Neural Additive Models for Explainable Heart Attack Prediction

Ksenia Balabaeva[1] and Sergey Kovalchuk[1]

[1] ITMO University, Saint Petersburg, Russia
kyubalabaeva@gmail.com
sergey.v.kovalchuk@gmail.com

Abstract. Heart attack (HA) is a sudden health disorder when the flow of blood to the heart is blocked, causing damage to the heart. According to the World Health Organization (WHO), heart attack is one of the greatest causes of death and disability globally. Early recognition of the various warning signs of a HA can help reduce the severity. Different machine learning (ML) models have been developed to predict the heart attack. However, patients with arterial hypertension (AH) are especially prone to this disorder and have several features that distinguish them from other groups of patients. We apply these features to develop a special model for people suffering from AH. Moreover, we contribute to this field bringing more transparency to the modelling using interpretable machine learning. We also compare the patterns learned by methods with prior information used in heart attack scales and evaluate their efficiency.

Keywords: heart attack prediction, XAI, Decision Tree, NAMs, heart attack risk, arterial hypertension, interpretable machine learning

1 Introduction

Heart diseases are one of the main reasons of a death rate increase each year. Heart Attack (HA) is a sudden disruption of a cardiac system. In healthcare sector tons of data are generated in electronic health records (EHR) about patients. These records describe main patients' characteristics helping clinicians make decisions on diagnostics and treatment. According to one of the surveys conducted by WHO, the clinicians can accurately predict only 67% of heart diseases [1]. There are many factors possible influencing the outcome of heart disease and it's impossible to take everything in mind but still there is potential to increase the quality.

EHRs are valuable sources of data for machine learning models and medical decision support systems. Such systems already proved their accuracy, however, the pace of implementation and practical use remains

ICCS Camera Ready Version 2022
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-08757-8_11

low especially in medical institutions. One of the reasons of such refuse of AI technologies is a lack of trust and understanding among users.

We implement a machine learning based system that can detect and predict heart attack for patients using the medical records. The proposed solution is based on Neural Networks and generalized additive models (GAMs) [2, 3].

The dataset used in our study was collected from Almazov Medical Research Center. It contains 17 features (such as age, gender, blood pressure, etc.) and 385 observations after preprocessing and data clearance. It was then split into 70% train sets and 30% test sets. Moreover, cross-validation was performed to compare methods and optimize hyperparameters. A series of experiments was conducted to examine the performance, accuracy, and interpretation stability of the proposed system. Experiments and were and modules were implemented in Python 3 programming language which predicts the risk of heart attack among patients with AH. The results show that the NAMs performance and accuracy is high enough for the task and it provides helpful interpretation.

2 Related Works

Different researchers have contributed to the development of digitalization and predictive analytics in medical domain. Prediction of heart disease based on machine learning algorithm is always curious case for researchers.

A concept of explainability and interpretation also plays an import role in healthcare [4]. Early works were devoted to IF-ELSE rules, where researchers modelled a set of diseases (lung cancer, asthma, and diabetes) based on electronic health records [5]. Researchers in [6] applied LIME to explain the prediction of heart failure by recurrent neural networks. They also provided explanations that allowed to identify the risk-factors such as kidney failure, anemia, and diabetes that increase the risk of heart failure.

In [7] authors used an algorithm named weighted association rule-based classifier (WAC), based on association rule mining to predict heart attack. Florence et al. [8] applied decision trees and artificial neural networks to the same task. In the work [9] authors used an algorithm based on graph association rules mining. One of the drawbacks of early works was poor accuracy and a lack of interpretation and transparency

DOI: 10.1007/978-3-031-08757-8_11

of the system. Therefore, recent advances of AI in medicine are achieved thanks to explainable artificial intelligence (XAI) using molecular data [10], deep meta-learning [11], and other methods [12]. In 2019 Madan M. et al proposed a technique based on combination of Genetic Algorithm (GA) and Adaptive Neural Fuzzy Inference System (ANFIS) for explainable heart attack prediction [13]. However, the proposed approach wasn't compared with other ML algorithms to conclude about predictive efficiency.

In this work we're testing a novel transparent machine learning approach named Neural Additive Models (NAMs) for the task of heart attack prediction. NAMs were proposed by X et al [3]. We also compare the predictive capacity of method with stability of explanations

3 Methods

For many years neural nets outperformed other ML algorithms mostly while applied on unstructured data (images, text, etc.). However, in 2021 Agarwal et al proposed a novel method bringing concept of neural networks (NN) to transparent class of generalized additive models (GAMs) [2, 3]. It allows to efficiently train neural nets on structured tabular data. GAMs have the form:

$$g(\mathbb{E}[y]) = \beta + f_1(x_1) + f_2(x_2) + \cdots + f_m(x_m) \qquad (1)$$

Where $x = (x_1, \dots, x_m)$ is a vector of Mfeatures, y is a target variable, $g(.)$ is a link function (exponential, logistic, etc.) with f_i being univariate shape functions with $\mathbb{E}[f_i] = 0$.

Compared to GAMs, NAMs learn a linear combination of networks, where each separate network is trained on a single feature $x_i \in M$. These networks are trained jointly, using classic backpropagation mechanism.

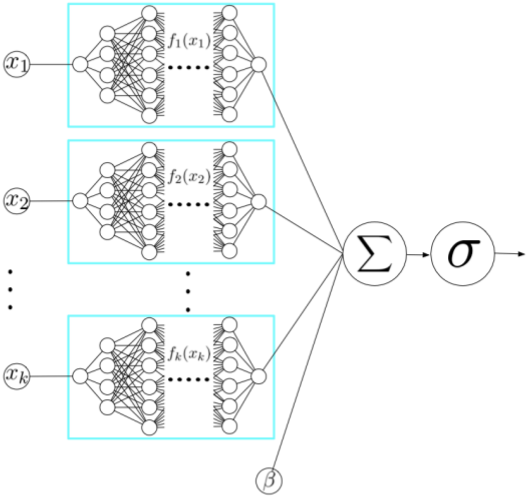

Figure 1 NAMs architecture for binary classification [3]

Interpretation of NAMs is possible in the form of feature importance since each subnet is independent from other features and can be calculated separately. Moreover, each single subnet in the architecture can be represented as graph and “exactly describe how NAMs computes the prediction”. In our study we compare NAMs to other self-explained ML methods, that can provide interpretation in the form of feature importances: Decision Tree, XGBoost and statistical model Logistic Regression. Moreover, we conduct additional experiments, to evaluate performance of widely-used clinical scale, named SCORE for evaluating heart disease risk development [15].

4 Experiments

The dataset was collected from Almazov Medical Research Center, Saint-Petersburg. It contains 385 observations representing patients with arterial hypertension. 281 have suffered from heart attack and 104 didn’t. The dataset includes men and women from 37 to 87 years old (fig. 1).

ICCS Camera Ready Version 2022
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-08757-8_11

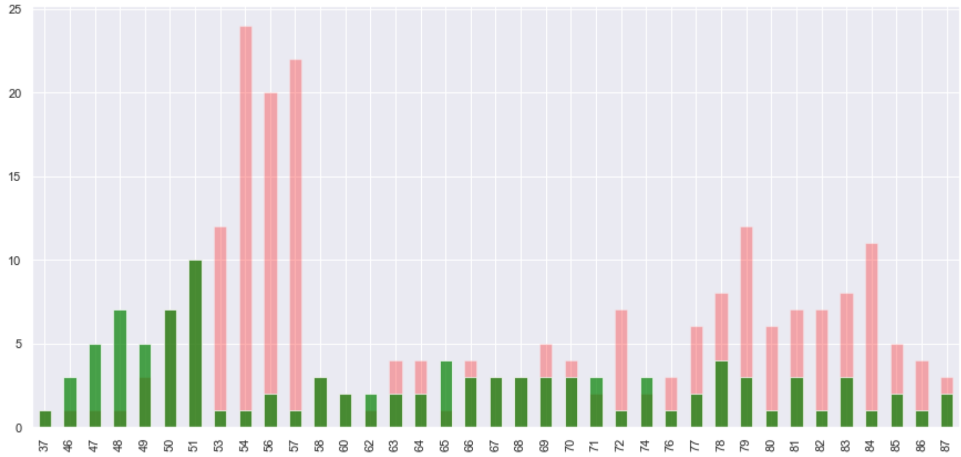

Figure 2 Age distribution for patients with (pink) and without (green) heart attack

The feature set includes the following information: gender, age, height, weight, body mass index (BMI), body square area (BSA), smoking (0, 1), diabetes (0, 1). Systolic and diastolic blood pressure measurements: dad_min, dad_max, sad_min, sad_max. Laboratory test results: lpvp, lpnp, alanine transaminase (ALT), aspartate aminotransferase (AST), urine. The descriptive statistics about features used for modelling is provided in table 1.

Table 1. Features' descriptive statistics

Feature Name	Mean Value	Standard Deviation
Gender	0.4857	0.50
Age	63.166	13.419
Height	167.71	7.872
Weight	80.65	14.32
BMI	29.17	8.815
BSA	1.943	0.2400
Smoking	0.127	0.333
Diabetes	0.945	0.227
DAD min	80.41	7.146
DAD max	89.48	7.589
SAD min	141.042	12.77
SAD max	151.22	15.109
LPVP	1.326	0.4105
LPNP	2.763	0.6609
ALT median	21.487	18.590

AST median	21.81	10.129
Urine	6.76	2.157
Stroke	0.729	0.44

The features for modelling were aggregated from several clinical episodes using descriptive statistics: mean, median, minimal, maximal and std values in cases where patients had previous episodes with necessary data collection. The outliers were detected with 3-sigma rule and clinical norms' limits. The missing values were filled with median.

The whole sample was randomly split into 70% training set and 30% test set. We selected 30% threshold for test to provide efficient data for model's testing and keep enough data for training.

Heart attack prediction can be solved as a binary classification task, with feature matrix X, target vector y, and model $f(X)$ that is trained on X to predict y. We selected a set of machine learning algorithms that can provide explanation in the form of feature importances: NAMs, Gradient boosting, Decision Tree, and Logistic Regression. In each approach, feature importances are calculated in different ways. Moreover, we compare ML algorithms to used scale SCORE for heart disease risk calculation that is widely used among clinicians. To adopt SCORE predictive capacity, we consider the highest possible risk (threshold >15%) as a positive class, and all other risks – as negative.

To compare predictive performance, we evaluate algorithms using F-score on 5-fold cross-validation and hold-out 30% test set. To test the interpretation stability, we train and test models on different random seeds (seed = 42; 2021; 2022) and evaluate the change in feature importance ranking using NDCG-metric. All experiments were conducted using Python 3.7.

5 Results and Discussion

Considering the predictive performance, best result both on test and validation was achieved by NAMs. Considering F1 on test set, XGB and logistic regression show similar but weaker results. That means that NAM outperforms other explainable ML algorithms in terms of accuracy and can be used for heart attack prediction in medical decision support systems (table 2).

Table 2. Predictive Performance of algorithms

Method	F1 Cross-Validation (STD)	Test
SCORE	0.3089 (±0.114)	0.419
XGB	0.7142 (±0.0183)	0.8349
Logistic Regression	0.7324 (±0.01)	0.8317
Decision Tree	0.6909 (±0.0528)	0.805
NAM	**0.8778 (±0.026)**	**0.8713**

Meanwhile, the SCORE scale showed the worst quality. Since this is a scale based on several rules, it can't observe dependencies in data. Moreover, since our dataset has bias (all patients have arterial hypertension), some of the rules in SCORE might not be valid. For instance, according to SCORE, the higher is the blood-pressure – the higher is the risk of a heart disease. But many patients take medical treatment to normalize systolic and diastolic blood pressure, which apparently doesn't guarantee the lower HA risk.

The consistency of interpretation was measured using NDCG - score (table 3), calculating the difference in rank and weight of feature importance trained on the same models using different random seeds (42, 2021, 2022) and taking the mean value of such comparisons. The change in the seed also influenced train-test split, so the score includes data perturbations.

Table 3. Evaluation of interpretation stability

Method	Feature Importance NDCG
NAM	0.7150
Logistic Regression	0.93937
XGB	0.9545
Decision Tree	0.7709

The highest stability was achieved by XGBoost and Logistic Regression, and the worst – by NAMs. We think that poor stability of NAM's interpretation might be caused by the lack of data, since usually neural nets require enormous data volumes for efficient and stable training.

Now let's discuss what exactly NAM's interpretation reveals. The most important risk-factors for prognosis were parameters related to blood pressure (SAD max, DAD max), height and age (fig. 3). Gender and AST have the lowest importance and doesn't influence model's prediction.

Mean Importance

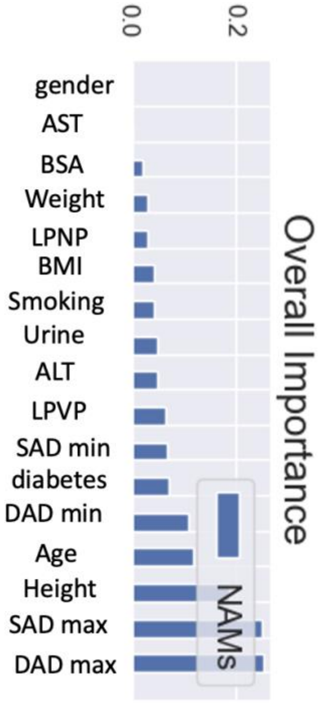

Figure 3 NAMs Feature importances

Using NAM for modelling, offers visualization of a detailed feature contribution in the form of graphs (fig. 4). Light-pink regions in graphs correspond to regions with low data density (few samples in a dataset), on the contrary, red regions correspond to high density in data. Blue line depicts shape functions, that tend to be smooth in regions with high density and serrated in regions of few data samples.

Thus, we see that low values of the blood pressure slightly increase the. It might seem counterintuitive, but extremely high values of maximum systolic and diastolic pressure may indicate the severe course of the disease and anti-AH treatment therapy. The high maximal value of arterial hypertension might be the evidence of a crisis when treatment fails to normalize the pressure. But at the same time, mean SAD and DAD might be much lower most of the time. Unfortunately, we can't test this hypothesis since blood pressure measurements were collected from anamnesis.

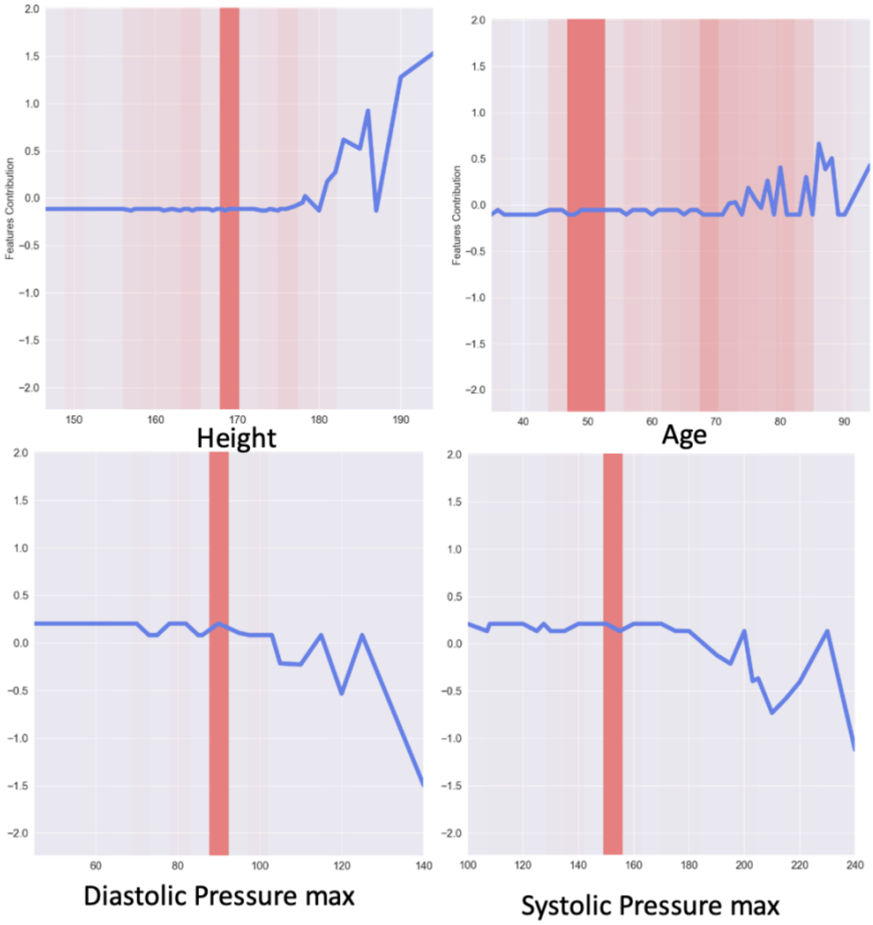

Figure 4 NAMs intrinsic interpretation

As for the patient's height, if it's lower that 178 cm, the height lowers the HA risk, however, if a person is higher than 180 cm the risk the positive feature contribution rises significantly. It is also clear that a model learned age patterns, the older a person – the more is a positive contribution of Age in a model.

Due to the transparency of NAM, we can find more insides concerning the connection between disease and risk factor, especially when it comes to specific patients' groups with different comorbidities, such as arterial hypertension or diabetes, where, as we saw, universal conservative methods of risk scoring (SCORE) may fail in terms of predictive performance.

6 Conclusion

To provide explainable and more transparent results for clinicians n a novel approach NAM was tested to predict heart attack among patients with arterial hypertension. The experiment confirmed that NAM predicts HA with a high accuracy and provide explanation in the form of graph and feature importances, showing redundant information on the influence of each risk factor for the prediction. However, the stability of interpretation might suffer from data perturbation and randomization parameter. This drawback can be potentially solved with multiple training of NAM on different seeds and averaging the feature importances.

Acknowledgement

This work was supported by the Ministry of Science and Higher Education of Russian Federation, goszadanie no. 2019-1339

References

1. V. Kirubha and S. M. Priya, “Survey on Data Mining Algorithms in Disease Prediction,” vol. 38, no. 3, pp. 124–128 (2016).
2. Hastie, T. J., & Tibshirani, R. J. Generalized additive models. Routledge (2017).
3. Agarwal, Rishabh, et al. "Neural additive models: Interpretable machine learning with neural nets." Advances in Neural Information Processing Systems 34 (2021).
4. Khedkar, S., Subramanian, V., Shinde, G., & Gandhi, P. (2019). Explainable AI in healthcare. In Healthcare 2nd International Conference on Advances in Science & Technology (ICAST) (2019).
5. H. Lakkaraju, S. H. Bach, and J. Leskovec, “Interpretable decision sets: A joint framework for description and pre-diction,” Proceedings of the ACM SIGKDD

InternationalConference on Knowledge Discovery and Data Mining, vol., pp. 1675–1684, 13-17 (2016).

6. S. Khedkar, V. Subramanian, G. Shinde, and P. Gandhi, “Explainable AI in Healthcare,” SSRN Electronic Journal (2019).
7. Soni J, Ansari U, Sharma D, Soni S. Intelligent and effective heart disease prediction system using weighted associative classifiers. International Journal on Computer Science and Engineering; 3: 2385-2392 (2011).
8. Florence S, Bhuvaneswari Amma NG, Annapoorani G, Malathi K. Predicting the risk of heart attacks using neural network and decision tree. International Journal of Innovative Research in Computer and Communication Engineering; 2: 7025-7030 (2014).
9. Jabbar MA, Deekshatulu BL, Chandra P. Graph based approach for heart disease prediction. Proceedings of ITC 2012, Bangalore, Springer-Verlag; 150: 465-474 (2012).
10. Westerlund, A. M., Hawe, J. S., Heinig, M., & Schunkert, H.). Risk Prediction of Cardiovascular Events by Exploration of Molecular Data with Explainable Artificial Intelligence. International Journal of Molecular Sciences, 22(19), 10291 (2021).
11. Dağlarli, E. Explainable artificial intelligence (xAI) approaches and deep meta-learning models. Advances and applications in deep learning, 79 (2020).
12. Duell, J., Fan, X., Burnett, B., Aarts, G., & Zhou, S. M. A comparison of explanations given by explainable artificial intelligence methods on analysing electronic health records. In IEEE EMBS International Conference on Biomedical and Health Informatics (BHI) (pp. 1-4). IEEE. (2021).
13. Aghamohammadi, Mehrdad, et al. "Predicting heart attack through explainable artificial intelligence." International Conference on Computational Science. Springer, Cham (2019).
14. Author, F.: Article title. Journal 2(5), 99–110 (2016).
15. Systematic COronary Risk Evaluation (SCORE): https://www.escardio.org/Education/Practice-Tools/CVD-prevention-toolbox/SCORE-Risk-Charts

DOI: 10.1007/978-3-031-08757-8_11