HEART FAILURE RISK PREDICTION USING AZURE DATA LAKE ARCHITECTURE WITH AUTOMATED MACHINE LEARNING AND MACHINE LEARNING APPROACHES

by

Ahmed M. ALGHAMDI[a*]***, Waleed Al SHEHRI***[b]***,***
Jameel ALMALKI[b]***, Najlaa JANNAH***[b]***,***
Adel BAHADDAD[c]***, and Abdullah M. BOKHARY***[a]

[a] Department of Software Engineering, College of Computer Science and Engineering, University of Jeddah, Jeddah, Saudi Arabia
[b] Department of Computing, College of Engineering and Computing in Al-Lith, Umm Al-Qura University, Makkah, Saudi Arabia
[c] Department of Information System, Faculty of Computing and Information Technology, King Abdulaziz University, Jeddah, Saudi Arabia

Original scientific paper
https://doi.org/10.2298/TSCI2406059A

Cardiovascular disease is a chronic disease that is a leading cause of death due to heart failure and blood stroke. The WHO records 17.9 million deaths yearly due to heart-related diseases. Heart failure occurs worldwide, especially having a significant impact in low and middle-income countries. Early diagnosis of heart disease is needed because a patient can face serious complexities if it is detected in the later stages of disease progression. In addition, if heart disease is identified early, it is likely to be cured. On the other hand, symptom identification of heart failure is necessary for an accurate and optimum solution. The model reported in this paper suggests a solution for the early diagnosis of heart disease. First, data analysis is performed, and pre-processing approaches are applied to prepare the dataset for model training. Raw data has noise and missing values, which are treated correctly before being passed to the model. Second, two types of algorithms are trained for the proposed solution. Traditional machine learning algorithms are used in the form of support vector machine, k-nearest neighbors, logistic regression, random forest, artificial neural networks, decision tree, xgboost, and catboost to train and test the model. In parallel, automated machine learning (AutoML) with an Azure machine learning cloud instance is used for model training and testing. Azure data lake cloud storage is utilized for model training and running the AutoML process. Finally, the performance of the models was evaluated using a University of California Irvine (UCI) machine learning open-source dataset for heart failure diagnosis. The AutoML outperformed when compared with traditional algorithms. The highest accuracy value obtained for the best machine learning algorithm was xgboost, with an accuracy of 82.22%, whereas the accuracy value obtained using AutoML was 88%. The proposed model can be used for clinical purposes due to its performance and the approach applied.

Key words: *heart failure, cardiovascular disease, Azure machine learning, Azure data lake, machine learning, AutoML*

* Corresponding author, e-mail: amalghamdi@uj.edu.sa

Introduction

Cardiovascular diseases (CVD) are a common and leading cause of death globally. CVD are recorded by the WHO as causing 17.9 million deaths each year worldwide. Heart failure and stroke are common in low and middle income countries due to unhealthy diets and low physical activity. Heart failure, coronary heart disease, cerebrovascular disease, and blood vessel disorders are common causes of CVD. Death commonly occurs in four out of five CVD patients due to heart failure, heart attacks, and blood vessel strokes [1]. One-third of fatalities in CVD patients occur in people under 70. Heart failure and CVD arise from unhealthy diets, smoking, alcohol use, and physical inactivity. Heart failure effects show in individuals in the form of high blood pressure, increased blood glucose, specifically in diabetes patients, and increased weight. If these factors are found earlier, lowering the number of heart failure and stroke cases may be possible. Symptoms of heart failure and strokes are pain in the center of the chest, pain in the arms, and pain in the left shoulder and jaw. Some precautions help save from heart failure and stroke attacks, such as reduced use of tobacco and alcohol, less salt in the diet, a healthy diet, and exercising for a healthy life [2].

Heart failure is an attack on the right or left side of the heart, but sometimes it occurs on both sides. Heart attacks have two forms: the first is acute, known as a short-term attack, and the second is chronic, known as an ongoing condition. Heart failure has been classified into four types: left-side heart failure, right-side heart failure, diastolic and systolic. Left-side heart failure occurs due to reduced left ventricular function. The heart usually pumps, but the bottom of the heart is thicker than normal, so the ventricular cannot relax properly. Right-side heart failure affects the right side of the heart and occurs due to left-side heart failure. A complication of heart failure is an abnormality in a heartbeat, kidney damage, liver damage, and lung fluid collection [3].

Heart failure is a chronic disease that becomes worse with time. Heart failure is divided into four stages: A through D. Heart failure Stage A is the pre-heart failure stage and may occur due to your family history of heart failure. The causes are diabetes, alcohol use, and fever. Stage B is the pre-heart failure in which your doctor diagnoses systolic left ventricular dysfunction, but you do not have heart failure symptoms. Stage C is when the patient has been diagnosed with heart failure and identified symptoms. Symptoms in this stage may be difficulty breathing, weak body, weak legs, and inability to exercise. Finally, in Stage D, people have significant apparent symptoms such as tiredness, shortness of breath, dry cough, and the need to urinate at night [4]. Heart failure can be cured with disease management and proper treatment of symptoms identified early [5].

Heart failure is caused due to reduced immunities, low glucose in the blood, and diabetes. Therefore, there is a need to address heart disease early to cover the symptoms and introduce proper treatment. In addition, early prediction of heart diseases such as heart failure and blood stroke is needed so that patients can be saved with proper care. However, clinical treatment cannot cover many patients presenting with heart diseases. As such, there is a need for automated solutions using other approaches like machine learning and deep learning. Such automated solutions can support the diagnosis of heart disease in more patients earlier.

The main goal of this research was to provide a solution the automated diagnosis of heart-related diseases like heart failure. Machine learning approaches are explored as a solution this problem. This solution can be used as part of early disease diagnosis to allow treatment to begin as soon as possible.

Literature review

This section presents extant solutions and approaches for heart disease diagnosis and analysis. Heart disease is complex and chronic and occurs globally. Therefore, there is a great need for early diagnosis to treat heart diseases such as heart failure before they result in death. Heart failure is not just total heart failure but also causes physiological functionalities to slow down, such as oxygen supply. Some approaches have already been applied to heart failure diagnosis [6]. Heart failure diagnosis approaches using machine learning, deep learning, and data mining are searched in [7]. A few approaches are discussed below, along with the problems they address, proposed solutions, and evaluations.

Accurate and timely identification of heart disease is needed to prevent complexities arising from the disease. Heart failure causes problems in the transport of oxygen within the whole body, which can cause serious issues. Jian *et al.* [8] proposed a system for identifying heart disease using machine learning approaches. They addressed the problem of timely and accurate identification of heart disease to play an important role in electronic healthcare, specifically in cardiology. This system was developed using machine learning classifiers such as support vector machine (SVM), decision tree (DT), ANN, k-nearest neighbor (kNN), naive Bayes, and logistic regression (LR). They also used standard algorithms such as minimal redundancy and maximum relevance, relief algorithm, least absolute shrinkage selection operator, and local learning algorithm. Irrelevant features were removed using these algorithms to save time and space during system execution in training. A model of best practice assessment was applied using leave-one-out cross-validation for hyperparameter training. The proposed system was evaluated using the University of California Irvine (UCI) open-source dataset. Evaluation measures showed promising results, indicating strong system performance. Their SVM-based system (FCMIM-SVM) performed best compared with the other applied approaches. This system can be applied in real-time healthcare for best practices.

Heart diseases are critical to human beings because they cause problems in the blood flowing to the body parts. The shortage of blood in body parts becomes dangerous for human health. Another research study addressed this problem to solve the on-time prediction of early-stage disease [9]. It addressed the issue that, for many reasons, more than a traditional patient medical history was required to provide the best solution for a heart disease diagnosis. The proposed system relies on machine learning approaches to give the best solution for heart disease diagnosis. The proposed system used seven popular machine learning algorithms for the best solution. The proposed system was evaluated using an open-source Cleveland heart disease dataset 2016, and seven evaluation measures, such as accuracy, specificity, sensitivity, Matthews' correlation coefficient, and execution time, were used to calculate their values. Statistical values after an evaluation showed their worth against different approaches. The LR and SVM classifiers outperformed compared to other classifiers, and their accuracy values after feature selection were 89% and 88%, respectively. Therefore, this system can be used in healthcare for commercial purposes after real-time data testing.

The CVD is a chronic disease affecting human health, leading to heart failure, stopping blood circulation, and may cause death. Models were proposed [10] to reduce death rates by supporting an early diagnosis of CVD. The proposed model suggested using several approaches to predict heart failure and heart disease. The proposed model was developed over several steps to manage the dataset, pre-process the dataset, and transform data to build the heart disease prediction system. This model used relief, least absolute shrinkage, and selection operator (LASSO) approaches to select suitable features. They used hybrid classifiers to classify heart disease patients from ordinary people. This model used the DT bagging method, random forest bagging method (RFBM), kNN bagging method, AdaBoost boosting method, and gradient boosting method for the prediction of heart disease and classification between normal and heart disease. The proposed model was evaluated using different evaluation measures on the open-source Cleveland dataset. The accuracy value for the RFBM and relief feature selection was 99.05%, which was best compared with the other machine learning methods.

Heart disease, also known as CVD, is complex due to the sensitivity of the heart. There may be a significant effect on the body before death due to a shortage of blood and oxygen to other body organs. Blockage of the blood in vessels due to the heart not working properly can cause death. Shah *et al.* [11] suggested a heart disease and CVD solution. This research paper addressed the problem of identifying heart disease symptoms early. Traditional approaches to heart disease diagnosis have already been applied using machine learning and data mining. However, this paper addressed heart failure diagnosis using machine learning algorithms such as naive Bayes, DT, kNN, and random forest (RF). These algorithms worked for the diagnosis of heart disease. The proposed model was evaluated using the UCI machine learning open-source dataset, and the evaluated results were high compared to the previous approaches. The kNN approach outperformed when compared to the other machine learning algorithms.

Alotaibi *et al.* [12] introduced an approach to address the significant problem of heart disease prediction. Due to heart failure, a human being is faced with many issues and potential death. Death rates due to heart failure have increased in the last few decades. The proposed model addressed this problem to solve heart failure diagnosis. Machine learning algorithms such as DT, LR, RF, SVM, and naive Bayes were used to predict heart failure in the open-source UCI heart disease dataset. Within the proposed system, the DT classifier outperformed when compared to the other algorithms. The accuracy values using the DT were computed at 93.19%.

Miao *et al.* [13] proposed a solution heart disease to diagnose patients accurately. Deep learning approaches were used to address heart disease problems. The WHO [14] has reported that the number of heart disease patients is increasing globally daily. In 2015, 30% of deaths occurred due to CVD, according to the WHO report. This research study applied enhanced deep neural networks (DNN) for heart disease detection and CVD diagnosis. The proposed system was evaluated using the Cleveland Clinic Foundation dataset. Evaluation measure values were calculated, resulting in an accuracy of 83.67%. The proposed system outperformed using the DNN algorithm for CVD and heart disease diagnosis. This system can be used for clinical purposes after testing on the real-time dataset.

Traditional machine learning approaches were used in the aforementioned methods. Furthermore, Other approaches used deep learning models for heart disease and CVD prediction. Alaa *et al.* [15] proposed a machine learning-driven based technique. This research study focused on heart disease diagnosis using automated machine learning concepts instead of traditional machine learning to get more accurate results. The study also identified how non-traditional variables could add value to the system for performing heart disease diagnosis. This research study was evaluated and outperformed with accuracy values of 77.23%. This system can be used for clinical purposes but has limitations due to the absence of cholesterol biomarkers.

These approaches address heart failure due to global disease and suggest solutions using open-source datasets. Some approaches used the UCI open-source dataset and evaluated their models using precision, accuracy, recall, and F1-measure measures. A few of them outperformed this dataset, but more accurate predictions for early diagnosis of heart disease still need to be made. Table 1 summarizes the state-of-the-art survey, including the approach used, the limitations of the proposed approach, and the advantages of the proposed methods.

Evaluation results were also compared to check their system performance. In this survey, we discovered that COVID-19 could also affect heart functionality. Bader *et al.* [16] discussed how COVID-19 has become more dangerous for patients already suffering from heart disease. Because two leading causes of heart disease are low glucose transformation and lower blood circulation, these also majorly impact COVID-19 patients.

Most approaches in the literature used machine learning for heart diagnosis [17]. This is because machine learning helps to predict heart-related diseases early. Moreover, they gave the best solution compared with traditional clinical methods because there is such a large number of heart failure patients that only automated solutions can diagnose quickly [18].

Table 1. Summary of the state-of-the-art survey

References	Technique	Limitations	Advantages	Acc [%]
[8]	FCMIM-SVM	Identification of heart disease	Timely and accurate heart disease identification	92.37
[9]	Hybrid intelligence system framework using machine learning classifiers	Heart disease prediction and analysis	Better to use machine learning approaches compared with traditional approaches	89.00
[10]	Machine learning algorithms with relief and LASSO feature selection	CVD and heart disease prediction	CVD and heart failure prediction performed with best performance values	99.05
[11]	Machine learning algorithms	Heart disease prediction	Prediction of heart disease and a solution for early diagnosis	90.78
[12]	Machine learning algorithms	Heart failure disease	Accurate detection of heart failure disease	93.19
[13]	DNN	CVD and heart disease diagnosis	CVD and heart disease were diagnosed using deep-learning algorithms	83.67
[15]	AutoML driven approach	Heart disease prediction	Heart disease prediction but the absence of cholesterol biomarkers	77.23

Material and design

This section discusses the dataset used for training and testing the model. The first part discusses the available open-source datasets and the heart failure dataset. The second part highlights the proposed methodology in detail, including how to use it for implementation.

Dataset description

The CVD is a necessary research subject in no small part due to the increasing number of patients worldwide. Heart failure and blood stroke are common forms of CVD. Therefore, there is a need for real-time clinical datasets that can be analyzed and used to evaluate novel ideas for improving the survival rates of patients with heart failure. Real-time datasets that were compiled for analysis from different regions are available online for use in experimental analysis. The National Health Interview Survey provides an open-source dataset for heart failure. The UCI machine learning repository provides a heart failure dataset [19]. The same heart failure open-source dataset is also available on the Kaggle site to use for experiments. The model described in this paper uses this dataset for its experimental analysis.

The dataset used for the experimental analysis has some characteristics. The data is multivariate; there are 299 instances, and the dataset's attributes are integer and real-valued.

There are 13 attributes, with the last attribute being the class or label. This dataset can be used for classification, regression, and clustering purposes. The details of each attribute are discussed in tab. 2.

Table 2. Each feature detail, including measurement and value range

Feature name	Feature detail	Measurement	Value range
Age	Patient age	Years	(40, 95)
Anemia	Red blood cell reduction or hemoglobin	Boolean	0, 1
High blood pressure	Hypertension of the patient	Boolean	0, 1
Creatinine phosphokinase	CPK enzyme level in blood	mcg [L]	(23, 7861)
Diabetes	Diabetes inpatient	Boolean	0, 1
Ejection fraction	Blood leaving percentage	[%]	(14, 80)
Sex	Man or woman	Binary	0, 1
Platelets	Platelets in blood	kilo platelets [mL]	(25.01, 850.00)
Serum creatinine	Creatinine level in the blood	[mg], [dL]	(0.50, 9.40)
Serum sodium	The sodium level in the blood	[mEqL^{-1}]	(114, 148)
Smoking	Patient smoking	Boolean	0, 1
Time	Period follow-up	Days	(4,285)
Target	During the follow-up period of patient death	Boolean	0, 1

The details in tab. 2 are used in the proposed model to analyze and diagnose heart failure in patients. The proposed model uses all the given features, and the last column represents the label or class. The features are integral and real-valued, including numeric and category features.

Proposed methodology

Timely heart failure analysis is a critical need, as evidenced by the applied approaches discussed in the literature review section. Researchers have focussed on machine learning, deep learning, and data mining algorithms to build heart failure and CVD analysis models. However, there is still a need for a solution that focuses on the accurate, early diagnosis of heart failure. The proposed model differs from previous approaches by using machine learning algorithms and AutoML to classify whether data indicates a heart failure patient or an average person. In addition, the Azure machine learning instance is utilized for the data architecture, and AutoML is applied to the given dataset. The details for each implementation step are shown in fig. 1. The implementation of the proposed model is divided into three phases: the data repository phase, the analysis phase, and the application development phase. Each phase is explained further with its supporting steps.

Data repository phase

In this phase, the data source for the training and testing of the model is defined, and an Azure instance for data storage is utilized to apply AutoML. The UCI machine learning open-source heart failure dataset was loaded using azure services for the data architecture to store the data for analysis. The proposed model accepts the dataset as a CSV file format for processing in the training model.

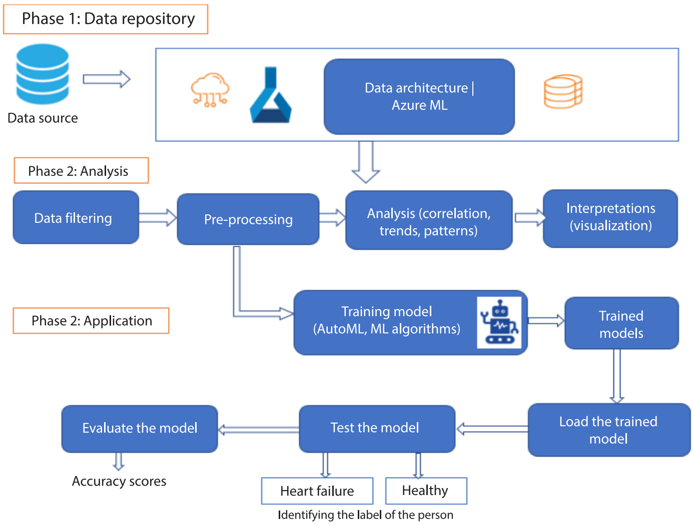

Figure 1. Proposed model methodology structure

Data analysis phase

This phase focused on the pre-analysis of the dataset. Pre-processing approaches are applied to prepare the dataset for model training and testing. We performed a few significant points for analysis. First, data filtering techniques are used to check for missing values in the data. Missing values, noise, and data scalability are addressed in the pre-processing. Next, the whole data is analyzed and examined to determine which techniques were better for this problem. Finally, correlation, checked trends, and data patterns are computed to gain insights about the data. After the data interpretation, it is passed to the model for further execution for training purposes.

Application phase

This phase developed the model across a few primary steps. The model is developed, trained, and tested using unseen data to evaluate the model's performance. The application development phase included five steps to complete the model implementation: model building, model training, export-trained model, model testing, and model evaluation. The details of each step are discussed below.

Model building

This step represents the development of the model. The most crucial point is that the data comes from an open-source dataset. After the model is developed, it moves to the training phase.

Model training

Using AutoML, data was stored in the Azure instance using Microsoft Azure's cloud services. The data was small enough for the machine learning algorithms to train the model on a standalone computer. The machine learning algorithms used were SVM, kNN, LR, RF, ANN, DT, XGboost, and Catboost.

Export trained model

Successfully trained models were moved toward the next steps. Trained models were exported to use at various times for testing, which is required to evaluate the model. The training was executed using Azure machine learning cloud services for AutoML, and the traditional algorithm models were stored standalone.

Testing model

The trained model was used to test the proposed model, and there is no need to train it again as it was just used in the testing module. Additionally, the label of a person is identified as someone who has or does not have heart failure. The dataset used for testing was labeled as heart failure or average person, allowing accuracy to be computed.

Model evaluation

The model was evaluated based on the training model. The evaluation measure calculated the values and proved whether the model was performing best based on the model evaluation's accuracy.

Accuracy is the evaluation measure that tests the ability of the model to correctly differentiate between a patient with heart failure and a healthy person. Accuracy is calculated as the ratio of actual results (true positive and true negative) to all results (including false positives and false negatives). The mathematical representation of the accuracy is shown:

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \tag{1}$$

The model was evaluated for both the traditional machine learning algorithms and AutoML using Azure machine learning cloud. The accuracy values are represented in [%] for each test case.

Azure Data Lake

Azure Data Lake is a repository for storing data in many formats. This data is usually stored in the form of files or blobs. The single data store, which includes the original copies of data sources, represents it. The data is acquired from any source, such as sensors, social, and transformed data, and used for machine learning, visualization, analysis, and reporting. Data Lake can contain structured data (relational databases that are rows and columns), unstructured data (PDF, documents, and emails), and semi-structured data (JSON, XML, HTML). It also has a binary form of data, such as images, audio, or video. Typically, Data lakes can be established in the cloud or data centers. Azure Data Lake stores the data in the cloud for model training and testing.

The automated machine learning

The AutoML is the process that automatically produces the test-set prediction. No human intervention is required to set the machine learning algorithms. This approach is an advanced form of machine learning but makes the model easy to use for larger or multiple models quickly without any extra human experts. In AutoML, we did not select the specific algorithm, but the model automatically determines the required machine learning algorithm based on the nature of the problem. The Azure machine learning cloud instance stores the dataset and uses the built-in feature to use AutoML. The AutoML helps to compute the required machine learning algorithm and saves time when using any traditional machine learning algorithm. AutoML can pre-process the data and select and create the appropriate features. The model selection helps to optimize the model's hyperparameters. Deep learning helps design the neural network's topology and post-process machine learning models.

Finally, the results obtained from these models were critically analyzed. AutoML, using the Azure cloud instance, uses the same dataset as traditional machine learning algorithms. The experimental analysis is discussed in the result section below.

Machine learning algorithms

Models are developed using different classifiers: SVM, kNN, LR, RF, ANN, DT, XG-boost, and Catboost. All these algorithms were trained and tested with the unseen dataset. Evaluation results are shown separately in the results section for each algorithm.

Result analysis and discussion

The proposed model was developed using the steps described previously and illustrated in fig. 1. We first developed models using traditional machine learning algorithms and then created a model using Azure machine learning for the AutoML implementation. When using machine learning algorithms, we stored the dataset standalone, but when we used AutoML, the dataset was stored in the cloud in the data architecture. Below, we discuss the results of using machine learning algorithms.

Analysis of the dataset revealed attributes of two types: integer and float. The distribution of datatypes in the dataset: 76.9% int64 and 23.1% float64. Analysis of the dataset was required to apply the most appropriate model. A complete study was conducted on each attribute, and death events were examined for correlating qualities such as age, sex, and diabetes.

After analyzing the dataset, the next step was to build the model. However, less valuable features were recognized with almost zero impact before developing the model, so we dropped them. The features with values practically equal to 0 were anemia, creatinine phosphokinase, diabetes, high pressure, platelets, sex, and smoking. This reduced computational requirements after finalizing the model because these features had no significant impact. Next, three features were selected that significantly impacted decision-making: time, section fraction, and serum creatinine. After that, the dataset is split for training and scaling the features to balance the labeled dataset. As discussed, machine learning algorithms and AutoML in the Azure cloud instance are used for classification.

Results using automated machine learning

The AutoML is a machine learning model. All the required services are used from the Azure cloud. The dataset is stored on Azure Data Lake, and the model is trained there. A dashboard of the Azure platform using AutoML is shown in fig. 2, where we can see the parameter set we require for the model.

The results are shown in tab. 3 using Microsoft Azure machine learning, where different machine learning algorithms were selected automatically. A random sampling policy was selected for the parameter sampling. SparseNormalizer and XGboost classifiers had the best scores compared with the other chosen algorithms. After completing 40 iterations, the best accuracy score of 0.88 was obtained.

Results using machine learning algorithms

The model was developed using traditional machine learning algorithms SVM, kNN, LR, RF, ANN, DT, XGboost, and Catboost. Evaluation results for each algorithm, as computed on the test data, are shown in tab. 4. These algorithms were trained independently and tested on an unknown dataset. For each algorithm, accuracy values were computed. Under the same circumstances, the XGboost algorithm outperformed the other machine learning algorithms.

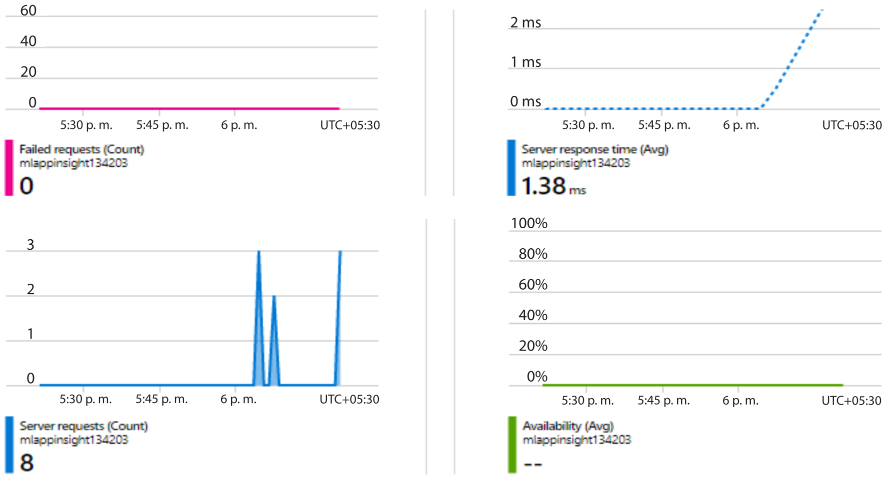

Figure 2. Application dashboard using Azure AutoML

Table 3. Each feature detail, including measurement and value range

Algorithm name	Accuracy
SparseNormalizer, XGboost Classifier	0.84
Sparse Normalizer, LightGBM	0.83
SparseNormalizer, XGboost Classifier	0.83
MaxAbsScater, LightGBM	0.82
AutoML	0.88

Table 4. Evaluated results from traditional algorithm

Algorithm name	Accuracy [%]
SVM	75.56
KNN	77.78
LR	81.11
RF	81.11
ANN	74.44
DT	78.89
XGboost	82.22
Catboost	80.00

After analysis, it can be seen that the statistical results for the AutoML model were best when compared with the traditional machine learning algorithms. The AutoML works based on auto hyperparameter tuning, which is performed automatically. In the conventional approach, XGboost performs better than the other machine learning algorithms. Overall, AutoML results were better than those of different techniques.

Conclusions

The CVD are a common and leading cause of death globally. The WHO has recorded 17.9 million deaths per year. Therefore, there is a need to address heart disease to identify symptoms and introduce proper treatment. Early prediction of heart diseases is vital so that heart failure and blood-stroke risks can be reduced with proper care. The proposed model suggests a solution for early diagnosis and prevention of heart disease. We first analyzed data and applied pre-processing approaches to prepare the dataset for model training. We used traditional machine learning algorithms SVM, kNN, LR, RF, ANN, DT, XGboost, and Catboost to train and test the model on an unseen dataset. In parallel, we used AutoML with the Azure machine learning cloud instance for model training and testing. The AutoML approach outperformed the traditional algorithms.

In the future, large-volume datasets can be used with many features to examine all possibilities. In addition, deep learning approaches can be used to diagnose and analyze heart disease. Heart diseases are dangerous to health due to their nature, so addressing them as early as possible is necessary.

Acknowledgment

This work was funded by the University of Jeddah, Jeddah, Saudi Arabia, under Grant No. (UJ-24-DR-04101247-1). Therefore, the authors thank the University of Jeddah for its technical and financial support.

References

[1] Banerjee, A., Mendis, S., Heart Failure: The Need for Global Health Perspective, *Current Cardiology Reviews*, *9* (2013), 97
[2] Khan, D. T., Cardiovascular Diseases, World Health Organization, https://www.who.int/health-topics/cardiovascular-diseases#tab=tab_1, 2022
[3] Hwang, I.-C., *et al*., Different Effects of SGLT2 Inhibitors According to the Presence and Types of Heart Failure in Type 2 Diabetic Patients, *Cardiovascular Diabetology*, *19* (2020), May, pp. 1-12
[4] Vasan, R. S., *et al*., Pidemiology of Heart Failure Stages in Middle-Aged Black People in the Community: Prevalence and Prognosis in the Atherosclerosis Risk in Communities Study, *Journal of the American Heart Association*, *10* (2021), e016524
[5] Rossignol, P., *et al.*, Heart Failure Drug Treatment, *The Lancet*, *393* (2019), Mar., pp. 1034-1044
[6] Iung, B., *et al.*, Contemporary Presentation and Management of Valvular Heart Disease: The EURObservational Research Programme Valvular Heart Disease II Survey, *Circulation*, *140* (2019), 14, pp. 1156-1169
[7] Ahsan, M. M., Siddique, Z., Machine Learning-Based Heart Disease Diagnosis: A Systematic Literature Review, *Artificial Intelligence in Medicine*, *14* (2022), 102289
[8] Jian, J. P., *et al*., Heart Disease Identification Method Using Machine Learning Classification in e-Healthcare, *IEEE Access*, *8* (2020), June, pp. 107562-107582
[9] Haq, A. U., *et al*., A hybrid Intelligent System Framework for the Prediction of Heart Disease Using Machine Learning Algorithms, *Mobile Information Systems*, *2018* (2018), 386046
[10] Ghosh, P., *et al*., Efficient Prediction of Cardiovascular Disease Using Machine Learning Algorithms with Relief and LASSO Feature Selection Techniques, *IEEE Access*, *9* (2021), Jan., pp. 19304-19326
[11] Shah, D., Heart Disease Prediction Using Machine Learning Techniques, *SN Computer Science*, *1* (2020), 345
[12] Alotaibi, F. S., Implementation of Machine Learning Model to Predict Heart Failure Disease, *International Journal of Advanced Computer Science and Applications*, *10* (2019), 6, pp. 261-268
[13] Miao, K. H., Miao, J. H., Coronary Heart Disease Diagnosis Using Deep Neural Networks, *International Journal of Advanced Computer Science and Applications*, *9* (2018), 10, pp. 1-8
[14] Nowbar, A. N., *et al.*, Mortality from Ischemic Heart Disease: Analysis of Data from the World Health Organization and Coronary Artery Disease Risk Factors from NCD Risk Factor Collaboration, Circulation: *Cardiovascular Quality and Outcomes*, *12* (2019), e005375
[15] Alaa, A. M., *et al*., Cardiovascular Disease Risk Prediction Using Automated Machine Learning: A Prospective Study of 423,604 UK Biobank Participants, PloS one, *14* (2019), e0213653
[16] Bader, F., *et al*., Heart failure and COVID-19, *Heart Failure Reviews*, *26* (2021), pp. 1-10
[17] Ramalingam, V., *et al*., Heart Disease Prediction Using Machine Learning Techniques: A Survey, *International Journal of Engineering and Technology*, *7* (2018), 2.8, pp. 684-687
[18] Ahsan, M. M., Siddique, Z., Machine Learning-Based Heart Disease Diagnosis: A Systematic Literature Review, *Artificial Intelligence in Medicine*, *19* (2022), 102289
[19] Chicco, D., Jurman, G., Machine Learning Can Predict Survival of Patients with Heart Failure from Serum Creatinine and Ejection Fraction Alone, *BMC Medical Informatics and Decision Making*, *20* (2020), Feb., pp. 1-16

Paper submitted: July 14, 2024
Paper revised: October 17, 2024
Paper accepted: October 21, 2024