Opening the Personal Gate between Technology and Health Care
M. Giacomini et al. (Eds.)

doi:10.3233/SHTI260131

Environmental Personal Exposure Clusters to Investigate Multiple Sclerosis and Amyotrophic Lateral Sclerosis Progression

Pietro BOSONI [a,1], Mahin VAZIFEHDAN [a], Helena AIDOS [b], Inês ALVES [c], Giovanni BIROLO [d], Guglielmo FAGGIOLI [e], Sergio GONZÁLEZ-MARTÍNEZ [f], Marta GROMICHO [c], Aleksandar JOVANOVIĆ [g], Borko KOSTIĆ [g], Enrico LONGATO [e], Umberto MANERA [h], Eleonora TAVAZZI [i], Erica TAVAZZI [e], Riccardo BELLAZZI [a], Roberto BERGAMASCHI [i], Maria Fernanda CABRERA [f], Adriano CHIÒ [h], Mamede DE CARVALHO [c], Barbara DI CAMILLO [e], Piero FARISELLI [d], Nicola FERRO [e], Sara C. MADEIRA [b], and Arianna DAGLIATI [a]
[a] *Dept of Electrical, Computer and Biomedical Engineering, University of Pavia, Italy*
[b] *LASIGE, Faculdade de Ciências, Universidade de Lisboa, Portugal*
[c] *Faculty of Medicine, University of Lisbon, Portugal*
[d] *University of Turin, Italy*
[e] *Dept of Information Engineering, University of Padova, Italy*
[f] *Life Supporting Technologies, Universidad Politecnica de Madrid, Spain*
[g] *Belit, Serbia*
[h] *Rita Levi Montalcini Dept of Neuroscience, University of Turin, Italy*
[i] *IRCCS Mondino Foundation, Pavia, Italy*

Abstract. Reliable prognosis in Multiple Sclerosis (MS) and Amyotrophic Lateral Sclerosis (ALS) is hampered by data scarcity and variability. Beyond clinical variables, evidence suggests that environmental data can help capture disease trajectories. We investigated whether personal environmental measures can be organized into stable patterns that inform prognosis. In a multicenter cohort, 293 patients with MS or ALS were equipped with Atmotube air-quality sensors. We normalized volatile organic compound (VOC) time series and computed Dynamic Time Warping distances to capture temporal similarity. Hierarchical clustering yielded five daily exposure clusters, which were profiled using Atmotube variables (season, day type, humidity, temperature) and patient self-reports (work status, time outdoors), and evaluated by day-level differences between personal and fixed-station variables. These clusters can support interpolation of missing wearable intervals and generation of context-aware exposure estimates, thereby strengthening environmental inputs for prognostic modeling in MS and ALS.

Keywords. Environmental Data, Wearable Device, Exposure Trajectories, Sclerosis

1. Introduction

Reliable prognostic models are essential for clinical care but remain difficult to develop when data are scarce, heterogeneous, or fragmented, particularly in rare or fast-

[1] Corresponding Author: Pietro Bosoni, pietro.bosoni@unipv.it.

progressing disorders [1]. Multiple Sclerosis (MS) and Amyotrophic Lateral Sclerosis (ALS) are progressive neurological diseases with wide variability in course and symptoms. Beyond standard clinical variables, growing evidence suggests that environmental data may help capture disease states and trajectories, motivating the integration of multi-source, real-world signals to enrich prognosis [2].

Environmental exposure reflects pollutant loads, meteorological conditions, and daily routines. Fixed monitoring stations provide continuous signals but lack individual context and do not reflect indoor conditions, while wearable sensors may be unavailable or unreliable for extended periods. Therefore, merging wearable-sensor readings with fixed air quality and weather data is crucial to generate continuous exposure estimates that distinguish indoor from outdoor settings and bridge gaps when wearable streams are missing or noisy.

In this work, we derived clusters of personal exposure that summarize profiles across individuals and contexts. Integrated with clinical records, lifestyle information, and wearable signals, these clusters can guide interpolation of missing intervals and assign unlabeled periods to the most similar profiles, reducing uncertainty in individual exposure estimates and strengthening their value for prognostic modeling.

2. Methods

2.1. Study Design and Disease Progression Modelling

Patients affected with MS and ALS were enrolled in a three-year, multicenter prospective study at four sites within the BRAINTEASER project: Servicio Madrileño de Salud (SERMAS, Madrid, Spain); Instituto de Medicina Molecular (iMM, Lisbon, Portugal); Azienda Ospedaliero-Universitaria Città della Salute e della Scienza di Torino (UNITO, Turin, Italy); and IRCCS Mondino Foundation (MNDN-PV, Pavia, Italy). Disease progression was assessed with the Expanded Disability Status Scale (EDSS) for MS [3] and the ALS Functional Rating Scale-Revised (ALSFRS-R) for ALS [4]. Additional endpoints included relapse occurrence (for MS) and time to Non-Invasive Ventilation (NIV), Percutaneous Endoscopic Gastrostomy (PEG), and death (for ALS).

Participants received two monitoring devices: an Atmotube Pro wearable sensor for environmental monitoring and a Garmin Vivoactive 4S smartwatch to capture daily activity, stress, sleep, physical activity, and heart rate. Patients and clinicians also used two different dedicated tools developed as part of the BRAINTEASER project to complete disease-specific and general questionnaires: the Patient App, a mobile application designed for patients, and the Clinical Tool, a web-based dashboard for clinicians.

2.2. Prospective Data

The BRAINTEASER prospective dataset integrated routine clinical information with continuous monitoring from three sources: (i) environmental exposures captured by the Atmotube sensor and augmented with fixed air-quality stations and remote weather data, downloaded from the Air Quality database [5]; (ii) physiological signals from the Garmin smartwatch; and (iii) questionnaires and ancillary metadata collected via the BRAINTEASER dedicated tools. Data were ingested via a dedicated API and stored on BRAINTEASER servers. The database linked environmental exposure to individual

profiles while ensuring anonymization. For privacy, GPS data from Atmotube users were not retained

Specifically, the Clinical Tool provided 13 questionnaires, while patients complete 13 additional questionnaires using the Patient App. The Patient App also recorded lifestyle and work-related context to refine exposure estimates, such as day/night schedule, indoor/outdoor work setting, workplace proximity to home (≤50 km), work modality (home/onsite/hybrid), and routine workday details. Patients could report time away from home in the previous four weeks and, optionally, the temporary postcode.

2.3. Data Labeling, Fairification: The BRAINTEASER Ontology

The clinical course and anamnestic history of patients, together with questionnaire data and continuous streams from wearable devices, were modelled within the BRAINTEASER Ontology (BTO) [6]. BTO is openly available, specified in OWL 1.2, and used a stable identifier space. It was designed to adhere to the Open Biological and Biomedical Ontologies (OBO) and the Findable, Accessible, Interoperable, and Reusable (FAIR) principles, and was developed through a co-design process in close collaboration with clinicians and domain experts, embedding their knowledge and validating each design decision. In particular, and in contrast to disease-specific schemas, BTO emphasizes patients' trajectories and lifestyles rather than underlying biological mechanisms, offering a single, coherent conceptual view across both conditions.

2.4. Computation of Personal Exposure Clusters

Patients' Atmotube measurements were matched to fixed-station measurements by residential postcode. Since Volatile Organic Compounds (VOCs) are, in general, strong discriminators of indoor versus outdoor environments [7], we restricted this exploratory analysis to days with a complete 24-hour Atmotube VOC time series and an available residential postcode, from February to August 2024 (winter–summer). To limit seasonal variability and subject bias, we randomly sampled up to 20 complete days per season per patient and stratified by weekday and weekend. We computed a Dynamic Time Warping (DTW) distance matrix across all pairs of daily normalized VOC time series to capture temporal similarity in environmental exposure. We then explored candidate clustering solutions, including Hierarchical Clustering, Partitioning Around Medoids (PAM), DBSCAN, and Spectral Clustering. Clusters were qualitatively profiled using additional Atmotube-derived variables (season, day type, humidity, temperature) and Patient App self-reported information (work status, time spent outdoors). In addition, we evaluated the clusters by quantifying day-level differences in humidity and temperature between Atmotube readings and fixed-station measurements.

3. Results and Discussion

3.1. Environmental Prospective Data

In this prospective cohort, 289 patients equipped with Atmotube sensors were enrolled on a rolling basis from September 2022 through December 2024. Sensor measurements and baseline demographics are summarized in Table 1. Median age differed across the

four clinical centers as assessed using the Kruskal–Wallis test. Post hoc pairwise Wilcoxon rank-sum tests with Benjamini–Hochberg adjustment indicated significant differences for MNDN-PV and UNITO compared to all centers, and for SERMAS and iMM (p-value < 0.0001). Age distributions also differed significantly by sclerosis type according to the Kruskal–Wallis test (p-value < 0.0001).

Table 1. Distribution of patients with the Atmotube sensor. Statistics are presented as count (percentage) and median [interquartile range].

Variable	MNDN-PV (N=61)	UNITO (N=78)	iMM (N=45)	SERMAS (N=109)	Overall (N=293)
Gender					
Female	41 (67%)	46 (59%)	14 (31%)	57 (52%)	158 (54%)
Male	20 (33%)	32 (41%)	31 (69%)	52 (48%)	135 (46%)
Sclerosis					
ALS	0 (0%)	24 (31%)	45 (100%)	39 (36%)	108 (37%)
MS	61 (100%)	54 (69%)	0 (0%)	70 (64%)	185 (63%)
Age (years)	42 [14]	45.5 [20.25]	61 [18]	49 [21]	49 [20]
Atmotube measure	154,542 [242,590]	49,641 [167,937]	219,129 [331,152]	275,347 [425,697]	162,489 [355,361]

3.2. Personal Exposure Clusters

Seventy-seven patients met the inclusion criteria, contributing 3,257 patient-days distributed across seasons, including 1,183 in spring (36.3%), 956 in summer (29.4%), and 1,118 in winter (34.3%), and across day types, including 2,342 weekdays (71.9%) and 915 weekends (28.1%). To partition daily exposure profiles, we selected Hierarchical Clustering with Ward linkage, which yielded low-variance groups and an interpretable dendrogram. Guided by dendrogram cut heights and the average Silhouette score, we identified five exposure clusters, which are reported with their counts in Table 2. As shown in Figure 1, Cluster C1 (pink) is characterized by the highest daily average VOC concentrations, followed by C2 (dark green). In contrast, C4 (light blue) and C5 (dark blue) exhibit lower VOC levels. Considering additional Atmotube-derived variables, humidity is highest in C1 and C5 and lowest in C3, while temperature is lowest in C1 and highest in C3 and C5. User-reported information indicates that C1 is enriched for currently working participants, predominantly onsite and spending less time outdoors. Finally, when comparing personal and fixed-station measurements, C1 shows the largest day-level discrepancies linked to residential postcodes, while C5 shows the smallest differences. Humidity discrepancies are broadly similar across clusters.

Table 2. Summary of the number of distinct days and patients in each identified cluster.

Cluster	Days	Unique Patients
C1 (Pink)	499	45
C2 (Dark Green)	447	56
C3 (Light Green)	572	65
C4 (Light Blue)	658	64
C5 (Dark Blue)	1,081	62

4. Conclusions

In 77 patients contributing 3,257 Atmotube-monitored days, we applied DTW to normalized daily VOC time series and identified interpretable clusters that capture recurrent exposure patterns across individuals and contexts. We profiled clusters using

clinical records, patient-app self-reports, and day-level differences in humidity and temperature between personal and postcode-matched fixed-station readings. Limitations include restricting analyses to adherent users, and relying on VOCs as indoor/outdoor discriminators without ground-truth labels. Nevertheless, these clusters can guide imputation of missing wearable intervals, enabling continuous, context-aware exposure estimates that distinguish indoor from outdoor conditions and strengthening environmental inputs for MS and ALS prognostic modeling.

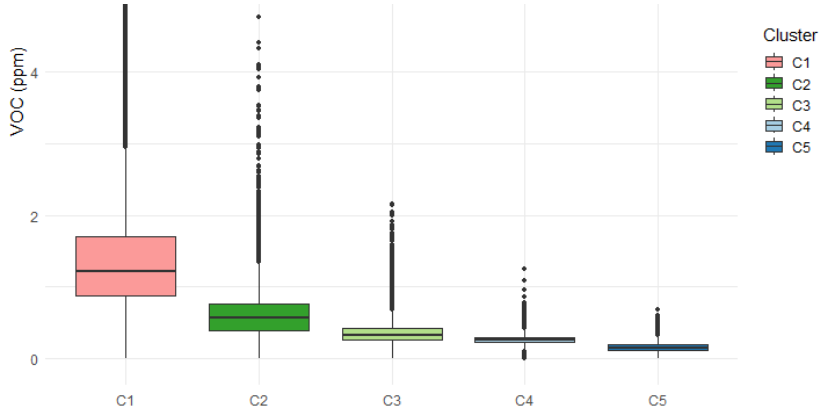

Figure 1. Distribution of Atmotube daily VOC average values (in parts per million) for each cluster.

Acknowledgments

This research was supported by the BRAINTEASER (BRinging Artificial INTelligencE home for a better cAre of amyotrophic lateral sclerosis and multiple SclERosis) project, funded by the European Union's Horizon 2020 research and innovation programme (Grant Agreement number: 101017598).

References

[1] Schaefer J, Lehne M, Schepers J, Prasser F, Thun S. The use of machine learning in rare diseases: a scoping review. Orphanet J Rare Dis. 2020 Jun 9;15(1):145.

[2] Gashi S, Oldrati P, Moebus M, Hilty M, Barrios L, Ozdemir F, Kana V, Lutterotti A, Rätsch G, Holz C. Modeling multiple sclerosis using mobile and wearable sensor data. Npj Digit Med. 2024 Mar 11;7(1):64.

[3] Kurtzke JF. Rating neurologic impairment in multiple sclerosis: an expanded disability status scale (EDSS). Neurology. 1983 Nov;33(11):1444–52.

[4] Cedarbaum JM, Stambler N, Malta E, Fuller C, Hilt D, Thurmond B, Nakanishi A. The ALSFRS-R: a revised ALS functional rating scale that incorporates assessments of respiratory function. BDNF ALS Study Group (Phase III). J Neurol Sci. 1999 Oct 31;169(1–2):13–21.

[5] Faggioli G, Menotti L, Marchesin S, Trescato I, Ahmad L, Aidos H, Alungulese AL, Bellazzi R, Bergamaschi R, Birolo G, Bosoni P, et al. The BRAINTEASER Datasets: Clinical, Wearable and Environmental Data for ALS & MS Progression Modeling. Sci Data. 2025 Nov 21;12(1):1854.

[6]. Faggioli G, Menotti L, Marchesin S, Chió A, Dagliati A, de Carvalho M, Gromicho M, Manera U, Tavazzi E, Di Nunzio GM, Silvello G, Ferro N. An extensible and unifying approach to retrospective clinical data modeling: the BrainTeaser Ontology. J Biomed Semant. 2024 Aug 30;15(1):16.

[7]. Su FC, Mukherjee B, Batterman S. Determinants of personal, indoor and outdoor VOC concentrations: An analysis of the RIOPA data. Environ Res. 2013 Oct 1;126:192–203.