Using an adaptive neuro-fuzzy inference system for the classification of hypertension

Gabriella Casalino, Giovanna Castellano and Gianluca Zaza

Department of Computer Science, University of Bari Aldo Moro, Italy

Abstract

In this work, neuro-fuzzy systems are compared to standard machine learning algorithms to predict the hypertension risk level. Hypertension is a cardiovascular disease, which should be continuously monitored to avoid the worsening of its symptoms. Automatic techniques are useful to support the clinicians in this task, however, most of the machine learning techniques behave like black boxes, thus they are not able to explain how their results have been obtained. In the medical domain, this is a critical factor, and explainability is demanded. Neuro-fuzzy systems, that combine Neural Networks (NNs) and Fuzzy Inference Systems (FISs), are used to obtain explainable results. Moreover, to enhance the explanation, a feature selection method has been used to reduce the number of relevant features and thus the overall number of fuzzy rules. Qualitative analyses have shown comparable results between the machine learning methods and the neuro-fuzzy systems. However, the neuro-fuzzy systems are able to explain the hypertension risk level with only nine fuzzy rules, which are easy to interpret since they use linguistic terms.

Keywords

Neuro-Fuzzy model, Hypertension classification, Decision Support System, Machine learning algorithm

1. Introduction

Hypertension is cardiovascular disease, consisting of a rise in blood pressure, that increases the risk for cerebral, cardiac, and renal events. Antihypertensive drugs are used to lower blood pressure, thus reducing cardiovascular risk. However, despite the availability of several effective drugs, hypertension and its concomitant risk factors remain uncontrolled in most patients, whilst continuous monitoring would help in preventing major cardiovascular events [1]. The World Health Organization (WHO), mentions cardiovascular diseases (CVDs) among the first causes of death [1]. Hypertension programs have shown to be effective at the primary care level, to reduce coronary heart disease and stroke. However, these programs are expensive in terms of human costs, since they involve clinicians and other medical staff, and in terms of facilities that need to be managed.

As an alternative, machine learning methods, have been shown to be effective tools to support medical decisions [2], particularly for hypertension diagnostics [3]. Moreover, low-cost sensors,

[1] WHO:https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds) (last access August 4, 2021)

and fast network connections have led to a new discipline called the Internet of Medical Things (IoMT), where smart devices are continuously connected and they are used for several purposes such as monitoring the status of patients [4] or for diagnosing a disease [5] Machine learning techniques, together with smart sensors, are combined in intelligent systems which are used for m-health, telemedicine, ambient assisted living, etc. [6].

In this context, photoplethysmography (PPG) is a great ally for continuous monitoring of vital signs parameters [7], and particularly, it is widely used for hearth rate monitoring [8]. It uses the light reflectance due to blood variations in vessels, for measurements.

In this work, a dataset of photoplethysmographic signals was collected to perform a quality assessment study of them and explore the intrinsic relationship between PPG waveform and cardiovascular disease [9].

Specifically, interpretability for hypertension prediction is studied, since the results returned by automatic processing need to be understood by physicians [10]. Fuzzy logic has shown to be effective in the medical domain since it uses linguistic terms and represents expert knowledge and reasoning [11, 12]. Usually, when expert knowledge is available, fuzzy rules are defined by hand. However, when it is missing or partially available, neuro-fuzzy networks are able to automatically learn the parameters of the fuzzy rules from the data. Indeed, they form an adaptive fuzzy system exploiting the similarities between fuzzy systems and some types of neural networks [13].

A feed-forward network or a set of interpretable fuzzy rules are suitable to represent the reasoning behind a classification model learned from a neuro-fuzzy network. This leads to the use of neuro-fuzzy networks suitable for classification tasks where the interpretability of the model, as well as the accuracy, are desirable. A neuro-fuzzy system has been compared with standard machine learning techniques since it combines the accuracy of neural networks with the interpretability characteristic of fuzzy inference systems [14].

The paper is organized as follows. Section 2 describes the data and the algorithms that have been used to assess the hypertension stage. Section 3 reports the results of experiments aimed to compare the derived neuro-fuzzy model with other machine learning methods, in terms of classification performance and interpretability. In section 4 we draw conclusions and outline future works.

2. Materials and methods

The goal of the work is to compare black-box machine learning algorithms with neuro-fuzzy systems to verify whether the use of the latter approach is more effective than classical machine learning algorithms, in terms of accuracy and interpretability. Indeed, neuro-fuzzy systems generate IF-THEN rules that constitute a model that is comprehensible to the user.

2.1. Data

A dataset composed of 219 subjects, aged between 21 and 86 years (mean age 58), has been used. The dataset collects the photoplethysmographic signals (PPG) together with the related physiological signals of the patients, to study the presence of possible correlations between them

Table 1
Statistics on the dataset.

Features	Range
Age	21-86
Height	145-196
Weight	36-103
Systolic Blood Pressure (SBP)	80-182
Diastolic Blood Pressure (DBP)	42-107
Heart Rate (HR)	52-106
Body Mass Index (BMI)	15-37

Classes	Frequency
Normal	85
Prehypertension	80
Stage 1 hypertension	34
Stage 2 hypertension	20

[9]. To the aim of this work, only physiological signals have been considered, and particularly a subset of seven features has been selected, as summarised in Table 1 [2]

Moreover, while four diseases are described in the dataset, namely hypertension, diabetes, cerebral infarction, and cerebrovascular disease, this work focuses on hypertension disease. Four output classes *Normal*, *Prehypertension*, *Stage 1*, and *Stage 2* have been defined for the prediction task. As Table 1 shows, the dataset is quite unbalanced, indeed, patients belonging to the two last classes (i.e. serious disease symptoms) are lower than those belonging to the first two classes (i.e. healthy subjects, and patients with low symptoms).

2.2. Classification algorithms

To solve the decision task, classification algorithms have been used. In particular, two variants of neuro-fuzzy systems (with Gaussian and Triangular membership functions) have been compared with standard machine learning algorithms.

A neuro-fuzzy network, i.e. a neural network encoding a set of fuzzy IF-THEN rules in its structure, was trained to learn fuzzy rules for assessing the level of hypertension from data. In particular, the form of fuzzy rules adheres to a zero-order Takagi-Sugeno (TS) fuzzy model [15] in which the antecedent of each rule is represented by fuzzy sets while the consequent part is defined by fuzzy singletons.

Given the collection of rules, the fuzzy model provides certainty degrees for each output class (risk level) by inference of fuzzy rules. The fuzzy knowledge base will contain fuzzy rules with the following structure:
IF (x_1 is A_{k1}) AND ... AND (x_n is A_{kn}) THEN (y_1 is b_{k1}) AND .... AND (y_m is b_{km})
 for $k = 1, .., K$, where K is the number of rules, A_{ki} are fuzzy sets defined over the n input variables $x_i (i = 1, ..., n)$ and b_{kj} are fuzzy singletons expressing the certainty degree of the m output class $y_j, j = 1...m$. Gaussian and Triangular membership functions have been used to design the fuzzy sets in the two variants of the system.

The neuro-fuzzy architecture is inspired by ANFIS (Adaptive-Network-Based Fuzzy Inference System) [16] which consists of a four-layer feed-forward neural network that reflects the fuzzy rules in its architecture, as shown in Fig. 1. The network performs the inference of fuzzy rules by computing for each layer: 1) the membership degree of input values to fuzzy sets, 2) the

---

[2]Three features have been removed (Num and Subject_ID) since not useful for the classification task, and Sex since we are modeling continuous features only.
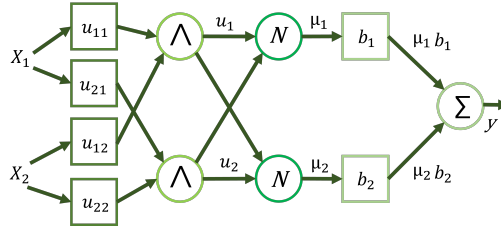
Figure 1: Architecture of the neuro-fuzzy network.

activation strength of each fuzzy rule, 3) the normalized activation strengths and 4) the certainty degree for output classes.

A Backpropagation learning procedure implementing the gradient descent on fuzzy rules parameters was used for the training of the neuro-fuzzy network.

Four standard classification algorithms have been used for comparison, namely Random Forest (RF), Multilayer Perceptron (MP), Multiclass support vector machine (SVC), XGBoost (XGB) [7]. Python's Scikit-Learn classification algorithms [3], with default parameters, have been used.

3. Results

Two sets of experiments have been conducted to compare the effectiveness of the neuro-fuzzy models, with the other classifiers, in terms of classification performance. Moreover, the interpretability of NFSs has been studied.

In the first one, all the features have been considered, while in the second one, a feature selection technique, based on ANOVA F-values [4], has been used.

This second experiment aimed to reduce the number of features, thus leading to more simple models, that is with a lower number of fuzzy rules. Of course, while increasing the interpretability of the neuro-fuzzy models, classification performance should be preserved or increased.

Since the dataset is unbalanced, to study the robustness of the different algorithms, in learning accurate models, three experimental setups have been considered, by using different splits for the training and test sets (60-40, 70-30, and 80-20).

Moreover, to evaluate which membership function is more suitable for the given problem, both Gaussian (NFG) and Triangular (NFT) membership functions have been compared.

Standard classification measures have been used to quantitatively evaluate the model performances, whilst both quantitative and qualitative evaluations have been discussed to evaluate the interpretability of the neuro-fuzzy systems.

Table 2 shows the qualitative evaluation of the neuro-fuzzy systems and the standard classifiers, without and with feature selection, varying the splits. Looking at the neuro-fuzzy models

---

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

[4]*f_classif*: https://scikit-learn.org/stable/modules/generated/sklearn.feature_selection.f_classif.html

Table 2

Qualitative evaluation of the classifiers, with and without feature selection, varying the split, in terms of Accuracy (A), Precision (P), Recall (R), and F1-measure (F1).

Split	Classifiers	No Feature selection				Feature selection			
		A	P	R	F1	A	P	R	F1
80-20	NFT	0.64	0.71	0.58	0.57	0.93	0.92	0.89	0.89
	NFG	0.80	0.83	0.78	0.79	0.89	0.89	0.84	0.85
	MLP	0.80	0.80	0.82	0.81	0.52	0.28	0.35	0.31
	RF	0.95	0.95	0.90	0.92	1.0	1.0	1.0	1.0
	SVC	0.39	0.10	0.25	0.14	0.39	0.10	0.25	0.14
	XGBC	**1.0**	**1.0**	**1.0**	**1.0**	**1.0**	**1.0**	**1.0**	**1.0**
70-30	NFT	0.68	0.56	0.56	0.55	0.92	0.88	0.89	0.88
	NFG	0.79	0.78	0.75	0.76	0.83	0.84	0.82	0.81
	MLP	0.79	0.75	0.74	0.73	0.47	0.26	0.31	0.28
	RF	0.95	0.90	0.89	0.90	0.98	0.98	0.96	0.97
	SVC	0.39	0.10	0.25	0.14	0.39	0.10	0.25	0.14
	XGBC	**1.0**	**1.0**	**1.0**	**1.0**	**1.0**	**1.0**	**1.0**	**1.0**
60-40	NFT	0.64	0.54	0.53	0.53	0.78	0.56	0.59	0.58
	NFG	0.80	0.80	0.75	0.77	0.84	0.85	0.75	0.78
	MLP	0.82	0.79	0.78	0.78	0.42	0.22	0.28	0.25
	RF	**0.97**	0.96	0.92	0.93	**0.99**	**0.98**	**0.97**	**0.97**
	SVC	0.39	0.10	0.25	0.14	0.39	0.10	0.25	0.14
	XGBC	0.95	**0.97**	**0.97**	**0.97**	0.95	0.97	**0.97**	**0.97**

(NFT and NFG), without feature selection, the Gaussian membership function returns better results than the Triangular one. Particularly, no significant differences are observed varying the splits.

This is also confirmed by the confusion matrices in Figure 2. Heatmap representation has been used to easily identify misclassifications. It can be seen that by using the Gaussian membership function the results are more accurate and a low number of false positives and negatives are returned. As it could be expected, the first two classes are easier to predict, since more samples are available. For the same reason, the more are the data in the training set (e.g. 80%), the better are the classification results. However, we can observe that whilst the neuro-fuzzy systems with Gaussian membership function have a low misclassification rate, and it confuses adjacent classes, that, in the medical domain, means subsequent stages of the disease. On the contrary, a higher number of errors is returned by the Triangular membership function, as suggested by the dark colors in the cells outside the principal diagonal. Moreover, in some cases, non-adjacent classes are confused. This is the case of Stage 2 that is predicted as Stage 1 or Normal (Fig. 3f), which is a very serious error, since suggesting that the patient is healthy while he is not.

Looking at the other classifiers, without feature selection, we can observe that, again, the best results are obtained with more data in the training set (split 80-20). In all configurations, the model that performs worse is SVC. Then there is MLP, followed by the RF. Finally, the best results are returned by XGBC, for all the splits.

In this first part of the experiments the black-box machine learning models (RF, MLP, and
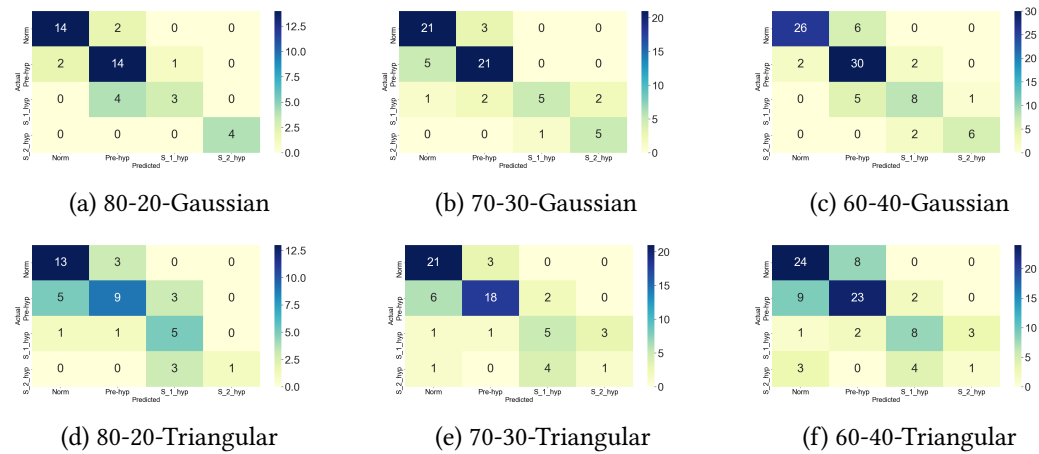
Figure 2: Comparison among the neuro-fuzzy models with different membership functions, experimentation setups, with and without features selection, in terms of confusion matrices.

XGBT), except for SVC, performed better than the neuro-fuzzy models, reaching in some cases an accuracy of 1.0 on the test set. However, as already said, they are not able to explain how predictions are derived. On the contrary, neuro-fuzzy models showed quite good results (the best accuracy achieved was 0.80), but they have the characteristic of being explainable and therefore a low decrease of accuracy could be preferred with an increase of explainability.

However, when using all the features, 2187 rules were returned by the neuro-fuzzy systems. This makes the system complex to understand, thus a feature selection process has been applied to reduce the number of rules and observe its influence on the classification performance.

Only two variables were selected as the most relevant by the feature selection process, namely SBD (Systolic Blood Pressure) and DBP (Diastolic Blood Pressure). The third section of Table 2 shows the qualitative results obtained by using these two features to learn the models.

The neuro-fuzzy models strongly improved their performance for all the splits. Particularly, the best improvements are obtained by the Triangular membership functions that return comparable results with the Gaussian membership function (the best accuracy is 0.93).

Figure 3 shows the confusion matrices of the neuro-fuzzy models. Almost all models are able to classify the normal hypertension class. As regards the classification of the other classes, also in this case the neuro-fuzzy models occasionally committed errors by confusing the adjacent class. The model with the most errors in classification has been the configuration with the Triangular membership function and with the split of the dataset into 60% for the training set and 40% for the test set (figure 3f).

Whilst the model based on neuro-fuzzy systems improved their performance, and a high reduction of accuracy is observed for MLP, the other classifiers were not affected by the feature selection.

However, it is worth pointing out that a strong reduction in the number of fuzzy rules has been obtained after the feature selection phase. Indeed, with 7 features 2187 fuzzy rules were returned (described by 3 membership functions each). With 2 features (again, with 3 membership functions each), the number of fuzzy rules has been drastically reduced to 9, as
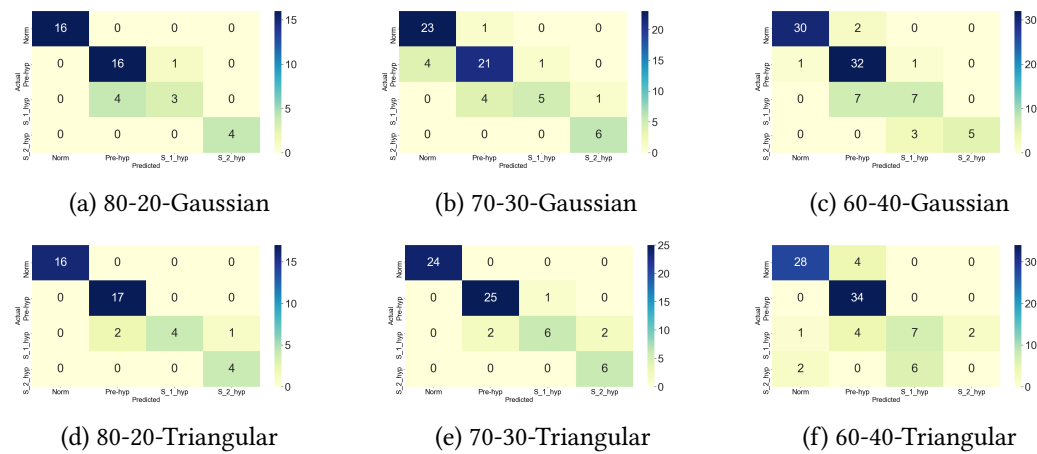
(a) 80-20-Gaussian

(b) 70-30-Gaussian

(c) 60-40-Gaussian

(d) 80-20-Triangular

(e) 70-30-Triangular

(f) 60-40-Triangular

Figure 3: Comparison among the neuro-fuzzy models with different membership function, experimentation setup and with features selection in terms of the confusion matrix.

shown in figure 4. The antecedents of the rules contain the two fuzzy variables returned by the feature selection (SBD and DBP) with all the configurations of the three fuzzy terms emerged by the neuro-fuzzy computation (low, medium, and high), as shown in figure 5. The consequents contain the four risk levels with the relative memberships. Thus, from the first four rules it is easy to understand that, the hypertension risk is Normal, if: SBD is low, and DBP is low, or SBD is low and DBP is medium, or SBD is low and DBP is high, or SBD is medium and DBP is low.

IF **SBD** is **low** and **DBP** is **low** THEN **Normal** = 1.0, **Prehypertension** = 0.0, **Stage_1_hypertension** = 0.16, **Stage_2_hypertension** = 0.20

IF **SBD** is **low** and **DBP** is **medium** THEN **Normal** = 1.0, **Prehypertension** = 0.16, **Stage_1_hypertension** = 0.0, **Stage_2_hypertension** = 0.10

IF **SBD** is **low** and **DBP** is **high** THEN **Normal** = 1.0, **Prehypertension** = 0.76, **Stage_1_hypertension** = 0.07, **Stage_2_hypertension** = 0.0

IF **SBD** is **medium** and **DBP** is **low** THEN **Normal** = 0.94, **Prehypertension** = 1.0, **Stage_1_hypertension** = 0.03, **Stage_2_hypertension** = 0.0

IF **SBD** is **medium** and **DBP** is **medium** THEN **Normal** = 0.42, **Prehypertension** = 1.0, **Stage_1_hypertension** = 0.40, **Stage_2_hypertension** = 0.0

IF **SBD** is **medium** and **DBP** is **high** THEN **Normal** = 0.0, **Prehypertension** = 0.88 , **Stage_1_hypertension** = 1.0, **Stage_2_hypertension** = 0.17

IF **SBD** is **high** and **DBP** is **low** THEN **Normal** = 0.0, **Prehypertension** = 0.95, **Stage_1_hypertension** = 0.67, **Stage_2_hypertension** = 1.0

IF **SBD** is **high** and **DBP** is **medium** THEN **Normal** = 0.0, **Prehypertension** = 0.34, **Stage_1_hypertension** = 0.70, **Stage_2_hypertension** = 1.0

IF **SBD** is **high** and **DBP** is **high** THEN **Normal** = 0.15, **Prehypertension** = 0.0, **Stage_1_hypertension** = 0.75, **Stage_2_hypertension** = 1.0

Figure 4: Example of fuzzy rules generated by the neuro-fuzzy model with the feature selection process.

Overall, whilst qualitative results are comparable to those obtained by the best machine learning models, the neuro-fuzzy systems are able to return interpretable results, that help clinicians in understanding and trusting the process behind the algorithms.
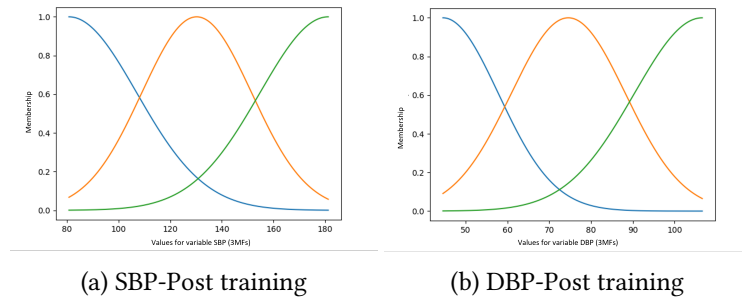
(a) SBP-Post training   (b) DBP-Post training

Figure 5: Example of Gaussian membership function generated by the neuro-fuzzy model with the feature selection process.

4. Conclusion

Four machine learning algorithms have been compared with two neuro-fuzzy systems (based on Gaussian and Triangular membership functions) for hypertension assessment. The experiments aimed to evaluate if the qualitative performance of the two NFS models were higher, or at least comparable, with those given by the ML methods, with the added value of the explainability that fuzzy logic allows.

Since the dataset is unbalanced, three different experimental settings have been used. Moreover, further experiments, with a reduced number of features, have been conducted, to enhance the explainability of the neuro-fuzzy systems.

Results have shown that, without feature selection, the Gaussian membership function obtains higher performance than the Triangular one, but still lower than the machine learning methods. However, by considering all the seven features in data, the number of rules is too high to be understandable. Thus, the two most relevant features have been selected, leading to a significant reduction of the number of rules (from 2187 to 9). Feature selection has also improved the performance of the neuro-fuzzy systems, while machine learning methods have preserved their quantitative values, or as for MLP they have been reduced.

Overall, experiments have shown that NFSs are useful support tools for hypertension risk assessment since while returning accurate results, they are also able to explain with linguistic terms how these results have been obtained. In the medical domain, this is crucial, since both patients and medical staff need to understand and trust the automatic tools. Future work will be devoted to better studying the model explainability. To this aim, different algorithms will be compared and domain experts will be involved in evaluating the explanations.

References

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

[2] 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.

[3] V. S. Kublanov, A. Y. Dolganov, D. Belo, H. Gamboa, Comparison of machine learning methods for the arterial hypertension diagnostics, Applied bionics and biomechanics 2017 (2017).

[4] A. Bajaj, M. Bhatnagar, A. Chauhan, Recent trends in internet of medical things: a review, Advances in Machine Learning and Computational Intelligence (2021) 645–656.

[5] M. T. Angelillo, F. Balducci, D. Impedovo, G. Pirlo, G. Vessio, Attentional pattern classification for automatic dementia detection, IEEE Access 7 (2019) 57706–57716.

[6] C. Ardito, T. Di Noia, C. Fasciano, D. Lofù, N. Macchiarulo, G. Mallardi, A. Pazienza, F. Vitulano, Management at the edge of situation awareness during patient telemonitoring, in: International Conference of the Italian Association for Artificial Intelligence, Springer, 2020, pp. 372–387.

[7] 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.

[8] 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.

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

[10] 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.

[11] C. Mencar, G. Castellano, A. M. Fanelli, On the role of interpretability in fuzzy data mining, International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems 15 (2007) 521–537.

[12] U. Kaymak, On using fuzzy sets in healthcare process analysis, in: International Conference on Theory and Applications of Fuzzy Systems and Soft Computing, Springer, 2018, pp. 24–24.

[13] J. Jang, Anfis: adaptive-network-based fuzzy inference system, IEEE Trans. Syst. Man Cybern. 23 (1993) 665–685.

[14] A. Abraham, Neuro fuzzy systems: State-of-the-art modeling techniques, in: International Work-Conference on Artificial Neural Networks, Springer, 2001, pp. 269–276.

[15] T. Takagi, M. Sugeno, Fuzzy identification of systems and its applications to modeling and control, IEEE transactions on systems, man, and cybernetics (1985) 116–132.

[16] J.-S. Jang, C.-T. Sun, Neuro-fuzzy modeling and control, Proceedings of the IEEE 83 (1995) 378–406.