: [10.1093/mnras/stu755](https://doi.org/10.1093/mnras/stu755)
- Xue, R., Liu, R.-Y., Petropoulou, M., et al. 2019, *ApJ*, 886, 23, doi: [10.3847/1538-4357/ab4b44](https://doi.org/10.3847/1538-4357/ab4b44)
- Yan, D., & Zhang, L. 2015, *MNRAS*, 447, 2810, doi: [10.1093/mnras/stu2551](https://doi.org/10.1093/mnras/stu2551)
- Yang, J., Fan, J., Liu, Y., et al. 2023, *Science China Physics, Mechanics, and Astronomy*, 66, 249511, doi: [10.1007/s11433-022-2062-2](https://doi.org/10.1007/s11433-022-2062-2)
- Yang, J. H., Fan, J. H., Liu, Y., et al. 2022a, *ApJS*, 262, 18, doi: [10.3847/1538-4365/ac7deb](https://doi.org/10.3847/1538-4365/ac7deb)
- . 2022b, *ApJS*, 262, 18, doi: [10.3847/1538-4365/ac7deb](https://doi.org/10.3847/1538-4365/ac7deb)
- Yang, W.-X., Xiao, H.-B., Wang, H.-G., et al. 2022c, *Research in Astronomy and Astrophysics*, 22, 085002, doi: [10.1088/1674-4527/ac712c](https://doi.org/10.1088/1674-4527/ac712c)
- Yang, X., Wang, Y., Byrne, R., Schneider, G., & Yang, S. 2019, *Chemical Reviews*, 119, 10520, doi: [10.1021/acs.chemrev.8b00728](https://doi.org/10.1021/acs.chemrev.8b00728)
- Yip, C. W., Connolly, A. J., Vanden Berk, D. E., et al. 2004, *AJ*, 128, 2603, doi: [10.1086/425626](https://doi.org/10.1086/425626)
- York, D. G., Adelman, J., Anderson, John E., J., et al. 2000, *AJ*, 120, 1579, doi: [10.1086/301513](https://doi.org/10.1086/301513)
- Youden, W. J. 1950, *Cancer*, 3, 32, doi: [https://doi.org/10.1002/1097-0142\(1950\)3:1<32::AID-CNCR2820030106>3.0.CO;2-3](https://doi.org/10.1002/1097-0142(1950)3:1<32::AID-CNCR2820030106>3.0.CO;2-3)
- Yuan, C., Murase, K., & Mészáros, P. 2020, *ApJ*, 890, 25, doi: [10.3847/1538-4357/ab65ea](https://doi.org/10.3847/1538-4357/ab65ea)

- Yuan, F., & Narayan, R. 2014, *ARA&A*, 52, 529, doi: 10.1146/annurev-astro-082812-141003
- Zamfir, S., Sulentic, J. W., & Marziani, P. 2008, *MNRAS*, 387, 856, doi: 10.1111/j.1365-2966.2008.13290.x
- Zech, A., Cerruti, M., & Mazin, D. 2017, *A&A*, 602, A25, doi: 10.1051/0004-6361/201629997
- Zha, M., Xiang, G., Xing, Y., Zhou, J., & Wang, X. 2019, 830, doi: 10.22323/1.358.0830
- Zhang, H., Fang, K., Li, H., et al. 2019, *ApJ*, 876, 109, doi: 10.3847/1538-4357/ab158d
- Zhang, J., Liang, E.-W., Zhang, S.-N., & Bai, J. M. 2012a, *ApJ*, 752, 157, doi: 10.1088/0004-637X/752/2/157
- . 2012b, *ApJ*, 752, 157, doi: 10.1088/0004-637X/752/2/157
- Zhang, J., & Mani, I. 2003, in *Proceedings of the ICML'2003 Workshop on Learning from Imbalanced Datasets*
- Zhang, L., Chen, S., Xiao, H., Cai, J., & Fan, J. 2020, *ApJ*, 897, 10, doi: 10.3847/1538-4357/ab9180
- Zhang, L., Fan, J., & Zhu, J. 2021, *PASJ*, 73, 313, doi: 10.1093/pasj/psaa122
- Zhang, Y., & Zhao, Y. 2003, *Publications of the Astronomical Society of the Pacific*, 115, 1006, doi: 10.1086/376847
- Zhang, Y.-W., & Fan, J.-H. 2008, *Chinese J. Astron. Astrophys.*, 8, 385, doi: 10.1088/1009-9271/8/4/02
- Zheng, Y. G., Yang, C. Y., & Kang, S. J. 2016, *A&A*, 585, A8, doi: 10.1051/0004-6361/201526703
- Zhou, B., Kamionkowski, M., & Liang, Y.-f. 2021, *Phys. Rev. D*, 103, 123018, doi: 10.1103/PhysRevD.103.123018
- Zhu, C. R., Wang, G., Liu, B. L., et al. 2013, *Phys. Rev. B*, 88, 121301, doi: 10.1103/PhysRevB.88.121301
- Zhu, J. T., Lin, C., Xiao, H. B., et al. 2023, *The Astrophysical Journal*, 950, 123, doi: 10.3847/1538-4357/acca85