# Explaining Predictions of Hypertension Disease through Anchors

Gabriella Casalino[1,*], Giovanna Castellano[1], Katarzyna Kaczmarek-Majer[2], Pietro Giovanni Rizzo[1] and Gianluca Zaza[1,*]

[1]Computer Science Department, University of Bari Aldo Moro Bari, Italy
[2]Systems Research Institute, Polish Academy of Sciences, Warsaw, Poland

## Abstract

Hypertension is a disease that stresses the arteries and can cause damage to vital organs. It is often asymptomatic, and timely diagnosis and management are crucial to prevent complications and mitigate the risks associated with the disease. Photoplethysmography has proven to be effective in capturing variations in blood volume within vessels and holds the potential for continuous monitoring of heart-related diseases to be adopted in real-time systems [1]. Using automated processing on "high-risk" medical data requires careful attention to regulations. The emergence of Explainable Artificial Intelligence (XAI) is especially important in this context because it can provide explanations that clarify the reasoning behind the results produced by automatic processing. This paper introduces the application of an agnostic algorithm called Anchors for explaining predictions related to hypertension levels through the use of concatenations of logic statements. This algorithm has been selected based on its ability to produce easily understandable explanations, which is particularly valuable in the medical domain, where the primary stakeholders are physicians and patients. Additionally, it has been chosen for its ability to balance classification and explanation accuracy. Furthermore, we have investigated the impact of varying the number of features utilized in the explanations on the quantitative measures. This exploration involved the application of diverse feature selection methods, and their outcomes were systematically compared. Experiments showed that reducing the number of features does not harm classification performance and significantly improves the quality of explanations.

## Keywords

Explainable Artificial Intelligence, XAI, Hypertension, classification, Decision Support System, Photoplethysmography, Feature selection,



## 1. Introduction

Explainable Artificial Intelligence (XAI) has gained a lot of attention in recent years due to AI's incredible and sometimes overwhelming capabilities. The increasing power of AI has made it necessary to establish regulations to ensure trustworthy, privacy-compliant, and ethical AI practices. XAI specifically refers to automated methods that can represent, in a way that

*EXPLIMED - First Workshop on Explainable Artificial Intelligence for the medical domain - 19-20 October 2024, Santiago de Compostela, Spain*
*Corresponding author.
gabriella.casalino@uniba.it (G. Casalino); giovanna.castellano@uniba.it (G. Castellano); k.kaczmarek@ibspan.waw.pl (K. Kaczmarek-Majer); p.rizzo7@studenti.uniba.it (P. G. Rizzo); gianluca.zaza@uniba.it (G. Zaza)
000-0003-0713-2260 (G. Casalino); 0000-0002-6489-8628 (G. Castellano); 0000-0003-0422-9366 (K. Kaczmarek-Majer); 0000-0003-3272-9739 (G. Zaza)

is understandable for humans, the hidden mechanisms guiding their processing [2]. The importance of XAI extends across various domains, with a particular emphasis on areas like healthcare. In such critical fields, understanding algorithms' inner workings has become essential [3]. Physicians and patients alike need insight into how specific results are generated by an algorithm. This transparency is crucial for establishing trust in the technology and ensuring that AI applications are accurate and understandable to end-users. This need for explainability has become an absolute requirement in the medical domain, highlighting XAI's pivotal role in fostering trust and confidence in AI-driven decision-making processes [4]. Explainable methods are broadly categorized into two groups: ante-hoc methods, which are inherently explainable by design, and post-hoc methods, which are applied to the outcomes of a machine learning method to extract explanations.

Depending on the type of data and methods employed, various XAI methods have been proposed in the literature, and they have been effectively used in the medical domain [5]. Some examples are: feature importance techniques such as SHAP (SHapley Additive exPlanations) [6] and LIME (Local Interpretable Model-agnostic Explanations) [7], Counterfactual Explanations [8], Layer-wise Relevance Propagation (LRP) [9], Rule-based Models [10, 11, 12, 13], Attention Mechanisms [14], Surrogate Models [15, 16, 17].

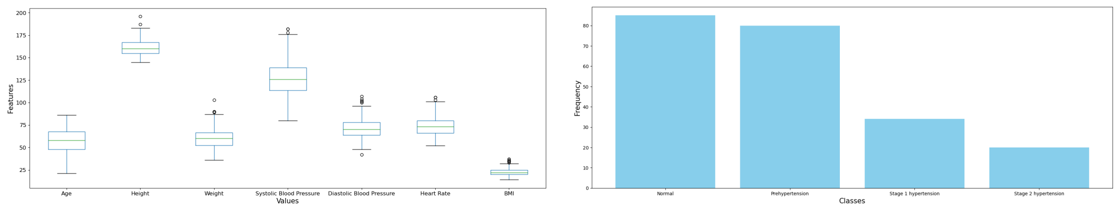
In this work, we used the Anchors algorithm. It is a model-agnostic technique used for generating explanations that are both easy to understand and reliable. It focuses on creating simple and clear conditions (anchors) that explain a model's prediction for a specific instance (local explanations). By analyzing each instance, the algorithm identifies a set of features that, when present, are highly likely to lead to the model's prediction. These anchor conditions are combined using a disjunction (OR combination) to form a complete explanation. This makes it easy for end-users to understand and accept the reasons behind a model's decision. Moreover, Anchors attempts to balance precision and recall, ensuring that the generated conditions are accurate and cover a significant portion of the decision space [18].

The quality and correctness of explanations are closely related to the number of features used to describe the data. To investigate how the number of features influences the accuracy of explainability, we compared four different feature selection methods. The objective of this comparison was twofold: to identify the most effective algorithm and to determine which subset of features were the most relevant for the predictive task.

In this paper, a case study on predicting hypertension is used to demonstrate deriving explainable predictions for enhancing decision support systems.

Hypertension is a heart condition with increased blood pressure, increasing the likelihood of cerebral, cardiac, and renal events. Doctors often prescribe antihypertensive drugs to lower blood pressure and reduce the risk of cardiovascular problems. However, many patients still have uncontrolled hypertension and related risk factors. To prevent major cardiovascular events, monitoring blood pressure continuously is crucial [19]. According to the World Health Organization (WHO), cardiovascular diseases (CVDs) are one of the leading causes of death[1]. Hypertension programs have proven effective in reducing the incidence of coronary heart disease and stroke, especially at the primary care level. However, these programs can be expensive,

[1]WHO:https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds) (last accessed on May 5, 2024




(a) (b)

**Figure 1:** Dataset information: (a) the statistics of the feature values and (b) the occurrence values of the target class.

requiring medical staff and resources to manage. To address these challenges, machine learning methods have emerged as useful tools to support medical decision-making [20], particularly in hypertension diagnosis [21].

In this scenario, photoplethysmography (PPG) emerges as a valuable tool for the continuous monitoring of vital sign parameters [22, 23]. Specifically, it finds widespread application in heart rate monitoring by utilizing light reflection due to blood variations in vessels [24]. The present study utilizes a dataset of patient information and vital signs obtained from photoplethysmographic signals related to hypertension [25]. There are various methods proposed in scientific works for explaining hypertension, such as in [26, 27, 28, 29], just to mention a few. However, none of them concentrate on the accuracy of the explanation. We can provide local explanations for previously unseen samples using the Anchors algorithm. At the same time, we can study the balance between coverage and precision of the explanations derived.

The main findings of this work are as follows:

- a study decision support systems for hypertension that takes into account both the classification and explanation accuracy;
- a study on the more effective subset of features useful for accurate predictions
- a study on the more effective subset of features useful for accurate and easy-to-understand explanations.

The paper is organized as follows. Section 2 describes the data, the Anchors algorithm, and adopted feature selection algorithms. Section 3 presents quantitative and qualitative results to evaluate accuracy in terms of classification performance and explainability. In Section 4, we summarize our findings, draw conclusions, and outline future work.

## 2. Materials and methods

The work aims to apply an explainable model, such as Anchors, to analyze tabular data on hypertension risk. The Anchors method, along with the hypertension dataset to be analyzed, is described in detail. Additionally, a brief overview of the feature selection methods is provided.

## 2.1. Data

We utilized a dataset containing the values of patients' photoplethysmographic (PPG) signals correlated with their respective physiological information. The study in [25] aimed to find a possible correlation between the two sets of information collected. The dataset included 219 subjects (115 female and 104 male) aged between 21 and 86, with an average age of 58. In our research, we considered a subset of 8 input features, namely *Sex*, *Age*, *Height*, *Weight*, *Systolic Blood Pressure* (SBP), *Diastolic Blood Pressure* (DBP), *Heart Rate* (HR), and *Body Mass Index* (BMI). Our selection was guided by identifying the most influential features in classifying hypertension. Consequently, we excluded the features $Num$ and $Subject\_ID$ as they did not contribute significantly to this classification. Figure 1(a) displays the box plots describing the input features. The plots show uniformly distributed feature values in the dataset for even representation in model processing. Finally, we can also observe that there are outlier points, as they fall outside the range defined by the box plot's whiskers. We found this problem consistent in *Weight*, *Diastolic Blood Pressure*, and *BMI*. The dataset contains four target classes, the healthy class *Normal*, and three classes representing the various disease states of hypertension, namely, *Prehypertension*, *Stage 1*, and *Stage 2*. As depicted in Figure 1(b), the dataset reveals a slight imbalance between the *Normal* and *Prehypertension* and the classes *Stage 1 hypertension* and *Stage 2 hypertension* classes, underscoring the challenge in the research and the need for a robust model.

## 2.2. Explainable algorithm

The Anchors algorithm is designed to provide explanations for the predictions made by any black-box classification model. This is done by identifying a decision rule that effectively describes the prediction process. Anchors [18] uses a perturbation-based strategy for predictions made by black-box machine learning models. This produces easily understandable IF-THEN rules, known as anchors, that precisely define the instances to which they apply, even for those that may not have been previously observed. A rule anchors a prediction when changes in feature values have no effect on the prediction itself.

For each instance being considered, perturbations are created and evaluated, allowing the approach to bypass the structural and internal parameters of the black-box model. As a result, Anchors are model-agnostic, enabling their application across diverse classes of models. Anchors uses reinforcement learning techniques alongside a graph search algorithm to reduce the computational costs and avoid local optima.

An anchor is formally defined as:

$$\mathbb{E}_{\mathcal{D}_x(z|A)}[1_{\hat{f}(x)=\hat{f}(z)}] \geq \tau, \quad A(x) = 1 \tag{1}$$

where $x$ represents the instance being explained; $A$ is a set of features, namely the resulting rule; $f$ indicates the classification model to be explained; $\mathcal{D}_x(z|A)$ indicates the distribution of neighbors of $x$, corresponding to $A$; $0 \leq t \leq 1$ specifies a precision threshold (only rules that achieve a local fidelity of at least $t$ are considered a valid result).

In [18], the *coverage* is introduced to determine the quality of rules. Coverage refers to identifying a set of rules that apply to a significant portion of a model's input space. This means

**Table 1**
Feature selection settings.

| Algorithm | Parameter | #Features | Acronym |
|---|---|---|---|
| Feature Importance | 62.5% | 5 | FI1 |
| | 50% | 4 | FI2 |
| | 37.5% | 3 | FI3 |
| | 25% | 2 | FI4 |
| Recursive Feature Elimination | 65.5% | 5 | RFE1 |
| | 50% | 4 | RFE2 |
| | 37.5% | 3 | RFE3 |
| | 25.5% | 2 | RFE4 |
| Information Gain | 62.5% | 5 | IG1 |
| | 50% | 4 | IG2 |
| | 37.5% | 3 | IG3 |
| | 25% | 2 | IG4 |
| Correlation based | 0.7 | 6 | CB1 |
| | 0.6 | 5 | CB2 |
| | 0.5 | 4 | CB3 |
| | 0.3 | 3 | CB4 |

that it calculates the probability of an anchor applying to its neighbors, which represents its perturbation space. The goal is to find a rule that has the highest coverage among all eligible rules that meets the precision threshold according to the probabilistic definition.

$$\operatorname{cov}(A) = \mathbb{E}_{\mathcal{D}(z)}[A(z)]. \tag{2}$$

Rules with more predicates are typically more precise than those with fewer predicates. On the other hand, a rule with many features is excessively specific and only applicable to a few instances, leading to low coverage values. Therefore, finding the right balance between precision and coverage is essential to identify the most significant rules that describe a larger portion of the model.

### 2.3. Feature selection methods

To strike a balance between precision and coverage, reducing the number of features is necessary. To achieve this, we employed four different feature selection algorithms, namely Feature Importance, Recursive Feature Elimination, Information Gain, and Correlation-Based. Each algorithm was tested with four different parameter settings, resulting in variations in the number of features considered, ranging from 2 to 6. Table 1 summarises the sixteen different settings.

## 3. Results

A set of experiments was carried out to achieve two objectives: firstly, to evaluate the effectiveness of the Anchors algorithm in explaining hypertension data while altering the number of features employed to represent the data, and secondly, to investigate the influence of feature reduction on classification performance. The primary goal is to identify the best feature selection configuration that leads to favorable results in both explainability and classification performance. A value $0.95$ was used for the Anchors' threshold $t$, resulting in highly precise rules. Empirical evaluation of this algorithm parameter has shown that this default value was optimal. The results were evaluated both quantitatively and qualitatively. The dataset was split into 33% for testing and the rest for training. A random forest classifier was implemented using the Scikit-learn library[2] with default parameters. We selected this classifier based on its performance demonstrated in a previous study [30]. In that work, it outperformed other classification algorithms, including the perceptron, support vector machine, and neuro-fuzzy systems. Additionally, it exhibited stability when subjected to variations in data splits and feature numbers. Indeed, in that study, classifiers were compared using only two features. This was done to simplify the set of explanations returned from transparent models, such as fuzzy neural networks, making it easier for physicians to interpret. In this work, we take a step forward by utilizing a post-hoc explanation method that leverages natural language to explain the decision-making process that led to a given result with a black-box algorithm. However, even in this case, a high number of features compromises the clarity and effectiveness of the explanations. Therefore, we have reduced the number of features.

### 3.1. Quantitative results

Quantitative evaluation of the classification performance was carried out using standard classification metrics, such as accuracy, precision, recall, and F1 score, on different subsets of data. The outcomes of this evaluation, along with the number of features obtained from four feature selection methods, each having four different parameter settings, are presented in Table 2. Sixteen distinct subsets of data were generated through various configurations of feature selections, resulting in a range of features from 2 to 6. Additionally, we examined the scenario involving all features to assess whether reducing the number of features affects accuracy negatively or, conversely, leads to performance enhancement by reducing noise in the data.

Qualitative results confirm the robustness of random forest to the reduction of the number of features. The results remain consistent across different feature selection settings. In fact, reducing the number of features leads to an improvement in the classification performance. This indicates that some features contribute to noise and are not required for classification. A detailed analysis of the subsets of features will be conducted in the following paragraph.

As previously discussed, Anchors provides coverage and precision for each explanation, enabling us to quantify its performance. Table 3 presents the average values of coverage and precision across samples for each feature selection setting, along with the number of returned features. We can observe a high value of precision, confirming the previous discussion. Moreover, for all the feature selection methods, we observe an increase in coverage as the number of adopted features is reduced, while still preserving classification performance. This analysis suggests that a lower number of features is preferable because it improves coverage values, and shorter explanations are easier to understand than longer ones. In particular, regarding the algorithm that returned the best performance in terms of explainability, all the algorithms with two features gave a coverage of 27%, which is the best. Thus, the quantitative analysis of the explanations is not able to identify the best setting of feature selection methods, but suggests that a lower number of features is better.

[2]Python's Scikit-Learn library: https://scikit-learn.org/

**Table 2**
Quantitative results of the classifier, for different subsets of data, varying the number of the selected features.

| | #F | Acc. | Prec | | | | Rec | | | | F1 | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | N | P | S1 | S2 | N | P | S1 | S2 | N | P | S1 | S2 |
| No FS | 8 | 93% | 90% | 92% | 100% | 100% | 100% | 88% | 93% | 86% | 95% | 90% | 96% | 92% |
| FI1 | 5 | 99% | 100% | 96% | 100% | 100% | 100% | 100% | 100% | 86% | 100% | 98% | 100% | 92% |
| FI2 | 4 | 100% | 100% | 100% | 100% | 100% | 100% | 100% | 100% | 100% | 100% | 100% | 100% | 100% |
| FI3 | 3 | 97% | 100% | 93% | 100% | 100% | 100% | 100% | 93% | 86% | 100% | 96% | 96% | 92% |
| FI4 | 2 | 100% | 100% | 100% | 100% | 100% | 100% | 100% | 100% | 100% | 100% | 100% | 100% | 100% |
| RFE1 | 5 | 97% | 96% | 96% | 100% | 100% | 100% | 96% | 100% | 86% | 98% | 96% | 100% | 92% |
| RFE2 | 4 | 99% | 100% | 96% | 100% | 100% | 100% | 100% | 100% | 86% | 100% | 98% | 100% | 92% |
| RFE3 | 3 | 95% | 96% | 89% | 100% | 100% | 100% | 96% | 86% | 86% | 98% | 93% | 92% | 92% |
| RFE4 | 2 | 100% | 100% | 100% | 100% | 100% | 100% | 100% | 100% | 100% | 100% | 100% | 100% | 100% |
| IG1 | 5 | 99% | 100% | 96% | 100% | 100% | 100% | 100% | 100% | 86% | 100% | 98% | 100% | 92% |
| IG2 | 4 | 100% | 100% | 100% | 100% | 100% | 100% | 100% | 100% | 100% | 100% | 100% | 100% | 100% |
| IG3 | 3 | 97% | 96% | 96% | 100% | 100% | 100% | 96% | 100% | 86% | 98% | 96% | 100% | 92% |
| IG4 | 2 | 100% | 100% | 100% | 100% | 100% | 100% | 100% | 100% | 100% | 100% | 100% | 100% | 100% |
| CB1 | 6 | 93% | 100% | 90% | 85% | 100% | 100% | 100% | 79% | 71% | 100% | 95% | 81% | 83% |
| CB2 | 5 | 99% | 96% | 100% | 100% | 100% | 100% | 96% | 100% | 100% | 98% | 98% | 100% | 100% |
| CB3 | 4 | 99% | 96% | 100% | 100% | 100% | 100% | 96% | 100% | 100% | 98% | 98% | 100% | 100% |
| CB4 | 3 | 41% | 58% | 36% | 0% | 0% | 54% | 62% | 0% | 0% | 56% | 45% | 0% | 0% |

**Table 3**
Average values of precision and coverage for different subsets of data, varying the number of the selected features.

| | #Feature | AvgPrec | AvgCov |
|---|---|---|---|
| No FS | 8 | 89% | 13% |
| FI1 | 5 | 87% | 16% |
| FI2 | 4 | 85% | 17% |
| FI3 | 3 | 85% | 18% |
| FI4 | 2 | 84% | 27% |
| RFE1 | 5 | 88% | 16% |
| RFE2 | 4 | 85% | 16% |
| RFE3 | 3 | 83% | 20% |
| RFE4 | 2 | 84% | 27% |
| IG1 | 5 | 87% | 16% |
| IG2 | 4 | 83% | 18% |
| IG3 | 3 | 86% | 19% |
| IG4 | 2 | 83% | 27% |
| CB1 | 6 | 88% | 13% |
| CB2 | 5 | 86% | 13% |
| CB3 | 4 | 81% | 15% |
| CB4 | 3 | 72% | 7% |

## 3.2. Qualitative results

The qualitative evaluation aims to better understand the influence of different feature selection settings on the explanations. We reported the explanations obtained with Anchors without using feature selection and with the FI2 feature selection setting, along with the features selected by the different settings.

Figure 2 illustrates the features selected from each setting. We can observe that, except for the correlation-based algorithm, which behaves completely differently, the other algorithms

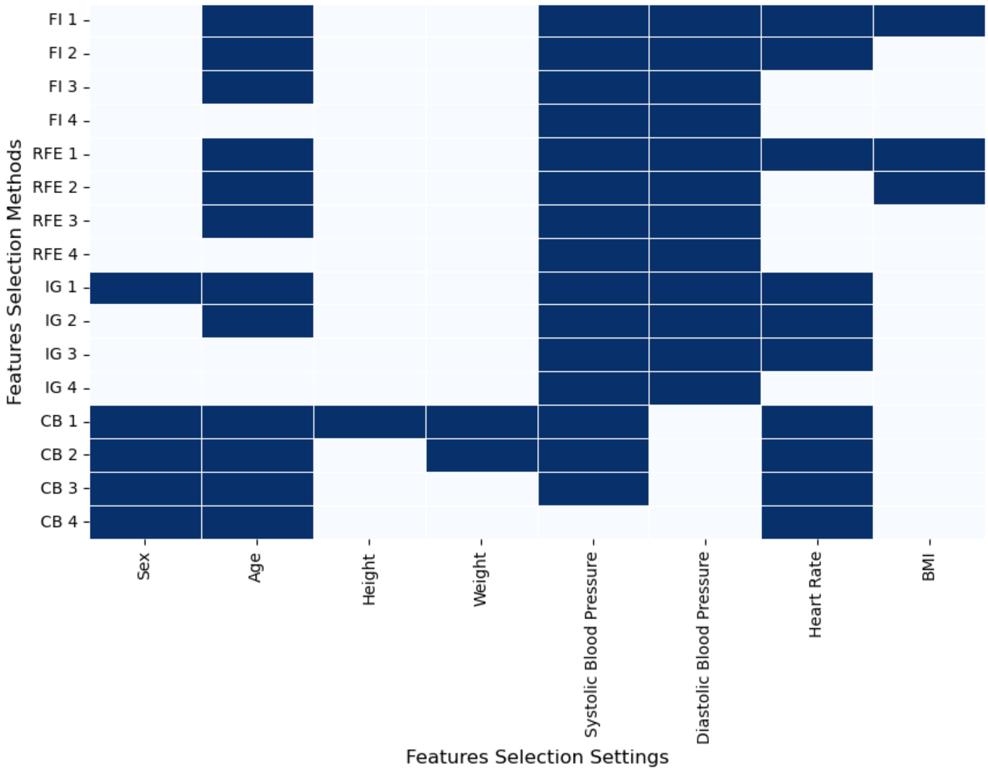



**Figure 2:** Comparison of the features selected by the different feature selection settings.

agree that the two most important features are Systolic and Diastolic blood pressure. When more features are added, they do not completely agree about the most important ones, but overall, age seems to be an important feature, as well as heart rate. Almost all the settings agree that height and weight are useless, as well as the BMI for most of them. The correlation-based algorithm returns completely different results; indeed, it is the only algorithm selecting sex and age as relevant other than the systolic blood pressure and the heart rate. Surprisingly, it does not select the diastolic blood pressure in any setting, resulting still in a good performance as previously discussed. These differences need further studies, so we will focus on the first three algorithms for the analysis of the explanations.

Figure 3 illustrates anchors generated for four instances belonging to the four classes: normal, Prehypertension, Stage 1 hypertension, and Stage 2 hypertension. Specifically, Figure 3(a) displays the explanations obtained without feature selection, while Figure 3(b) showcases the explanations after applying the feature selection setting FI2. When feature selection is applied, we observe an increase in coverage for all classes while the precision remains comparable or even increases. It was found that for the *Normal* class, only one feature, Systolic blood pressure, was enough to describe the class. However, very complex explanations were generated for the disease classes without feature selection. But, when feature selection was applied, the explanations became clear and easy-to-understand, with a reduced number of anchors. The

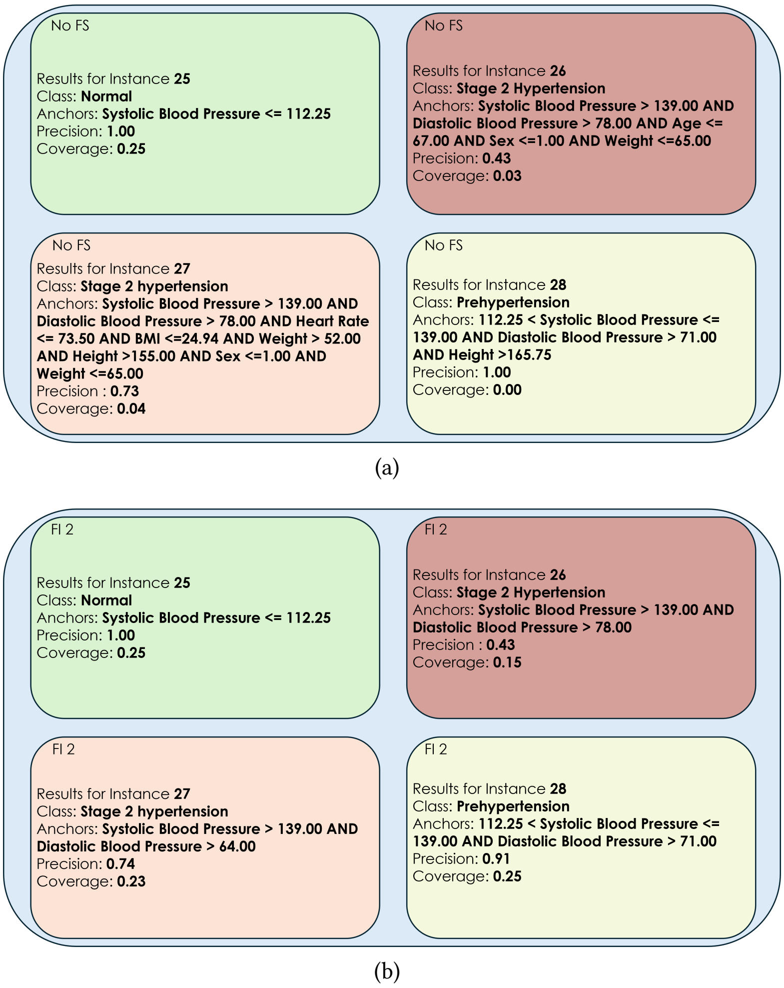



**Figure 3:** Examples of explanations obtained from the Anchors algorithm, for the four classes Normal, Stage 2 Hypertension, Stage 1 Hypertension, Prehypertension, without feature selection (a), and with the feature importance method FI 2 (b).

algorithm also identified differences between samples belonging to the three classes, where even if the two features involved were the same, the values of these features varied across classes. Similar results have been observed with the other feature selection settings.

## 4. Conclusion

This study aimed to assess the effectiveness of the Anchors algorithm in explaining hypertension data. To do this, we used a dataset that included patient personal information and vital signs obtained through photoplethysmography.

Previously, we used explainable algorithms to generate IF-THEN rules for classification explanations, which yielded lower results than black-box models. For this study, we used a post-hoc method that derives IF-THEN rules to explain a black-box model's decisions.

Our results showed that the Anchors algorithm was sensitive to the number of features in the data. An increased number of features led to decreased rule reliability (coverage). We compared sixteen different feature selection settings to understand this impact on classification performance and explainability. The results indicated that our chosen classification algorithm (random forest) remained robust even with reduced features.

Interestingly, only two out of eight features in the original data space yielded the best coverage values. This suggests a preference for concise explanations both quantitatively and qualitatively. Qualitative analysis also highlighted the agreement among feature selection algorithms, except for the correlation-based algorithm, regarding the importance of photoplethysmographic signal-derived features, specifically Systolic Blood Pressure and Diastolic Blood Pressure. Moreover, the anchors generated for the hypertension data with feature selection are more compact and, thus, more understandable than those generated without feature selection. Additionally, the algorithm was able to correctly identify differences among the four classes and explain them in terms of conjunctions of anchors.

Overall, Anchors proved to be a viable solution for explaining black-box models in natural language, facilitating comprehension for humans. However, its effectiveness relies on a limited number of features. Therefore, the adoption of feature importance methods becomes essential when utilizing Anchors.

Future research will delve into exploring the impact of various feature selection settings on Anchors using different datasets. This investigation aims to determine whether any feature selection algorithm outperforms others or if an optimal approach exists for maximizing Anchors' effectiveness.

## 5. Acknowledgments

Giovanna Castellano and Gianluca Zaza acknowledge the support of the PNRR project FAIR - Future AI Research (PE00000013), Spoke 6 - Symbiotic AI (CUP H97G22000210007) under the NRRP MUR program funded by the NextGenerationEU. The research objectives of this paper are in partial fulfilment of the project EXPLICIT (CUP H93C23000890005). Gabriella Casalino acknowledges funding from the European Union PON project Ricerca e Innovazione 2014-2020, DM 1062/2021. G. Casalino and G. Castellano are with the CITEL - Centro Interdipartimentale di Telemedicina, University of Bari Aldo Moro. All authors are members of the INdAM GNCS research group. This paper has been partially supported by the "INdAM – GNCS Project", CUP_E53C22001930001.

## References

[1] G. Coviello, A. Florio, G. Avitabile, C. Talarico, J. M. Wang-Roveda, Distributed full synchronized system for global health monitoring based on flsa, IEEE Transactions on Biomedical Circuits and Systems 16 (2022) 600–608.

[2] N. Díaz-Rodríguez, J. Del Ser, M. Coeckelbergh, M. L. de Prado, E. Herrera-Viedma, F. Herrera, Connecting the dots in trustworthy artificial intelligence: From ai principles, ethics,

and key requirements to responsible ai systems and regulation, Information Fusion (2023) 101896.

[3] L. Aversano, M. L. Bernardi, M. Cimitile, D. Montano, R. Pecori, L. Veltri, Explainable anomaly detection of synthetic medical iot traffic using machine learning, SN Computer Science 5 (2024) 1–15.

[4] K. Kaczmarek-Majer, G. Casalino, G. Castellano, M. Dominiak, O. Hryniewicz, O. Kamińska, G. Vessio, N. Díaz-Rodríguez, Plenary: Explaining black-box models in natural language through fuzzy linguistic summaries, Information Sciences 614 (2022) 374–399.

[5] H. Hagras, Towards true explainable artificial intelligence for real world applications, 2023, p. 5 – 13.

[6] J. Budzianowski, K. Kaczmarek-Majer, J. Rzeźniczak, M. Słomczyński, F. Wichrowski, D. Hiczkiewicz, B. Musielak, Ł. Grydz, J. Hiczkiewicz, P. Burchardt, Machine learning model for predicting late recurrence of atrial fibrillation after catheter ablation, Scientific Reports 13 (2023) 15213.

[7] S. Bouazizi, H. Ltifi, Enhancing accuracy and interpretability in eeg-based medical decision making using an explainable ensemble learning framework application for stroke prediction, Decision Support Systems 178 (2024) 114126.

[8] J. M. Metsch, A. Saranti, A. Angerschmid, B. Pfeifer, V. Klemt, A. Holzinger, A.-C. Hauschild, Clarus: An interactive explainable ai platform for manual counterfactuals in graph neural networks, Journal of Biomedical Informatics 150 (2024) 104600.

[9] P. R. Bassi, S. S. Dertkigil, A. Cavalli, Improving deep neural network generalization and robustness to background bias via layer-wise relevance propagation optimization, Nature Communications 15 (2024) 291.

[10] G. Casalino, G. Castellano, U. Kaymak, G. Zaza, Balancing accuracy and interpretability through neuro-fuzzy models for cardiovascular risk assessment, in: 2021 IEEE Symposium Series on Computational Intelligence (SSCI), IEEE, 2021, pp. 1–8.

[11] P. V. de Campos Souza, E. Lughofer, Efnn-nulluni: An evolving fuzzy neural network based on null-uninorm, Fuzzy Sets and Systems 449 (2022) 1–31.

[12] D. Leite, A. Silva, G. Casalino, A. Sharma, D. Fortunato, A.-C. Ngomo, Egnn-c+: Interpretable evolving granular neural network and application in classification of weakly-supervised eeg data streams, in: 2024 IEEE Conference on Evolving and Adaptive Intelligent Systems (EAIS), 2024, pp. 1–8.

[13] M. Daole, A. Schiavo, J. L. C. Bárcena, P. Ducange, F. Marcelloni, A. Renda, Openfl-xai: Federated learning of explainable artificial intelligence models in python, SoftwareX 23 (2023) 101505.

[14] Y. Zhang, J. Chen, X. Ma, G. Wang, U. A. Bhatti, M. Huang, Interactive medical image annotation using improved attention u-net with compound geodesic distance, Expert Systems with Applications 237 (2024) 121282.

[15] I. Matei, W. Piotrowski, A. Perez, J. de Kleer, J. Tierno, W. Mungovan, V. Turnewitsch, System resilience through health monitoring and reconfiguration, ACM Transactions on Cyber-Physical Systems 8 (2024) 1–27.

[16] R. Avogadro, F. D'Adda, M. Cremaschi, Feature/vector entity retrieval and disambiguation techniques to create a supervised and unsupervised semantic table interpretation approach, Knowledge-Based Systems (2024) 112447.

[17] D. Schicchi, D. Taibi, Ai-driven inclusion: Exploring automatic text simplification and complexity evaluation for enhanced educational accessibility, in: International Conference on Higher Education Learning Methodologies and Technologies Online, Springer, 2023, pp. 359–371.

[18] M. T. Ribeiro, S. Singh, C. Guestrin, Anchors: High-precision model-agnostic explanations, in: Proceedings of the AAAI conference on artificial intelligence, volume 32, 2018.

[19] F. H. Messerli, B. Williams, E. Ritz, Essential hypertension, The Lancet 370 (2007) 591–603.

[20] G. Quer, R. Arnaout, M. Henne, R. Arnaout, Machine learning and the future of cardiovascular care: Jacc state-of-the-art review, Journal of the American College of Cardiology 77 (2021) 300–313.

[21] H. Koshimizu, R. Kojima, Y. Okuno, Future possibilities for artificial intelligence in the practical management of hypertension, Hypertension Research 43 (2020) 1327–1337.

[22] G. Casalino, G. Castellano, G. Zaza, On the use of fis inside a telehealth system for cardiovascular risk monitoring, in: 2021 29th Mediterranean Conference on Control and Automation (MED), IEEE, 2021, pp. 173–178.

[23] G. Casalino, G. Castellano, A. Nisio, V. Pasquadibisceglie, G. Zaza, A mobile app for contactless measurement of vital signs through remote photoplethysmography, in: 2022 IEEE international conference on systems, man, and cybernetics (SMC), IEEE, 2022, pp. 2675–2680.

[24] A. Gudi, M. Bittner, R. Lochmans, J. van Gemert, Efficient real-time camera based estimation of heart rate and its variability, in: Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops, 2019, pp. 0–0.

[25] Y. Liang, Z. Chen, G. Liu, M. Elgendi, A new, short-recorded photoplethysmogram dataset for blood pressure monitoring in china, Scientific Data 5 (2018).

[26] D. Yun, H.-L. Yang, S. G. Kim, K. Kim, D. K. Kim, K.-H. Oh, K. W. Joo, Y. S. Kim, S. S. Han, Real-time dual prediction of intradialytic hypotension and hypertension using an explainable deep learning model, Scientific Reports 13 (2023) 18054.

[27] E.-S. A. El-Dahshan, M. M. Bassiouni, S. K. Khare, R.-S. Tan, U. R. Acharya, Exhyptnet: An explainable diagnosis of hypertension using efficientnet with ppg signals, Expert Systems with Applications 239 (2024) 122388.

[28] G. Coviello, G. Avitabile, A. Florio, C. Talarico, J. M. Wang-Roveda, A novel low-power time synchronization algorithm based on a fractional approach for wireless body area networks, IEEE Access 9 (2021) 134916–134928. doi:10.1109/ACCESS.2021.3115440.

[29] R. Elshawi, M. H. Al-Mallah, S. Sakr, On the interpretability of machine learning-based model for predicting hypertension, BMC medical informatics and decision making 19 (2019) 1–32.

[30] G. Casalino, G. Castellano, G. Zaza, Using an adaptive neuro-fuzzy inference system for the classification of hypertension., in: WILF, 2021.