Shapley Value as an Aid to Biomedical Machine Learning: a Heart Disease Dataset Analysis

Daniele Scapin[1], Giulia Cisotto[2,3], Elvina Gindullina[1], and Leonardo Badia[1,3]

[1] Dept. of Information Engineering (DEI), University of Padova, Italy
[2] Dept. of Informatics, Systems and Communication (DISCo), University of Milano-Bicocca, Italy
[3] Inter-University Consortium for Telecommunications (CNIT), Italy
email: daniele.scapin.1@studenti.unipd.it, giulia.cisotto@unimib.it, {elvina.gindullina, leonardo.badia}@unipd.it

Abstract—This paper investigates the decision making process aided by machine learning for biomedical problems and how to improve it through meta assessments of the most relevant features. Classification algorithms are usually trained and exploited with high dimensional datasets (i.e., with an extremely large number of features), which is inefficient and costly. It would be beneficial to identify the most meaningful features that contribute the most to assigning a category to a subject, and in particular, diagnosing a pathological condition. A helpful support can come from cooperative game theory, through the computation of the Shapley value, an indicator of desirable properties according to which the players, in our case the input features, can be ranked. We apply such a framework to a supervised machine learning scenario of a random forest tree classifier applied to heart disease detection. From a publicly available dataset, we identify the most relevant features that can affect the decision, thus obtaining practical guidelines for a compact yet efficient description based on an analytical rationale.

Keywords-Artificial intelligence; Game theory; Heart disease; Machine learning; Parametric analysis; Random forest classifier; Shapley value.

I. INTRODUCTION

In the last years, we faced a convergence of Information and Communication Technologies (ICT) towards supporting healthcare and medical practices [1], [2]. A meeting point of particular relevance is represented by Machine Learning (ML) aided medical diagnostics, where artificial intelligence techniques are used as effective tools to support medical decisions and make them systematic [3]. These procedures can be applied on top of medical instrumentations that already bridge ICT engineering principles, such as imagining, signal processing, and electromagnetic technologies, and biological systems, so as to realize a comprehensive stack of diagnostic procedures through ML that can be fully automated [4].

However, the most common approaches in this field just make use of a brute-force approach where large datasets with multi-feature information are fed to the decision engine, which makes for a generally cumbersome method, difficult to acquire, manage, and possibly extend with a broader clinical sample. For this reason, an analysis of the involved input features is advisable, which can lead to a cross-layer optimization of the entire system, also involving acquisition and storage [5].

Among other feature selection algorithms (the interested reader can refer to [6]), cooperative Game Theory, in particular using the *Shapley value* [7], can be used to identify the most meaningful features that contribute the most to the decision process. The general framework of cooperative games involves players explicitly cooperating by forming *coalitions*, from which they obtain a certain overall gain. The Shapley value is a quantitative evaluation of how players inside a coalition are instrumental to reaching that outcome and are therefore more important in keeping the coalition together. It is computed as the average surplus generated by the presence of players in the coalitions [8]. If we regard the different input features as players in a coalitional game, whose outcome is to attain a certain classification, the Shapley value can be regarded as how critical a given parameter was in eventually reaching the decision. Compared to other feature selection methods, this approach is the de facto state of the art of explainable machine learning, thus being able to provide both local and glocal explanations on the decisions taken by the classification algorithm, with very good explainability properties, such as high expressive power, high portability, good human comprehensibility, agnostic modelling [9].

We apply this approach in the context of supervised ML, for a random forest classifier [10], whose goal is to obtain a binary classification of general pathological heart conditions. In other words, our classifier recognizes output values in $\{0, 1\}$ for each patient, i.e., free from or affected by a heart disease, respectively. We consider a dataset of clinical features, containing for example values of "cholesterol" and "resting blood pressure", and we compute the Shapley value of each of them to the final diagnosis, calculated as the average marginal contribution of a feature value across all possible coalitions, meant as the possible subset of features.

These computations are performed through SHAP (SHapley Additive exPlanations), a software tool to explain the output of any machine learning model through the classic Shapley value from game theory [11]. From a data analytics standpoint, our investigation is capable of gaining additional

insights from raw data, highlighting those who are most important and better able to contribute to the decision. It allows for more compact representations and/or uncovering correlations [12]. On a more medical level, this can also shed light on what are the critical factors (i.e., features) that are highly correlated with a heart disease, or what parameters are less relevant from a diagnostic perspective. More in general, our investigation provides insight on the value of information, hinting at the derivation of powerful instruments to effectively capture outliers and corrupted entries, and what type of new data to acquire to improve the prediction [13].

The rest of this paper is organized as follows. Section II reviews the theoretical fundamentals of collaborative games and the idea of Shapley value, as well as related work. In Section III, we describe our approach and expand the details of our data, their processing, and the application of the classification procedure. In Section IV, we discuss numerical evaluations, detailing the analysis performed with two approaches according to the availability of a key feature (the "cholesterol" attribute in our dataset). Finally, Section V concludes the paper and discusses future work.

II. STATE OF THE ART

In this section, we provide a background on cooperative games and the Shapley value, and we review related work about the application of these paradigms to machine learning approaches for biomedical problems.

A. Cooperative games

A cooperative (often called coalitional) game consists of a finite set of players $\mathcal{N} = \{x_1, x_2, \ldots, x_N\}$, interacting with the goal of forming coalitions. It differs from standard (non-cooperative) game theory setups, where players instead act individually and collaboration is only indirectly obtained through bargaining or credible threats [14]. There, game theory is used to predict the outcome of a certain interaction. Conversely, in collaborative games the very interaction that is under scrutiny is the formation itself of coalitions, that are subsets of $\mathcal{N}$ (sometimes called the *grand coalition*, since it can be seen as a coalition of all the players).

The overall *gain* of a coalition is described through a characteristic function $v : \wp(\mathcal{N}) \to \mathbb{R}$, such that $v(\mathcal{S})$ corresponds to the collective payoff that the set of players $\mathcal{S} \subseteq \mathcal{N}$ creates by forming a specific coalition. The characteristic function satisfies $v(\emptyset) = 0$; if a non-empty coalition $\mathcal{S} \neq \emptyset$ is formed, and player $j \in \mathcal{S}$, it is interesting to investigate whether j can advocate a share of $v(\mathcal{S})$ as a deserved contribution, for simply being instrumental in forming the coalition [15]. This is a sensible point if $v(\mathcal{S} \setminus \{j\}) < v(\mathcal{S})$, as will be argued next; that is, if the presence of j in the coalition increases its payoff, player j can rightly claim a share of the payoff, which can be determined through the Shapley value.

Before discussing how this happens, we remark that this abstract rationale can be translated in a quantification of the importance of a given feature in a ML-aided classification problem [16]. Given a classifier that uses a set of input features $\mathcal{N} = \{x_1, x_2, \ldots, x_N\}$, these can be considered as players in a coalitional game where the payoff $v(\mathcal{S})$ of a given subset of features is some measure of influence of that coalition in the decision process. Thus, the goal is to ultimately relate this to an individual quantification of importance for each specific feature through the Shapley value.

B. The Shapley value

The Shapley value of the jth player under characteristic function v is denoted as $\phi_j(v)$, and we will interpret it as the importance of the jth *feature* used for classification, where v is the payoff achieved by the classifier, i.e., some measure of goodness of the overall assignment to a class.

The value of $\phi_j(v)$ can be seen as a way to split v among the members of a coalition, which is referred to as the efficiency property. It also satisfies additional properties such as symmetry, null player, and additivity axioms, discussed in the following. Indeed, the idea of the Shapley value is to be a fair retribution to the members of a coalition $\mathcal{S}$, assuming they get a physical reward $v(\mathcal{S})$ from the existence of the coalition and want to share it. This reconnects coalitional games with standard non-cooperative games where players are moved by selfish objectives [17]; in particular, they join a coalition $\mathcal{S}$ because they are interested in getting a share of $v(\mathcal{S})$.

The Shapley value $\phi_j(v)$ of player j under payoff rule v is computed as the incremental benefit that j brings to possible coalitions, weighted and averaged over all of them, and the ways in which they can be formed through subsequent inclusions of one player at a time. This leads to [8]

$$\phi_j(v) = \sum_{\mathcal{S} \subseteq \mathcal{N} \setminus \{j\}} \frac{|\mathcal{S}|!(N - |\mathcal{S}| - 1)!}{N!} \left(v(\mathcal{S} \cup \{j\}) - v(\mathcal{S}) \right).$$

The Shapley value can be shown to satisfy the following properties [16].

Symmetry: The contributions of two players (features) j and k should be the same if they contribute equally to all possible coalitions, if

$$v(\mathcal{S} \cup \{j\}) = v(\mathcal{S} \cup \{k\}) \qquad \forall \mathcal{S} \subseteq (\mathcal{N} \setminus \{j, k\})$$

then $\phi_j(v) = \phi_k(v)$.

Null player: A player (feature) j that does not change the goodness of the coalition (i.e., the classification performance), regardless of which coalition it is added to, should have a Shapley value of 0, i.e., if

$$v(\mathcal{S} \cup \{j\}) = v(\mathcal{S}) \qquad \forall \mathcal{S} \subseteq (\mathcal{N} \setminus \{j\})$$

then $\phi_j(v) = 0$.

Table I

DATASET SAMPLE. THE FIRST 10 PATIENTS ARE REPORTED WITH DEMOGRAPHIC INFORMATION AND VALUES FOR THE 11 AVAILABLE FEATURES.

	Age	Sex	ChestPainType	RestingBP	Cholesterol	FastingBS	RestingECG	MaxHR	ExerciseAngina	Oldpeak	ST_Slope	HeartDisease
0	40	M	ATA	140	289	0	Normal	172	N	0.0	Up	0
1	49	F	NAP	160	180	0	Normal	156	N	1.0	Flat	1
2	37	M	ATA	130	283	0	ST	98	N	0.0	Up	0
3	48	F	ASY	138	214	0	Normal	108	Y	1.5	Flat	1
4	54	M	NAP	150	195	0	Normal	122	N	0.0	Up	0
5	39	M	NAP	120	339	0	Normal	170	N	0.0	Up	0
6	45	F	ATA	130	237	0	Normal	170	N	0.0	Up	0
7	54	M	ATA	110	208	0	Normal	142	N	0.0	Up	0
8	37	M	ASY	140	207	0	Normal	130	Y	1.5	Flat	1
9	48	F	ATA	120	284	0	Normal	120	N	0.0	Up	0

Additivity: given two coalitional games on the same set of players $\mathcal{N}$, described by characteristic functions v and w, if we consider a third game with payoff rule $v + w$, this summation reflects on the Shapley value of the players

$$\phi_j(v + w) = \phi_j(v) + \phi_j(w) \qquad \forall j \in \mathcal{N}$$

A cost-sharing rule based on the Shapley value, as defined above, minimizes the price of anarchy in the case of concave costs (or concave utilities) and thus it can be reinterpreted as a good strategic decision rule to combine individual criteria into the best choice from a joint standpoint [7].

C. Related work

Despite the aforementioned theoretical properties of the Shapley value that make it very desirable in principle, there are some issues in its computation and also different ways to evaluate it based on the model assumptions and the way training data are used. First of all, the basic idea of the Shapley value is better understood when the feature element is binary, that is, it can be seen as a vote in favor or against a certain decision. Generally, the involved features for a ML approach, especially in the biomedical field, give instead continuous values. In addition, other computational issues are present such as the NP-hardness of an exact exhaustive evaluation [18].

Moreover, for computations based on a preliminary training through part of the dataset, it is possible to re-evaluate the Shapley values after a retraining, or apply similar artifices to improve the computation [12]. Fortunately, as we will show in the following, our analysis is almost unaffected by this issue.

For our analysis, we make use of SHAP, as done by many studies in the field [11], [19]. Based on the general theory of the Shapley value, SHAP is a practical package in Python that is able to calculate the Shapley value for different ML models, in particular, tree-based classifiers [13].

There are many ways to implement the Shapley value in ML, with differences about the ML approach, the use of training data, and the overall offered explanations [20]. Hence, the implications about using Shapley value in a given biomedical problem are non-trivial and, as a result, our analysis can only be seen as one piece of a larger

puzzle, where the conclusions that we draw from our specific data are clearly worth of more investigations with broader datasets. Nevertheless, we believe that the consequences for clinicians are relevant.

In the biomedical field, a similar approach has been performed, for example, in relation to anxiety and its triggers when automatically classified through ML [21]. This is a relevant analysis since mental illness states, such as the anxiety disorder (as characterized by irrational fear of certain things or events) can be related to myriads of underlying causes and subsequent effects, and trigger different quantitative disfunctions. Thus, achieving an efficient representation and relating the parameters to one another is surely important.

Another similar study is [22], where Shapley values are employed in EEG classification task, i.e., motor tasks such that of [23], [24], for early detection of active movements from EEG and support by prosthetics and rehabiltation tools.

The study of [25] uses the Shapley value derived through a similar framework for a meta analysis of ML applied to biomedical data to forecast post-traumatic stress disorder, allowing for preventive interventions in the aftermath of a stressful event. Another analysis using SHAP to a similar end is [26], where the goal is to highlight hidden gene-gene and gene-environment interactions in genome-wide association.

III. PROPOSED APPROACH

We combined multiple datasets from Kaggle [27] by considering 5 datasets of cardiovascular diseased patients with 11 common features for a total of 918 observations.

A. Data overview

A general overview of the dataset is now provided. Its 11 features are reported in Table I, including: Age in years; Sex; ChestPainType (TA: Typical Angina, ATA: Atypical Angina, NAP: Non-Anginal Pain, ASY: Asymptomatic); the resting blood pressure (RestingBP) in mmHg; the serum Cholesterol in mg/dl; a parameter labeled as FastingBS that describes the fasting blood sugar f, specifically as $\chi(f > 120 \text{ mg/dl})$, where χ is the characteristic function, equal to 1 if the condition is true, and 0 otherwise; the resting ECG result RestingECG [Normal or ST, i.e., having ST-T

Table II
FIRST 10 DATA NORMALIZED

	Age	RestingBP	Cholesterol	FastingBS	MaxHR	Oldpeak	Sex_M	ChestPainType_ASY	RestingECG_Normal	ExerciseAngina_Y	ST_Slope_Up
0	0.244898	0.433962	0.596215	0.0	0.770992	0.015873	1.0	0.0	1.0	0.0	1.0
1	0.428571	0.622642	0.252366	0.0	0.648855	0.174603	0.0	0.0	1.0	0.0	0.0
2	0.183673	0.339623	0.577287	0.0	0.206107	0.015873	1.0	0.0	0.0	0.0	1.0
3	0.408163	0.415094	0.359621	0.0	0.282443	0.253968	0.0	1.0	1.0	1.0	0.0
4	0.530612	0.528302	0.299685	0.0	0.389313	0.015873	1.0	0.0	1.0	0.0	1.0
5	0.224490	0.245283	0.753943	0.0	0.755725	0.015873	1.0	0.0	1.0	0.0	1.0
6	0.346939	0.339623	0.432177	0.0	0.755725	0.015873	0.0	0.0	1.0	0.0	1.0
7	0.530612	0.150943	0.340694	0.0	0.541985	0.015873	1.0	0.0	1.0	0.0	1.0
8	0.183673	0.433962	0.337539	0.0	0.450382	0.253968	1.0	1.0	1.0	1.0	0.0
9	0.408163	0.245283	0.580442	0.0	0.374046	0.015873	0.0	0.0	1.0	0.0	1.0

wave abnormality (T wave inversions and/or ST elevation or depression of more than 0.05 mV)]; the binary variable LVH showing probable or definite left ventricular hypertrophy according to Romhilt-Estes criteria [28]; the maximum heart rate (MaxHR) in Hz as an integer between 60 and 202; the presence of exercise-induced angina; and finally, two parameters related to the ST segment of the ECG, namely, oldpeak = the ST [Numeric value measured in depression] and the slope of the ST peak segment [Up, Flat, or Down]. The last column is 1 for patients affected from a heart disease, 0 otherwise.

B. Data Preprocessing

To manipulate these data, we preprocessed them in order to convert them all into numerical values and a normalization was applied, so having all features values belonging to the interval $[0, 1]$, see Table II. Additionally, some values were removed because they were found to be of little significance or leading to needless complications in the evaluation. For example, the Down value of the parameter "ST slope" is rarely encountered in the database and its implications are very similar to the value Flat. Assigning a separate value was found to making less efficient predictions, so we just considered a binary choice as of "Flat" (merging it with "Down") or "Up." Analogous procedure was applied to "Chest pain type", "Exercise angina," and "RestingECG", reporting only the "ASY", "Yes'," and "Normal" type, respectively, or their negative corresponding values.

An important remark concerns the feature "Cholesterol". In the dataset, some data (about 170 entries) have this input parameter equal to 0, meaning that data was not measured in the patient. Thus, to provide a correct model prediction we would have two alternatives: either to discard the patients who lack this specific feature, or to remove the feature for all the data. Since this parameter may actually be significant for the diagnosis of a heart condition, we decided to consider both approaches and therefore, perform two different investigations; apart from this parameter, everything else follows the same approaches and, includes the same features in both cases. This also gives us the opportunity to discuss a further detailed exploration of the importance of a specific feature and how it may affect the evaluation of others through the

Shapley values, to see if, for the problem at hand, they are mutually interdependent, or redundant, or disconnected [13].

C. Classifier

A random forest classifier with 20 decision trees was trained to classify the data. It uses multiple decision trees with random permutations of $\lfloor \log_2 N \rfloor = 3$ combinations of the N features, which are considered as the players of the coalitional games, for which we compute the Shapley value through the SHAP toolbox [11]. We train each tree on a different part of the training set and average the final prediction output. This produces a small increase of the prediction bias, but also a lower variance.

To have a correct classification, we randomly split the dataset into 80% of training and 20% of validating data. We trained a RF classifier which obtained training score very close to 1, a validation score ~ 0.85 and mean square error of about 0.4. Then, we considered the feature ranking after validation of the RF model. We also remark that the convergence of the training phase is fast and robust, which seems to imply that there is no special need for a retraining in the problem at hand [12].

IV. RESULTS AND DISCUSSION

We now discuss the results of the Shapley values for the trained model. These first results refer to the dataset with discarded data, i.e., patients without the Cholesterol feature.

Fig. 1 explains the "Global feature importance", i.e., for each feature, the mean absolute Shapley value is plotted.

A more detailed plot is the one in Fig. 2. In this figure, we arrange the features in decreasing order of their importance (as in Fig. 1), the horizontal location of the dot corresponds to the Shapley value of the feature, indicating also if it contributes positively or negatively to the prediction. The color of each dot represents its normalized value. We have values close to 1 plotted in red, while values close to 0 are plotted in blue, with various shades indicating intermediate values.

Based on the Shapley value, we can establish a hierarchy of features by their importance. The "ST_Slope" appears to be the most important, as also hinted by clinical trials such as [29], followed by an asymptomatic pain type and
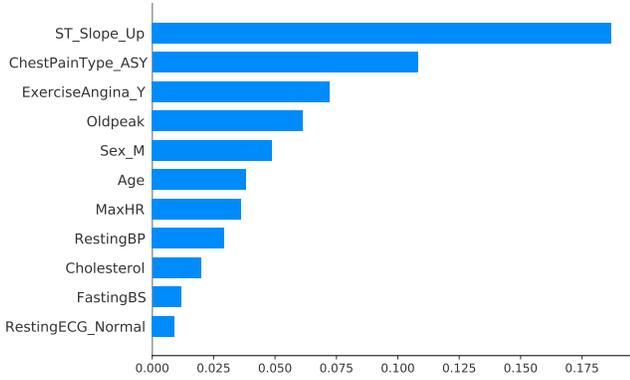
Figure 1. Global feature importance for the dataset without subjects with zero-valued Cholesterol feature.
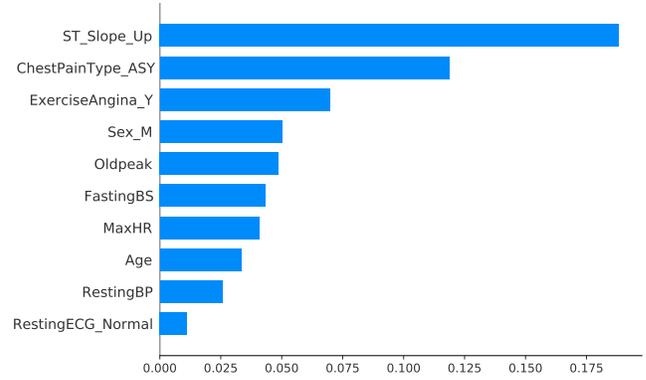

Figure 3. Global feature importance for the dataset with the complete removal of the Cholesterol features.
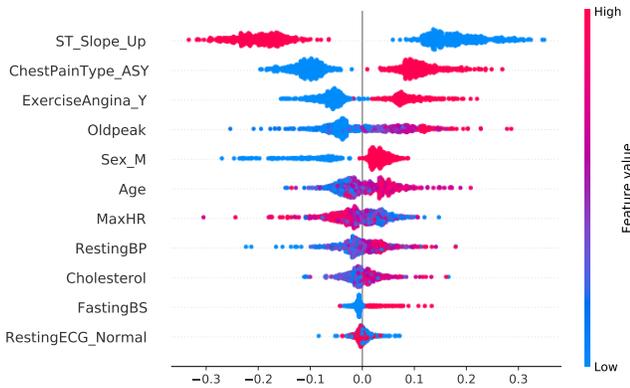

Figure 2. Local explanation summary for the dataset without subjects with zero-valued Cholesterol feature.
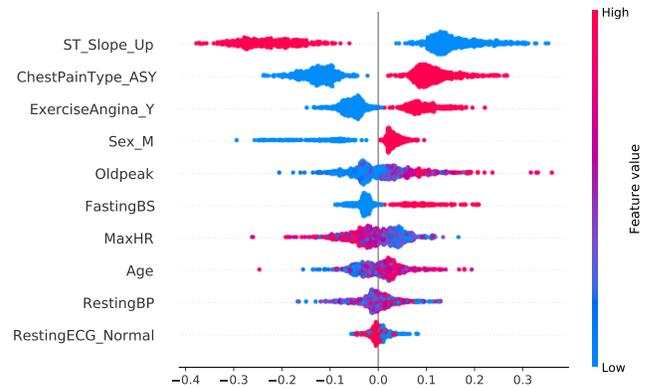

Figure 4. Local explanation summary for the dataset with the complete removal of the Cholesterol features.

the presence of angina induced by exercise. It is expected to find a patient suffering from a heart disease whenever "ST_Slope" is "Up", no matter what the values of the other features are. The other two binary features are also, considered alone, highly correlated with a disease. In fact, these are generally regarded to be dangerous symptoms and, as can be seen in Fig. 2, we have a clear distinction on the contribution of each of them. In this study case, we have almost no confusing measurements for the predicted label (i.e., red dots inside a blue bulk or vice-versa), or when we do, their value is extremely low.

Focusing on the other attributes, we give a short explanation of one of them (the reasoning is the same for the other features). Taking for example the maximum heart-rate (MaxHR): the lower its value, the higher the risk for the subject to be affected by a heart condition. We can notice also a coalitional effect for the less important features (with lower Shapley value), that is, they correlate with a heart condition of the patient only when an abundance of them has extreme values.

Since the Cholesterol parameter does not hold one of the primary roles, we can also analyze the second dataset, i.e., the data without this attribute (Figs. 3 and 4). With this setup, we are able to augment the total number of data, therefore we can have a more precise insight on some features.

As can be seen, the first 3 attributes importance remained the same, but there are some differences in the remaining ones. The Fasting blood sugar plays a more important role (4 times more important than in the previous analysis). Nevertheless, the global feature importance remains unchanged, so we can consider the first analysis valid.

Finally, Fig. 5 plots the Shapley value for the specific cases of patients #3 and #5 (with only the former being diagnosed as suffering from a heart condition) to show how the parametric analysis can be repeated on an individual scale.

V. CONCLUSIONS AND FUTURE WORK

We used the game theoretic concept of the Shapley value to explain the results of a tree-based classifier for heart
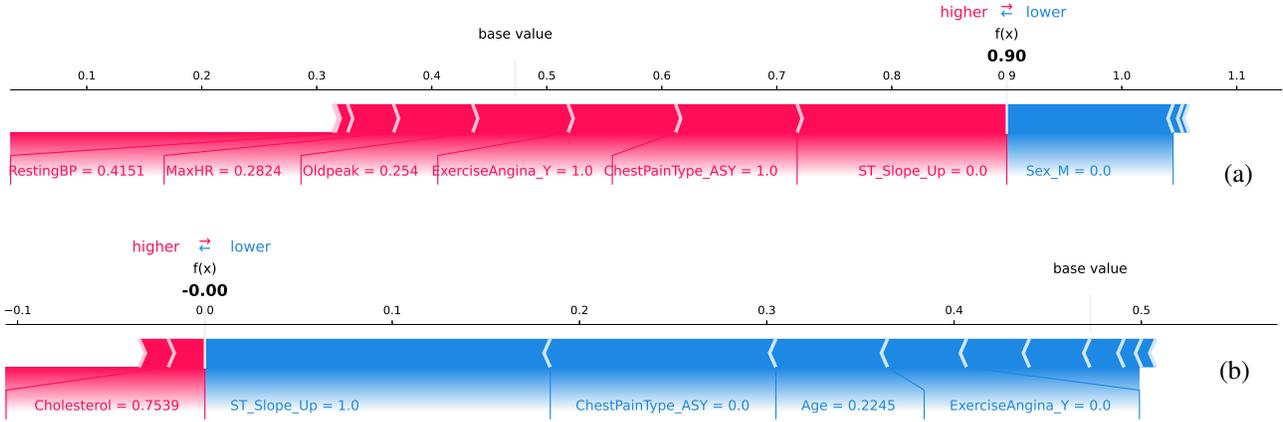
Figure 5. SHAP diagram for (a) patient #3 with output 1 and (b) patient #5 with output 0. Each plot shows the baseline value for the prediction and the final prediction ($f(x)$) for each specific patient. In red, it reports those normalized features (with their values for the specific patient) that contribute to increase the prediction from the base value to the final value (moving that value to the right, as indicated by the arrows). In blue, the opposite.
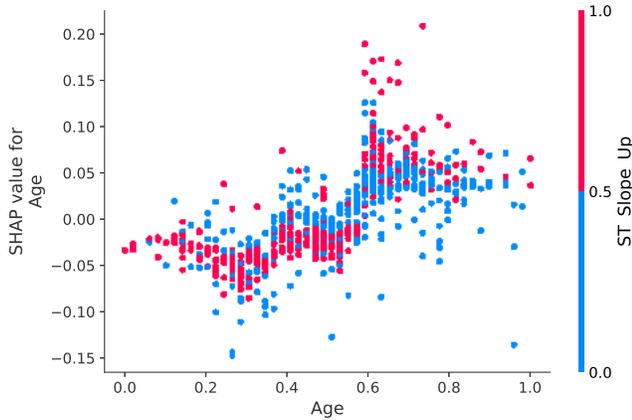

Figure 6. Dependence plot: value of the feature on the x-axis vs. the Shapley value of the same feature on the y-axis. Vertical dispersion of the data points represents interaction effects.

diseases. We found that some markers represent a high risk factor and contribute the most in the decision of the classifier. More than the presence of a single indicator, a heart condition can be diagnosed from a coalition of parameters, even if they have lightweight impact when taken individually.

It is confirmed that many parameters related to heart conditions correlate to one another, see for example maximum heart-rate and age. The latter is a particularly relevant driver, since the higher the age, the more likely the presence of a heart condition. This is highlighted in Fig. 6, where it is shown the strong correlation of age and the flat behavior of the ST-slope feature (i.e., well-established clinical marker for heart diseases [30]). However, a full-fledged causal model, which would be the best explanation, is hard to derive

and possibly not very meaningful in scenarios like the one under evaluation. On the contrary, the impact of unavoidable parameters such as an increasing age should be factored not as a risk element in itself, but properly accounted for when drawing conclusions. The Shapley value actually allows for a better transparency of the correlations picked up by predictive ML models.

To sum up, this can be a first step towards a better model building for understanding the relationships among physiological parameters and a more efficient diagnosing process (not limited to angina, but including also other heart phenomena, e.g., arrhythmias and extrasystoles). Possibly, the expansion of the dataset with more independent features can be instrumental to this task, to sort the parameters in order to avoid redundancies and/or confounding factors. Also, other game theoretic concepts (e.g., the inessential player axiom) could be used to complement the present analysis on relevant features and increase its general validity, as recently suggested by [31].

REFERENCES

[1] H. Zhu, C. K. Wu, C. H. Koo, Y. T. Tsang, Y. Liu, H. R. Chi, and K.-F. Tsang, "Smart healthcare in the era of internet-of-things," vol. 8, no. 5, pp. 26–30, 2019.

[2] G. Cisotto, A. V. Guglielmi, L. Badia, and A. Zanella, "Joint compression of EEG and EMG signals for wireless biometrics," in *Proc. IEEE Globecom*, 2018, pp. 1–6.

[3] K. Shailaja, B. Seetharamulu, and M. Jabbar, "Machine learning in healthcare: A review," in *Proc. IEEE ICECA*, 2018, pp. 910–914.

[4] S. Silvoni, M. Cavinato, C. Volpato, G. Cisotto, C. Genna, M. Agostini, A. Turolla, A. Ramos-Murguialday, and F. Piccione, "Kinematic and neurophysiological consequences of an assisted-force-feedback brain-machine interface training: a case study," *Frontiers in neurology*, vol. 4, p. 173, 2013.

[5] M. T. Ribeiro, S. Singh, and C. Guestrin, ""Why should i trust you?" explaining the predictions of any classifier," in *Proc. ACM SIGKDD*, 2016, pp. 1135–1144.

[6] A. M. Alhassan and W. M. N. W. Zainon, "Review of feature selection, dimensionality reduction and classification for chronic disease diagnosis," *IEEE Access*, vol. 9, pp. 87 310–87 317, 2021.

[7] A. Ghorbani and J. Zou, "Data Shapley: Equitable valuation of data for machine learning," in *Proc. Int. Conf. Mach. Learning*, 2019, pp. 2242–2251.

[8] L. S. Shapley, "A value for n-person games," *Contrib. Th. Games II, Ann. Math. Stud.*, vol. 28, 1953.

[9] L. Gianfagna and A. Di Cecco, *Explainable AI with Python*. Springer, 2021.

[10] A. T. Azar, H. I. Elshazly, A. E. Hassanien, and A. M. Elkorany, "A random forest classifier for lymph diseases," *Comp. Meth. Prog. Biomed.*, vol. 113, no. 2, pp. 465–473, 2014.

[11] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in *Adv. Neur. Inf. Proc. Sys.*, I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, Eds. Curran Associates, Inc., 2017, vol. 30, pp. 4765–4774.

[12] E. Štrumbelj and I. Kononenko, "Explaining prediction models and individual predictions with feature contributions," *Knowledge and information systems*, vol. 41, no. 3, pp. 647–665, 2014.

[13] S. M. Lundberg, G. Erion, H. Chen, A. DeGrave, J. M. Prutkin, B. Nair, R. Katz, J. Himmelfarb, N. Bansal, and S.-I. Lee, "From local explanations to global understanding with explainable AI for trees," *Nature Machine Intelligence*, vol. 2, no. 1, pp. 2522–5839, 2020.

[14] L. Badia, M. Levorato, F. Librino, and M. Zorzi, "Cooperation techniques for wireless systems from a networking perspective," vol. 17, no. 2, pp. 89–96, 2010.

[15] F. Guidolin, L. Badia, and M. Zorzi, "A distributed clustering algorithm for coordinated multipoint in LTE networks," vol. 3, no. 5, pp. 517–520, 2014.

[16] S. Cohen, E. Ruppin, and G. Dror, "Feature selection based on the Shapley value," in *Proc. Int. J. Conf. Artif. Intell.*, 2005, pp. 665–670.

[17] L. Prospero, R. Costa, and L. Badia, "Resource sharing in the Internet of things and selfish behaviors of the agents," *IEEE Trans. Circuits Syst. II*, vol. 68, no. 12, pp. 3488–3492, 2021.

[18] I. Covert and S.-I. Lee, "Improving KernelSHAP: Practical Shapley value estimation using linear regression," in *Proc. Int. Conf. Artif. Intell. Statist.*, 2021, pp. 3457–3465.

[19] D. Dave, H. Naik, S. Singhal, and P. Patel, "Explainable AI meets healthcare: A study on heart disease dataset," *arXiv preprint arXiv:2011.03195*, 2020.

[20] M. Sundararajan and A. Najmi, "The many Shapley values for model explanation," in *Proc. Int. Conf. Mach. Learning*, 2020, pp. 9269–9278.

[21] N. Jothi, W. Husain, and N. Abdul Rashid, "Predicting generalized anxiety disorder among women using Shapley value," *J. Infect. Pub. Health*, vol. 14, no. 1, pp. 103–108, 2021.

[22] H. Alsuradi, W. Park, and M. Eid, "Explainable classification of EEG data for an active touch task using Shapley values," in *Int. Conf. Human-Comp. Inter.*, 2020, pp. 406–416.

[23] G. Cisotto, A. V. Guglielmi, L. Badia, and A. Zanella, "Classification of grasping tasks based on EEG-EMG coherence," in *Proc. IEEE Healthcom*, 2018, pp. 1–6.

[24] G. Cisotto, S. Pupolin, S. Silvoni, M. Cavinato, M. Agostini, and F. Piccione, "Brain-computer interface in chronic stroke: An application of sensorimotor closed-loop and contingent force feedback," in *Proc. IEEE ICC*, 2013, pp. 4379–4383.

[25] K. Schultebraucks, M. Sijbrandij, I. Galatzer-Levy, J. Mouthaan, M. Olff, and M. van Zuiden, "Forecasting individual risk for long-term posttraumatic stress disorder in emergency medical settings using biomedical data: A machine learning multicenter cohort study," *Neurobiol. Stress*, vol. 14, p. 100297, 2021.

[26] P. V. Johnsen, S. Riemer-Sørensen, A. T. DeWan, M. E. Cahill, and M. Langaas, "A new method for exploring gene–gene and gene–environment interactions in GWAS with tree ensemble methods and SHAP values," *BMC bioinformatics*, vol. 22, no. 1, pp. 1–29, 2021.

[27] F. Soriano, "Kaggle heart failure prediction dataset." [Online]. Available: https://www.kaggle.com/fedesoriano/ heart-failure-prediction

[28] E. H. Estes, Z.-M. Zhang, Y. Li, L. G. Tereschenko, and E. Z. Soliman, "The Romhilt-Estes left ventricular hypertrophy score and its components predict all-cause mortality in the general population," *Am. Heart J.*, vol. 170, no. 1, pp. 104–109, 2015.

[29] K. Wang, R. W. Asinger, and H. J. Marriott, "ST-segment elevation in conditions other than acute myocardial infarction," *New Eng. J. Medic.*, vol. 349, no. 22, pp. 2128–2135, 2003.

[30] A. H. Kashou, H. Basit, and A. Malik, "ST segment," in *StatPearls [Internet]*. StatPearls Publishing, 2021.

[31] D. Fryer, I. Strümke, and H. Nguyen, "Shapley values for feature selection: The good, the bad, and the axioms," *IEEE Access*, vol. 9, pp. 144 352–144 360, 2021.