Model for diagnosing cardiovascular diseases using AI

Denis Mušić, Muhamed Karajić, Lejla Hećo, Emir Osmanović   
"Džemal Bijedić" University of Mostar

Faculty of Information Technologies  
Mostar, Bosnia and Herzegovina  
[{denis, muhamed.karajic,](mailto:{denis}muhamed.karajic@edu.fit.ba) [lejla.heco,](mailto:lejla.heco@edu.fit.ba)[emir.osmanovic}@edu.fit.ba](mailto:emir.osmanovic@edu.fit.ba)

*Abstract* — Cardiovascular disease (CVD) requires expert supervision to be effectively treated. In addition to the necessity to diagnose CVD as soon as possible, there are often cases in which patients with a substantial risk of contracting this disease are very often categorized and treated in the same way as the ones with a far lower risk. Specific symptoms often make it difficult for physicians to recognize certain forms of CVDs, and therefore correct diagnosis requires many years of experience and additional education. To contribute to the application of modern algorithms of Artificial Intelligence (AI) for the purpose of diagnosing diseases, this paper analyzed the practical application of certain algorithms and proposes a model for diagnosing CVDs. The model showed satisfactory performance, and with minor extensions and adjustments, it could be used in cardiology and family medicine clinics as an assistant to doctors.

Keywords— cardiovascular diseases, diagnosing, machine learning, artificial intelligence

# Introduction

Cardiovascular diseases (CVDs) is a group of disorders of heart and blood vessels, including coronary artery disease, cerebrovascular disease, rheumatic heart disease, and other conditions [3]. CVDs are the leading cause of death worldwide. In 2019, an estimated 17.9 million deaths were attributed to cardiovascular disease (CVD), comprising approximately 32% of all global fatalities. Among these, 85% were specifically linked to heart attacks and strokes [10]. The global mortality rate from CVDs has shown a significant increase, rising from 12.1 million in 1990 to 20.5 million in 2021, as documented by the World Heart Federation (WHF) [1].

More than three-quarters of CVD deaths occur in low-income and middle-income countries. The highest mortality rates from cardiovascular diseases are in Africa, Asia, Eastern Europe, and South America [2]. The major risk factors for CVD and stroke are unhealthy diet, insufficient physical activity, smoking, and harmful use of alcohol. The effects of behavioral risk factors can manifest as high blood pressure, high blood glucose levels, high blood lipid levels, and overweight or obesity [3].

Quitting smoking, cutting back on salt, eating more fruits and vegetables, staying physically active, and moderating alcohol intake have all been demonstrated to lower the chances of developing CVD. Recognizing individuals with the greatest risk of CVD and ensuring they get suitable treatment can help prevent early deaths. It's vital to have access to medications for chronic diseases and necessary health technologies in all primary healthcare facilities to guarantee that those requiring treatment and counseling receive it [3].

AI has the potential to completely transform the process of categorizing diseases, and automatically interpreting Electrocardiogram[[1]](#footnote-16465) (ECGs), facilitating early disease detection. This should ease the patient-physician journey, reduce treatment costs, and encourage the belief that the application of AI will stop the trend of increasing CVD mortality worldwide.

Based on the analysis of contemporary trends this study focuses on exploring the application of AI to prevent CVDs through early detection. To accomplish such a goal, the project presented in this paper provides Application Programming Interface (API) that offers the trained model as a *Software as a Services* (SaaS) which will be available to public and private medical institutions to influence individual health outcomes by optimizing the workflow of healthcare professionals, physicians, and cardiologists. This is intended to improve their ability to prioritize patients at higher risk which will have a side effect to deal with CVDs more efficiently.

For the implementation of AI into diagnostic processes, API uses ML.NET library which serves as the implementation foundation for our research paper and allows use of machine learning (ML) algorithms. The proposed AI model, validated through training and implementing the web application, will be elaborated further in the subsequent sections of this paper.

# Previous Research

Application of AI in medicine has enviable results in recent years and is also becoming the subject of increasing interest of researchers. Some of those studies already explore the extensive possibilities of AI in diagnosing CVDs. Although it has been shown that the potential of AI in CVDs is promising, these studies are lacking practical use of the trained models while others were missing crucial information to understand how well the model performs in context of specificity and sensitivity [4].

The fact that the large amount of data is the main prerequisite for the creation of any AI prediction model, a database of cardiovascular diseases was created as part of the research [5] which focuses on the early detection of risk factors, developing predictive models for early diagnosis, and improving patient care. The creation of this dataset involved 750,000 anonymized patients for CVD, from Asan Medical Center and Ulsan University Hospital. The database's main purpose was to provide clinicians and engineers with easy access to key patient information.

Interesting research [6] resulted in development of the framework called Machine Learning based Cardiovascular Disease Diagnosis (MaLCaDD) which provides basis for early prediction and diagnosis of cardiovascular diseases. Through four key phases, from addressing missing value problems to using ensemble models, MaLCaDD demonstrates acceptable level in prediction, surpassing many existing studies.

Research that focuses on discussion of the contribution of AI and ML in healthcare, followed by an exploration of the benefits and challenges was presented in [7]. It also discusses the experimental approach to predicting CVDs ML techniques. Their study has demonstrated that among all classifiers evaluated in CVD classification, the Support Vector Machine (SVM) algorithm exhibited superior performance, achieving an accuracy rate of 82.5%. In-depth analyses of XAI techniques, including feature selection, explaining feature weight initialization, normalization, and optimization, are explored for proper explanation and interpretation of results. XAI-guided models for classifiers for predicting CVDs are evaluated on the Cleveland dataset. Future research dimensions could include statistical analysis of models on different datasets, such as the UK Biobank dataset, Federico II University dataset, Statlog heart disease dataset, and NEU hospital dataset for heart diseases.

# Proposed Model

In this part of the paper, we present an ML model tailored for the prediction of CVDs. The model is thoroughly designed, incorporating a range of algorithms considered by the library and parameters such as: inputs and outputs to understand how well the model performed.

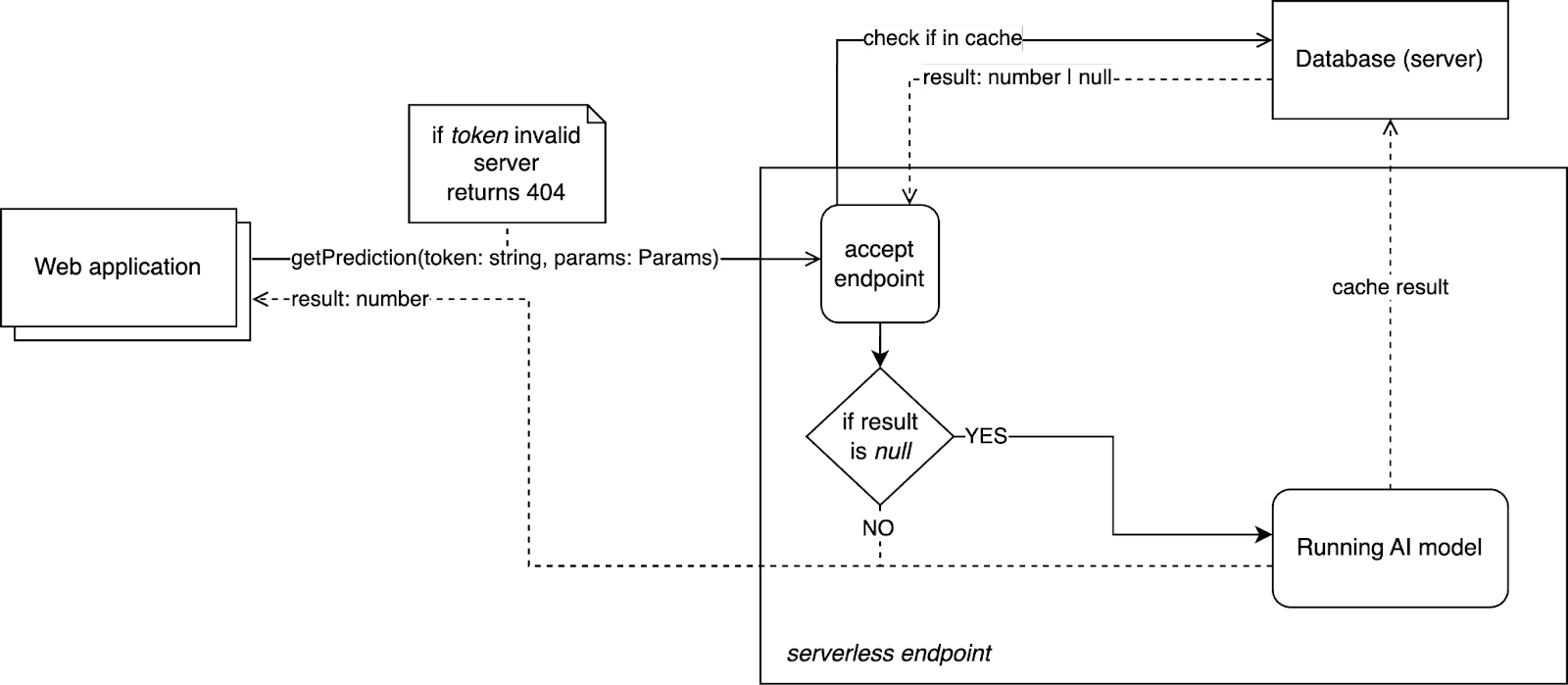


Fig. 1. *Proposed model for diagnosing cardiovascular diseases using AI*

The model of the complete implementation infrastructure, which also served for validation, is presented in Fig.1.

## *Framework and Technologies Utilized*

The model is developed using the ML.NET library, a versatile and efficient ML tool that integrates seamlessly with the .NET ecosystem. ML.NET provides a rich set of functionalities for building, training, and deploying ML models, making it an ideal choice for developing predictive models such as the CVD model desired in this paper.

## *Parameterization*

The model's parameters are carefully chosen to enhance its performance and predictive accuracy. Important highlights encompass various aspects, including:

* *Feature Columns*: These parameters represent the input variables used to train the model. In this context, features such as cholesterol levels, resting blood pressure, and various Electrocardiogram (ECG) measurements are utilized to capture relevant information for CVD prediction.
* *Label Column*: The label column specifies the target variable to be predicted by the model. In this case, it is the presence or absence of CVD, encoded as binary values where 0 means absence and 1 stands for presence.
* *Algorithm Configuration*: The model employs the FastTree binary classification algorithm, a decision tree-based ensemble learning technique known for its efficiency and effectiveness in handling classification tasks.

## *Algorithm Description*

The FastTree algorithm utilized in this model operates by recursively partitioning the feature space into regions, aiming to minimize the classification error at each step. Through iterative splitting of feature values, the algorithm constructs a decision tree that can efficiently classify new instances based on their feature values. FastTree's adaptability to large datasets and its ability to handle fast non-linear relationships between features make it well-suited for the CVD prediction task.

## *Data Loading and Training Process*

The model loads training data from a CSV file and prepares it for training. This data includes clinical parameters and corresponding labels for CVD presence or absence. Then it constructs an ML pipeline and trains the FastTree algorithm to optimize decision tree structure based on the input data. After training, the model obtains a trained transformer encapsulating learned patterns for making predictions.

## *Inputs*

The model takes as input a set of clinical parameters and measurements typically collected during a patient's medical examination. These inputs are described in [8] and they include:

* Cp (Chest Pain Type): Categorical variable representing the type of chest pain experienced by the patient.
* TrestBps (Resting Blood Pressure): Numerical variable indicating the patient's resting blood pressure.
* Chol (Serum Cholesterol Level): Numerical variable describing the patient's serum cholesterol level.
* Fbs (Fasting Blood Sugar): Binary variable indicating whether the patient's fasting blood sugar is above a certain threshold.
* RestEcg (Resting Electrocardiographic Results): Categorical variable describing the results of the resting electrocardiogram.
* Thalach (Maximum Heart Rate Achieved): Numerical variable representing the patient's maximum heart rate achieved during the exercise stress test.
* Exang (Exercise Induced Angina): Binary variable indicating whether the patient experienced exercise-induced angina.
* OldPeak (ST Depression Induced by Exercise): Numerical variable representing the ST depression induced by exercise relative to rest.
* Slope: Categorical variable describing the slope of the peak exercise ST segment.
* Ca (Number of Major Vessels Colored by Fluoroscopy): Numerical variable indicating the number of major vessels colored by fluoroscopy.
* Thal (Thalassemia): Categorical variable describing a blood disorder called thalassemia.

## *Outputs*

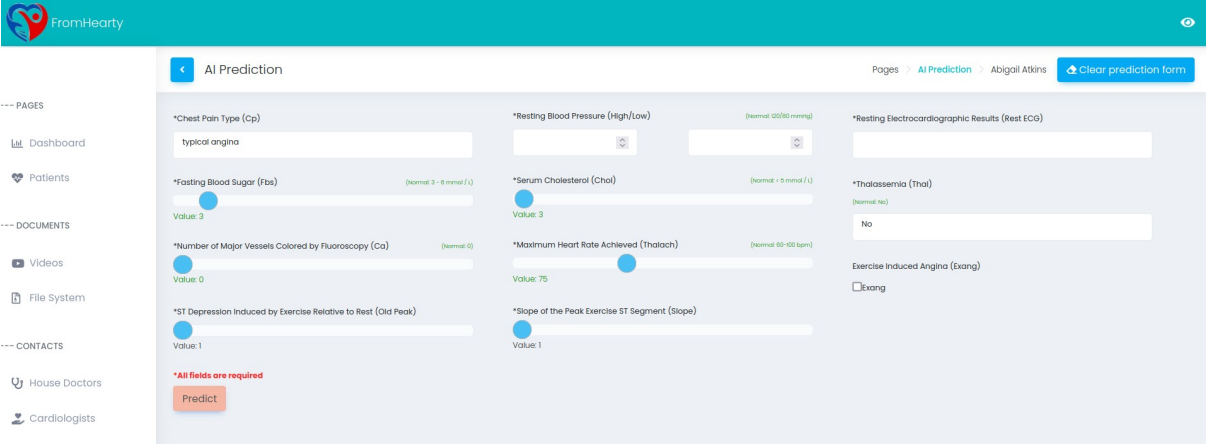
The model generates predictions regarding the likelihood of CVD based on the input parameters. The output is encapsulated within one class which includes three main components:

* *Prediction Label*: A boolean value indicating the predicted presence or absence of CVD.
* *Probability*: A float value representing the confidence level of the prediction.
* *Score*: Internal scoring or ranking assigned to the prediction by the FastTree algorithm.

# Model Validation

As illustrated in Fig. 1., the AI model developed in this research is deployed as an API accessible to and type of application seeking to present users with the likelihood of a patient's illness or to support diseases classification. Our methodology leverages the deterministic nature of the ML algorithm FastTree [1] enabling the database to function as a cache which optimizes cost efficiency by returning precomputed AI results. The interface of the web application interface used to test the model is shown on Fig. 2.

The model evaluation process involves partitioning the dataset into distinct training and testing subsets. Approximately 70% of the dataset comprises the training set, dedicated to model training. The remaining 30% constitutes the testing set, deliberately excluded during model creation to serve as an independent measure of the AI's performance. This separation ensures an unbiased assessment of the AI's efficacy.

  
Fig. 2. *Interface of the Web application implemented for model validation purposes*

As described with the confusion matrix in Table I. the proposed AI model achieved an accuracy rate of 0.981. While accuracy serves as a primary metric, specificity and sensitivity are equally essential parameters, assessing false positives and false negatives, respectively.

1. Confusion Matrix

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Absence** | **Presence** | **Recall** |
| **Absence** | 155 | 3 | 0.9810 |
| **Presence** | 3 | 161 | 0.9817 |
| **Precision** | **0.9810** | **0.9817** |  |

While sensitivity is more commonly used in biological research, in ML.NET it is called Recall, and precision could be interpreted as specificity.

# Conclusion

The presented AI model offers a robust and reliable tool for predicting CVDs. By leveraging advanced algorithms using ML.NET, carefully selected parameters as well as training and test sets – even in the early parts of the development that the model shows promising results with accuracy, specificity and sensitivity reaching over 98%. We retrained the model multiple times with different magnitudes for the training and test set while the accuracy, specificity and sensitivity were at reasonable levels. Therefor additional techniques such as cross validation was not necessary, saving time and effort as well as making it possible to focus on further improving our solution by expanding our dataset by negotiating with hospitals. Therefore, it can be concluded that the model is usable for identifying patients with potential risks of CVDs is serving its ultimate purpose which is improving healthcare.

1. Deaths from cardiovascular disease surged 60% globally over the last 30 years (Report), published: 20th May 2023, Published online at world-heart-federation.org. Retrieved from: ‘https://world-heart-federation.org/news/deaths-from-cardiovascular-disease-surged-60-globally-over-the-last-30-years-report/’, [Online Resource] [Accessed: 3rd February 2024]
2. Saloni Dattani (2023) - What are the different types of cardiovascular diseases, and how many deaths do they cause? Published online at OurWorldInData.org. Retrieved from: 'https://ourworldindata.org/ cardiovascular-diseases-types-and-death-tolls' [Online Resource] [Accessed: 15th December 2023]
3. Cardiovascular diseases, Overview, Sypmtoms, Treatment, Published online at www.who.int. Retrieved from: ' https://www.who.int/health-topics/cardiovascular-diseases' [Online Resource] [Accessed: 10th January 2024]
4. Karajić, Muhamed & Begic, Edin & Hrvat, Emina & Gurbeta Pokvic, Lejla. (2021). Application of Artificial Intelligence Tools in Classification and Diagnosis of Heart Disease: General Review. 10.1007/978-3-030-73909-6\_31.
5. Ahn I, Na W, Kwon O, Yang DH, Park GM, Gwon H, Kang HJ, Jeong YU, Yoo J, Kim Y, Jun TJ, Kim YH. CardioNet: a manually curated database for artificial intelligence-based research on cardiovascular diseases. BMC Med Inform Decis Mak. 2021 Jan 28;21(1):29. doi: 10.1186/s12911-021-01392-2. PMID: 33509180; PMCID: PMC7842077.
6. A. Rahim, Y. Rasheed, F. Azam, M. W. Anwar, M. A. Rahim and A. W. Muzaffar, "An Integrated Machine Learning Framework for Effective Prediction of Cardiovascular Diseases," in IEEE Access, vol. 9, pp. 106575-106588, 2021, doi: 10.1109/ACCESS.2021.3098688.
7. Guleria, Pratiyush, Parvathaneni Naga Srinivasu, Shakeel Ahmed, Naif Almusallam, and Fawaz Khaled Alarfaj. 2022. "XAI Framework for Cardiovascular Disease Prediction Using Classification Techniques" Electronics 11, no. 24: 4086. https://doi.org/10.3390/electronics11244086
8. Yazdani A, Varathan KD, Chiam YK, Malik AW, Wan Ahmad WA. A novel approach for heart disease prediction using strength scores with significant predictors. BMC Med Inform Decis Mak. 2021 Jun 21;21(1):194. doi: 10.1186/s12911-021-01527-5. PMID: 34154576; PMCID: PMC8215833.

1. Electrocardiogram – a diagnostic test that measures the electrical activity of the heart over a period of time. [↑](#footnote-ref-16465)