I.J. Intelligent Systems and Applications, **2024, 3, 35-46**
Published Online on June 8, 2024 by MECS Press (http://www.mecs-press.org/)
DOI: 10.5815/ijisa.2024.03.03

A Proposed Stacked Machine Learning Model to Predict the Survival of a Patient with Heart Failure

Md. Raihan Mahmud*
American International University-Bangladesh, 408/1 (Old KA 66/1), Kuratoli, Khilkhet, Dhaka 1229, Bangladesh
E-mail: raihan.mahmud514@gmail.com
ORCID iD: https://orcid.org/0009-0004-5936-5079
*Corresponding author

Dip Nandi
American International University-Bangladesh, 408/1 (Old KA 66/1), Kuratoli, Khilkhet, Dhaka 1229, Bangladesh
E-mail: dip.nandi@aiub.edu
ORCID iD: https://orcid.org/0000-0002-9019-9740

Md. Shamsur Rahim
School of Computer Science, University of Technology Sydney, Australia
E-mail: mdshamsur.rahim@uts.edu.au
ORCID iD: https://orcid.org/0000-0002-7721-079X

Christe Antora Chowdhury
Popular Medical College, 25 Road no 2, Dhanmondi, Dhaka 1205, Bangladesh
E-mail: c.nandi006@gmail.com
ORCID iD: https://orcid.org/0009-0003-8637-4352

Received: 02 August 2023; Revised: 05 January 2024; Accepted: 01 March 2024; Published: 08 June 2024

Abstract: Now a days heart failure is one of the most common chronic diseases that cause death. As it possesses high risk of death, it is important to predict patient's survival and optimize treatment strategies. Machine learning techniques have come to light as useful tools for evaluating enormous quantities of patient data and deriving important patterns and insights in recent years. The purpose of the study is to investigate the feasibility of using the machine learning methods for predicting heart failure patient's chances of survival. We have worked on a dataset with 2029 heart failure patients and the dataset comprises 13 features. To conduct this research, we suggested a model (Stacked machine learning model using scikit-learn using Decision Tree, Naive Bias, Random Forest, Linear Regression, SVM, XGBoost, ANN) using which we got better results than previously existed researches. We believe the suggested model will help advance our understanding of heart attack prediction.

Index Terms: Machine Learning, Heart Failure, Stacked Machine Learning Model, Scikit-learn.

1. Introduction

Heart failure is a potentially fatal state that affects masses of people around the globe. Several variables, such as diabetes and high blood pressure, increase the risk of heart failure. Numerous symptoms, including weariness, limb oedema, and shortness of breath, may result from this. For treatment plans to be optimized and patient outcomes to be improved, accurate and on-time estimation of a patient being survived in heart attack situations is essential. In recent years, different machine learning techniques have been employed to predict a patient's survival after a heart attack. To predict heart failure using machine learning, it is necessary to collect significant patient information and medical record datasets and train supervised machine learning algorithms such as neural networks, random forests, or logistic regression on the input data.

Machine learning approaches extract significant insights from massive and multiple data sets. Machine learning techniques can identify hidden patterns and associations that may predict a patient's likelihood of survival by exploiting the enormous quantity of patient-related data, including medical imaging results, demographic data, clinical documents, and diagnostic tests.

Because of the potential impact on patient health outcomes, detecting heart failure through machine learning techniques is essential. Rapid identification and risk classification can result in immediate interventions, enhancing the prognosis for at-risk individuals. Personalized medicine is possible by adapting therapies to individual patient features, optimizing effectiveness, and avoiding side effects. Detecting people at high risk who need specialized care can help maximize the distribution of healthcare resources. The combination of enhanced care for patients, the resource allocation, and the advancement of science make heart failure prediction employing machine learning necessary. Various machine learning approaches can be used to estimate the chance of survival for those with heart failure. Decision tree, random forest, and SVM are some of the most common strategies. These methods can be used to determine which factors are most likely to predict a patient's survival.

Because of the disease's complexity, it can be difficult to predict heart failure with machine learning techniques. The challenge is exacerbated by various patient's multiple underlying causes, demographics and varying signs and symptoms. Obtaining and combining extensive and high-quality datasets, such as genetic information, medical records, as well as lifestyle factors, is also difficult. Furthermore, the comprehensibility of machine learning algorithms in the framework of heart failure prediction remains an issue, as understanding the reasoning behind the predictions is critical. Some models, such as naïve methods, fail to predict because of the oversimplification of the disease's complexity. They frequently overlook critical variables and interactions, such as the interaction of hereditary factors, lifestyle decisions, and medical history. Naive techniques can also not handle and analyze extensive and diverse datasets, resulting in low accuracy and predictive potential.

Several fundamental elements contribute to improved machine learning (ML) accuracy. First and foremost, high-quality datasets are essential for training ML models. Clean, representative data reduces bias and increases the model's generalizability. Second, feature engineering is required to pick and transform relevant features efficiently. Domain expertise and data comprehension help in selecting features and extracting them. Finally, choosing a model and hyperparameter tuning guarantee that the ML model suits the task and that its parameters are adjusted for maximum performance.

Explainable AI refers to the capability to comprehend and analyze AI judgments, which is critical for enhancing trust in the model. Explainability is difficult to achieve, especially with complicated machine learning (ML) algorithms such as neural networks. These models operate like black boxes, making it challenging to grasp the connection between inputs and outputs. Simpler models, such as decision trees or linear models, can be used to achieve explainability, feature importance analyses, surrogate models, and local interpretation approaches. Finding the optimal trade-off between accuracy and interpretability remains an important field of AI research.

Several papers attempted to solve this issue by applying machine learning algorithms to the dataset. Due to a lack of a quality datasets and the inability to select the correct machine learning algorithm, no other researchers could achieve higher accuracy than 93.19, nor did they employ explainable AI in their research. In this study, we applied machine learning techniques to a dataset of heart failure patients collected from UC Irvine Machine Learning Repository. We have considered the dataset because of the accuracy, completeness, consistency, reliability, low noise, data preprocessing etc parameters were present in the dataset. We have identified the factors most likely to predict a patient's survival using the stacked machine learning model with sci-kit-learn, decision tree, naive bias, random forest, linear regression, SVM, XGBoost, and ANN techniques. We have then used explainable AI to determine why the findings are accurate. The results contribute to improved clinical decision-making, enhanced patient care, and a more efficient allocation of healthcare resources to treat heart failure.

This paper illustrates the possibility of using the machine learning methods for predicting heart failure patient’s chances of survival. The introduction highlights the significance of the research as well as the lacking of the existing research. Following this, the literature review delves into the existing researches related to the topic. The methodology outlines the approach of selecting research papers, data collection, data preprocessing and research questions. Subsequently, the results section discusses about the results and findings including the explainable AI and the finally the conclusion encapsulates the summary of the whole research including the impact of the research in the medica sector.

2. Literature Review

Several research papers have been conducted on using machine learning algorithms to detect the prognosis of heart failure patients. This section examines some of the most influential works in this field, focusing on the machine learning techniques used.

Pedro A. Moreno-Sanchez et al. aimed to boost clinical prediction system adoption among doctors in order for them to better understand the justification for black-box AI medical solutions while making more informed decisions based on data. They improved a prediction algorithm for heart failure survival using explainable artificial intelligence in 2021. They presented modeling and analysis, as well as a review of a heart failure survival detection model based on an accuracy-explainability ratio, in their study [1]. The model uses a data processing pipeline and posthoc approaches to identify the best ensemble tree method and feature selection strategy. The best-balanced explainable detection model employed an Extra Trees classifier on 5 of 12 factors (follow-up time, age, serum creatinine, diabetes, and ejection fraction), with adaptive accuracy of 85.1% and 79.5%, respectively. The follow-up time was the most important factor, next to ejection fraction and serum creatinine.

To predict the survival of heart failure patients, in 2019, John Mbotwa et al., Marc de Kamps et al., Paul D. Baxter et al., and Mark S. Gilthorpe et al. suggested a latent class regression strategy using the Cox model. The authors of this study [2] investigated modeling the survival of CHF patients using a Cox proportional hazard (PH) framework within a latent class framework. Using the available covariables, they determined patient types based on risk. They selected the optimal number of classes (BIC) depending on the supplied Bayesian information criteria. They measured the discriminative power of the selected model using AUC. They carried out a simulation experiment to evaluate the precision of their models' predictions. The performance of their suggested latent class model exceeds that of the standard one-class Cox PH model.

To analyze and predict cardiac stroke using ejection fraction and serum creatinine, Md Ershadul Haque et al., Salah Uddin et al., Md Ariful Islam et al., Amira Khanom et al., Abdulla Suman et al., Manoranjan Paul et al., and others used LSTM deep learning technique in 2022 [3]. The writers looked at a set of medical records for 299 people with heart failure from the Faisalabad Institute of Cardiology and Allied Hospital in Pakistan. The heart failure collection had thirteen characteristics that looked at physical traits, eating habits, and medical information. They found a growing trend in their data that will help to learn more about how to tell if someone has heart failure.

In 2018, another research on data mining methods for heart disease prediction was conducted by David H et al., Benjamin & Belcy et al., and S. et al.. Finding the classification algorithm that will deliver the maximum level of accuracy when identifying normal and abnormal people was the main objective of this research's efforts [4]. It was therefore possible to halt the loss of life at an early stage. The experimental environment for assessing the efficiency of algorithms was created using a standard data set for cardiac disease, which was collected from the UC Irvine Machine Learning Library. The Random Forest technique had the highest accuracy of 81% when compared to other machine learning models for predicting heart disease.

Xi Yang, Yan Gong, Nida Waheed, Keith March, Jiang Bian, William R. Hogan, and Yonghui Wu employed machine learning to evaluate cancer patients' heart failure risk [5]. Researchers tested four machine learning approaches on 143,199 UF Health Integrated Data Repository (IDR) cancer patients. The gradient boosting (GB) model had the highest AUC, sensitivity of 0.8520, and specificity of 0.8138. Cancer patients receiving chemotherapy exhibited lower specificity (0.7089). The investigation showed that machine learning can identify clinical features linked to heart failure and cancer patients at risk of heart failure due to cancer treatment.

Another study [6] was carried out in 2016 to identify people with active heart failure. Natural language processing was employed in this study by Shu Dong et al., R. Kannan Mutharasan et al., and Siddhartha Jonnalagadda et al. In this study, the authors proposed two approaches for using electronic health records (EHR) to find occurrences of heart failure that are now occurring. According to the degree of heart failure, the rule-based technique extracted cardiovascular data items from patient records and allocated people to various colors. The machine learning technique used four different models, with SVM with a linear kernel achieving the best outcome with 87.5% accuracy and a 0.86 F1-Score. Combining machine learning and rule-based approaches will enable hospital-wide monitoring of active heart failure with improved accuracy and interpretability of the results.

Researchers M.Akhil Jabbar et al., B. L. Deekshatulu et al., Priti Chandra et al. performed a study [7] on a genetic algorithm and associative categorization to detect cardiac disease in 2013. In this study, a genetic approach-based quick associative classification method for heart disease prediction was proposed. The main benefit of using evolutionary algorithms to create high-level forecasting criteria is that the rules they uncover are highly interpretable, highly predictively accurate, and highly intriguing. According to the trial results, the majority of classifier algorithms provide the best heart disease prediction, which helps doctors make treatment decisions.

The relationship between tumor severity (as defined by TNM staging) and survival in patients with TNM stages 2 and 3 was examined by Chris Roadknight et al., Durga Sury et al., Uwe Aickelin et al., John Scholefield et al., and Lindy G. Durrant et al in 2015 [8]. With a variety of selection criteria and straightforward classification algorithms, it used machine learning techniques to more accurately predict survival rates. Once agreement was found between several models, the performance of each model was next assessed using smaller sets of data. This innovative method of selective ensembling revealed that patients who achieve model agreement may experience significant improvements in prediction performance on an unknown test set. And finally, a strategy for figuring out whether or not a patient's prognosis can be accurately predicted was offered.

According to a study [9] a stroke prediction analytics strategy employing machine learning and neural networks was recommended, which was conducted by Soumyabrata Dev et al., Hewei Wang et al., Chidozie Shamrock Nwosu, Nishtha Jain et al., Bharadwaj Veeravalli et al., and Deepu John et al. This study systematically looked at the different facets of electronic medical data for effective stroke prediction in 2022. The most important signs for identifying stroke in patients are cardiac issues, age, hypertension, and a normal blood glucose level. A perceptron-based neural network with these four characteristics achieves the highest accuracy and least failure rate compared to all available input characteristics and multiple benchmarking strategies. Because the sample was heavily skewed in terms of the frequency of stroke, a balanced dataset was constructed using subsampling techniques.

Yilin Yin et al. and Chun-An Chou et al. predicted early ICU mortality in 2021 and conducted a survival analysis for respiratory failure in a separate study.This work [10] provided a dynamic modeling strategy for the early risk of death prediction for respiratory failure patients based on the first twenty-four-hour period of ICU physiological data. This method showed a decent AUROC efficacy (80–83%) and raised AUCPR by 4% on Day 5 after ICU admission. For the early ICU admission mortality study, the survival curve incorporates time-varying information.

These related works illustrated the application of various machine learning techniques, such as machine learning algorithms and interpretable models, to predict the heart failure patient's survival. Additionally, the studies displayed that some of the researches were able to obtain good accuracy but on the other hand, despite of using different methods, most of the studies did not get the exceptional accuracy. Also, experimental AI was not implemented in any reseaches.

3. Methodology

Here we are going to describe the methodology to predict the survival of a heart failure patient by using machine learning techniques.

3.1. Primary Studies Selection

Searching is essential for finding all relevant information on a given topic. We looked at two search approaches: the first is a manual search of specific journal papers and conference papers using different search engines. The second is keyword searches, a systematic way to find and gather articles from electronic databases. The research papers were chosen because they had previously been utilized as sources for prior reviews relevant to this field and were known to provide either empirical research or literature surveys. To begin the word search method, we picked digital libraries by reviewing past research, and we evaluated their most commonly used databases for the study. We retrieved keywords from relevant publication's titles, abstracts, and keywords, including alternative phrases and synonyms, and used Boolean "AND" and "OR" expressions to connect them. We came up with the following keyword search string:

- Data Mining OR Machine learning Techniques to predict the survival of a patient with heart failure

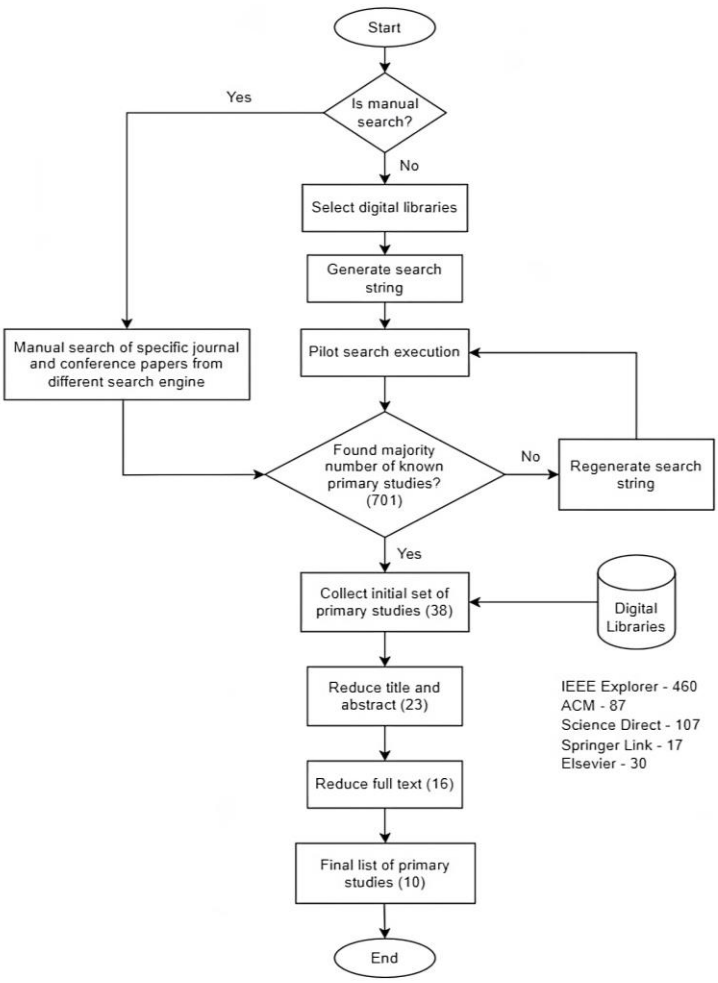

Fig.1. Search process and primary study selection process

As illustrated in Fig.1, the search approach yielded a starting set of 38 primary studies. However, the list may include some papers that do not add value to the research or do not fit the review's objectives. From this vantage point, we had to create the inclusion-exclusion criteria for the initial research so that they avoided studies that did not align with the research aims.

Inclusion Criteria:

- Studies that focus on predicting the survival of a patient with heart failure using Data Mining or Machine learning techniques are included.
- The study's conclusion must include information about the survival of a patient with heart failure.

Reduction Criteria:

- Research that did not consider the survival of heart failure patients.
- Non-English-language studies.
- Papers that are shorter than three pages in length
- Research that lacks empirical analysis and precise data on experimental results

The exclusion process is depicted in Fig.1 in two stages reducing studies by title and abstract and the full text of the studies. 38 primary studies were analyzed based on their titles and abstracts; 23 articles were identified. The whole text of each of these 23 articles was then evaluated. The full-text analysis procedure further reduces the list into 16 primary studies and finally results in a final list of 10 primary studies.

3.2. Data Collection

For research, data collection is one of the most important things, and it is one of the preliminary jobs that need to be done at the beginning of research. To conduct this research, we have collected a structured dataset from UCI Machine learning Repository in CSV format. This dataset featured the medical information of 299 heart failure patients collected over their follow-up period, with each patient profile containing 13 clinical variables.

3.3. Description of the Attributes

There are 13 clinical factors in the dataset. Table 1 displays the details of these characteristics.

Table 1. Description of attributes

Attribute	Description	Range	Example
age	the patient's age (in years).	40-95	75
anaemia	red blood cell or hemoglobin decrease (Boolean)	0, 1	0
high blood pressure	if the patient possesses high blood pressure (Boolean)	0, 1	1
creatinine phosphokinase (CPK)	Blood CPK enzyme concentration (mcg/L).	23-7861	146
diabetes	if the patient suffers from diabetes (Boolean)	0, 1	0
ejection fraction	the proportion of blood exiting the heart with each contraction (percentage)	14-80	38
Platelets	Blood platelet density (kiloplatelets/mL)	25.01-850.00	263358.03
sex	(Binary) Woman or Man	0, 1	1
serum creatinine	serum creatinine concentration (mg/dL)	0.50-9.40	1.2
serum sodium	Blood serum sodium concentration (mEq/L)	114-148	116
smoking	whether or not the patient smokes (Boolean)	0, 1	0
time	Follow-up time (in days)	4-285	10
death event	if the patient died during the observation period (Boolean).	0, 1	1

Here in Table 1, mcg/L: micrograms per liter. mL: microliter. mEq/L: milliequivalents per liter.
Here “death event” is the targeted attribute.

3.4. Data Preprocessing

In activities involving data analysis and machine learning, preprocessing of the data is essential. It entails cleaning up and converting raw data into a format appropriate for additional analysis and modeling. Data preparation refers to various methods for handling missing values, dealing with outliers, standardizing variables, and lowering noise. The objective is to improve the precision, efficacy, and efficiency of following modeling or data analysis procedures. After collecting data from the UCI Machine learning Repository, we manually transformed the data, and to confirm its quality, we randomly cross-checked a number of instances. the dataset, collected from the UCI Machine learning Repository, had 13 attributes and 299 records.

At first, we read the CSV file, which consists of the dataset, and then converted the numbers in a column (or a set of columns) to strings. We then used a wildcard pattern for the targeted column (DEATH_EVENT) and replaced “No” with “0”. We again used a wildcard pattern for the targeted column (DEATH_EVENT) and replaced “Yes” with “1”. We used SMOTE twice to oversample the dataset. After performing SMOTE for the first time, we gathered 406 rows of data. To get better accuracy, we performed SMOTE, and finally, for the second time, we gathered 2029 records. After that, we changed the age, anaemia, creatinine_phosphokinase, diabetes, ejection_fraction, high_blood_pressure, sex, smoking, and time columns from double numbers to integers and wrote the CSV file. Then we performed numeric outliers to detect and treat the outliers for all the columns. We partitioned the data and selected a relative 70%, which means 70% of the data is in the first partition. We also selected to use a random seed where we entered 1685150517569 as a fixed seed to get reproducible results upon re-execution. We induced a classification decision tree in the main memory. Afterwards, we used a decision tree predictor to predict the class value for new patterns. Finally, we used the Scorer node to compare two columns, where we selected DEATH_EVENT as the first column and Prediction (DEATH_EVENT) as the second column.

This research was conducted by answering the following questions:

Table 2. Research questions

RQ No.	Question
RQ: 1	What is the result of the previous research?
RQ: 2	What accuracy can the suggested model provide?
RQ: 3	Is it possible to introduce explainable AI?

Fig.2 is a screenshot of the KNIME workflow of the experiment.

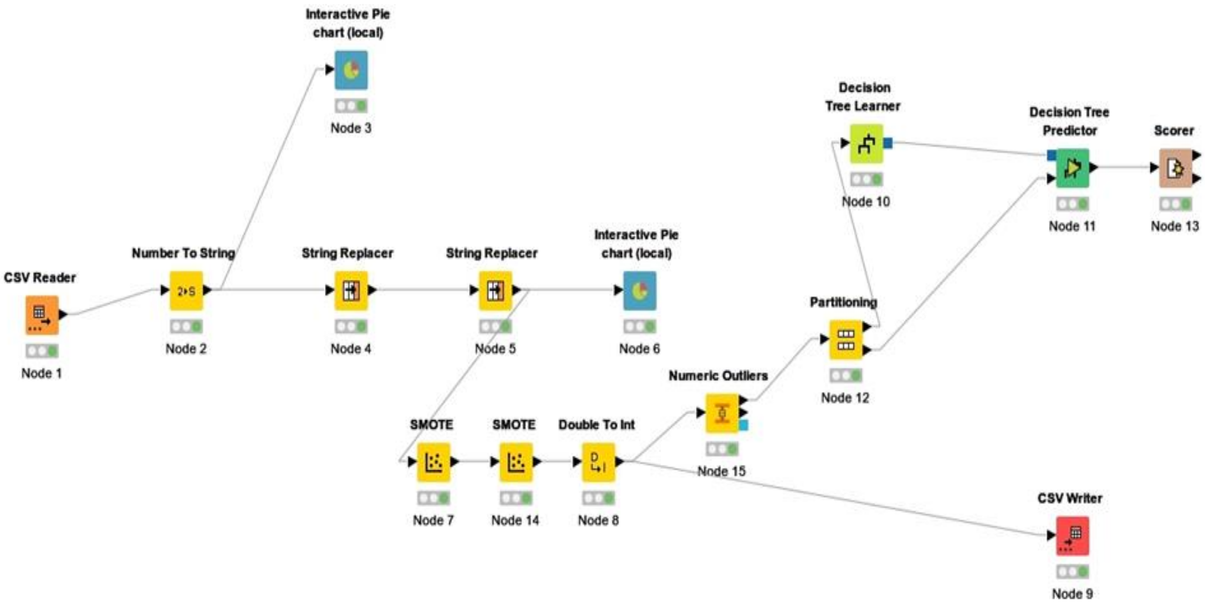

Fig.2. KNIME workflow

4. Results

We designed this research with three research questions. To answer RQ: 1, at first, we reviewed different related studies [11-17]. Then we list down some relevant findings mentioned in those studies in tabular form with their existing solutions.

To answer RQ: 2, we found out the Accuracy, MCC, Precision, Recall, F-Measure and AUC of different models such as Decision Tree, Naive Bias, Random Forest, SVM, XGBoost, ANN, Voting Classifier using scikit-learn with Decision Tree, Naive Bayes, Random Forest, SVM, XGBoost, and ANN, along with the proposed model Stacked machine learning model using scikit-learn using Decision Tree, Naive Bias, Random Forest, Linear Regression, SVM, XGBoost, ANN.

Table 3. Result of the previous research

Paper	Model name	Accuracy	MCC	Precision	Recall	F-Measure	AUC
[11]	RF (Random forests)	0.74	0.384	-	0.491	0.547	0.8
	DT (Decision tree)	0.737	0.376	-	0.532	0.554	0.681
	GB (Gradient boosting)	0.738	0.367	-	0.477	0.527	0.754
	LR (Linear regression)	0.73	0.332	-	0.394	0.475	0.643
	OR (One rule)	0.729	0.319	-	0.383	0.465	0.637
	ANN (Artificial neural network)	0.68	0.262	-	0.428	0.483	0.559
	NB (Naïve Bayes)	0.696	0.224	-	0.279	0.364	0.589
	RSVM (Radial SVM)	0.69	0.159	-	0.122	0.182	0.749
	LSVM (Linear SVM)	0.684	0.107	-	0.072	0.115	0.754
	KNN (k-nearest neighbours)	0.624	-0.025	-	0.121	0.148	0.493
[12]	GB (Gradient boosting)	0.78	0.55	-	0.85	0.81	-
	DT (Decision tree)	0.77	0.53	-	0.88	0.82	-
	KNN (k-nearest neighbours)	0.8	0.48	-	0.92	0.87	-
	LSVM (Linear SVM)	0.76	0.41	-	0.86	0.82	-
	RSVM (Radial SVM)	0.67	0.29	-	0.86	0.75	-
[13]	RF (Random Forest) classifier	0.82	0.64	-	0.75	0.8	-
	DT (Decision tree)	0.79	0.59	-	0.72	0.77	-
	DT (Decision tree (applied only to lung side & platelet count))	0.68	0.41	-	0.58	0.63	-
	Perceptron	0.62	0.23	-	0.95	0.71	-
	OR (One rule)	0.57	0.15	-	0.47	0.55	-
	PNN (Probabilistic neural network)	0.53	0.1	-	0.5	0.5	-
[14]	J48	56.76	-	-	-	-	-
	LMT (Logistic Model Tree)	55.77	-	-	-	-	-
[15]	OR (One rule)	0.73	-	-	-	0.83	-
	RF (Random Forest) classifier	0.75	-	-	-	0.82	-
	SVM	0.71	-	-	-	0.82	-
	Multi-Layer Perceptron	0.78	-	-	-	0.84	-
	NB (Naïve Bayes)	0.70	-	-	-	0.81	-
[16]	DT (Decision tree)	0.8778	-	0.83	0.83	0.83	-
	AdaBoost	0.8852	-	0.89	0.89	0.89	-
	LR (Linear regression)	0.8442	-	0.84	0.84	0.85	-
	SGD (Stochastic Gradient Descent)	0.5491	-	0.54	0.54	0.53	-
	RF (Random Forest) classifier	0.9188	-	0.92	0.92	0.92	-
	GBM (Gradient Boosting Machine)	0.8852	-	0.84	0.84	0.84	-
	Extra Tree Classifier	0.9262	-	0.93	0.93	0.93	-
	Gaussian NB (Naive Bayes) classifier	0.7540	-	0.75	0.75	0.75	-
	SVM (Support Vector Machine)	0.7622	-	0.76	0.76	0.76	-
[17]	DT (Decision tree)	93.19	-	-	-	-	-
	LR (Linear regression)	87.36	-	-	-	-	-
	RF (Random Forest) classifier	89.14	-	-	-	-	-
	NB (Naïve Bayes)	87.27	-	-	-	-	-
	SVM (Support Vector Machine)	92.30	-	-	-	-	-

At first, we performed all the algorithms on the imbalanced dataset and found the results as follows:

Table 4. Result using imbalanced dataset

Model name	Accuracy	MCC	Precision	Recall	F-Measure	AUC
DT (Decision tree)	0.6666666667	0.3012272126	0.619047619	0.52	0.5652173913	0.6457142857
NB (Naïve Bayes)	0.7333333333	0.4732438206	0.9090909091	0.4	0.5555555556	0.84
RF (Random Forest) classifier	0.7166666667	0.4129664151	0.7857142857	0.44	0.5641025641	0.8291428571
SVM (Support Vector Machine)	0.5833333333	0	0	0	0	0.4777142857
XGBoost	0.7333333333	0.442981195	0.7142857143	0.6	0.652173913	0.8697142857
ANN (Artificial neural network)	0.75	0.4841648319	0.8125	0.52	0.6341463415	0.8182857143
Voting Classifier using scikit-learn with Decision Tree, Naive Bayes, Random Forest, SVM, XGBoost, and ANN	0.7333333333	0.4420995869	0.7368421053	0.56	0.6363636364	0.8422857143
Suggested Model (Stacked ml model using scikit-learn using Decision Tree, Naive Bias, Random Forest, Linear Regression, SVM, XGBoost, ANN)	0.75	0.5070925528	0.9166666667	0.44	0.5945945946	0.8731428571

From the data, it is visible that the proposed method outperformed all the other algorithms for the imbalance dataset. However, the result was not up to par. So, we used SMOTE to solve the class imbalance problem. After oversampling the dataset and solving class imbalance problem, we found the following outcomes:

Table 5. Result using balanced dataset

Model name	Accuracy	MCC	Precision	Recall	F-Measure	AUC
DT (Decision tree)	0.9433497537	0.8866226534	0.9523809524	0.9278350515	0.9399477807	0.9426911107
NB (Naïve Bayes)	0.8201970443	0.6433516357	0.8622754491	0.7422680412	0.7977839335	0.9159696557
RF (Random Forest) classifier	0.9802955665	0.9608841744	0.9946808511	0.9639175258	0.9790575916	0.9983101537
SVM (Support Vector Machine)	0.5270935961	0.07154584211	0.5037593985	0.6907216495	0.5826086957	0.5534429099
XGBoost	0.9827586207	0.9655326282	0.9895287958	0.9742268041	0.9818181818	0.9990031122
ANN (Artificial neural network)	0.9630541872	0.9261644557	0.9735449735	0.9484536082	0.9608355091	0.9928758996
Voting Classifier using scikit-learn with Decision Tree, Naive Bayes, Random Forest, SVM, XGBoost, and ANN	0.9704433498	0.9407902252	0.9739583333	0.9639175258	0.9689119171	0.996595993
Suggested Model (Stacked ml model using scikit-learn using Decision Tree, Naive Bias, Random Forest, Linear Regression, SVM, XGBoost, ANN)	0.9901477833	0.9804309001	1	0.9793814433	0.9895833333	0.9978603385

The proposed methodology surpassed all the results.
The followings are the comparison of different models:

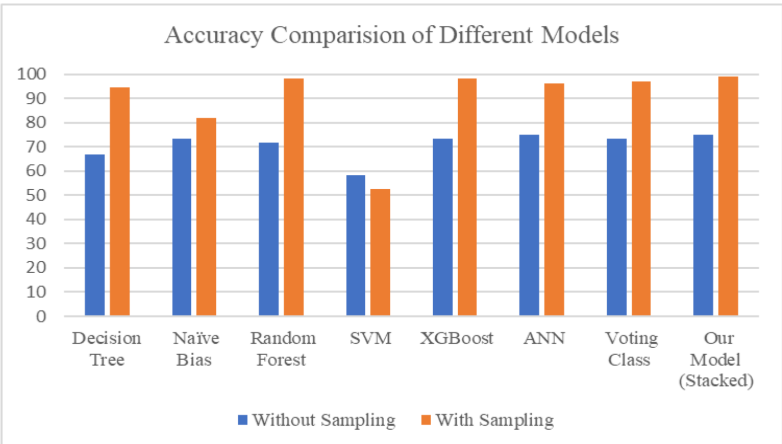

Fig.3. Accuracy comparison of different models

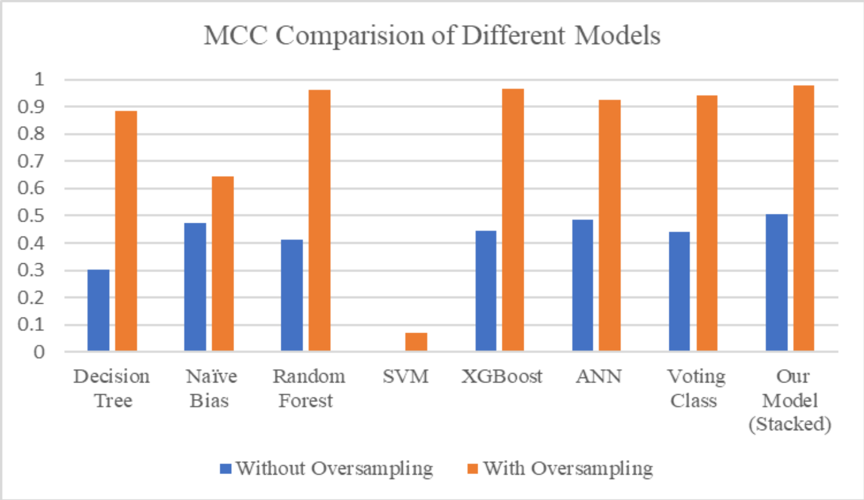

Fig.4. MCC comparison of different model

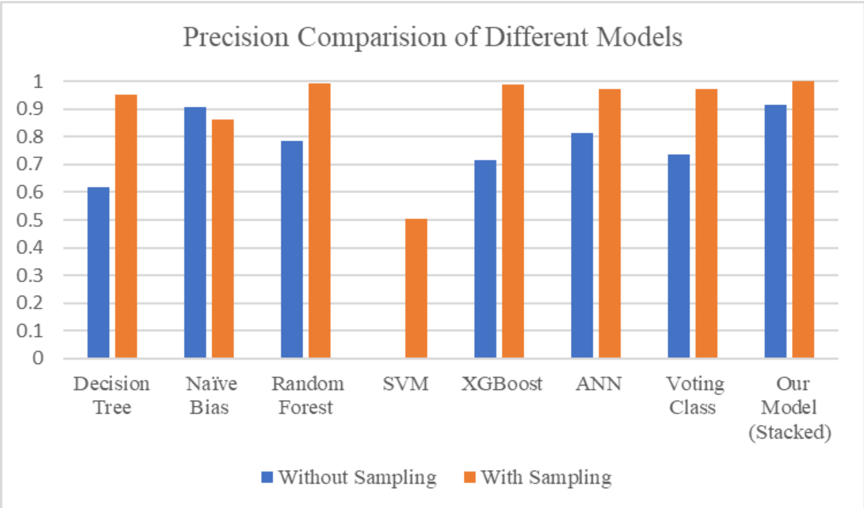

Fig.5. Precision comparison of different models

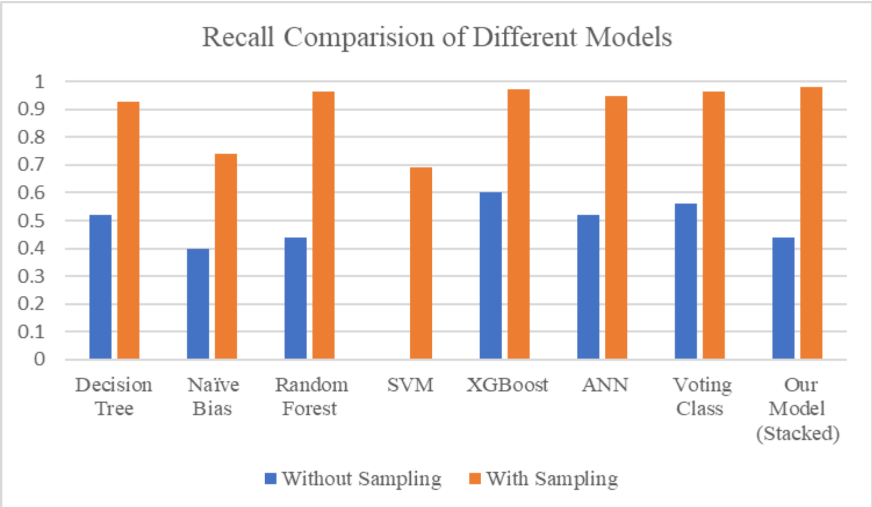

Fig.6. Recall comparison of different models

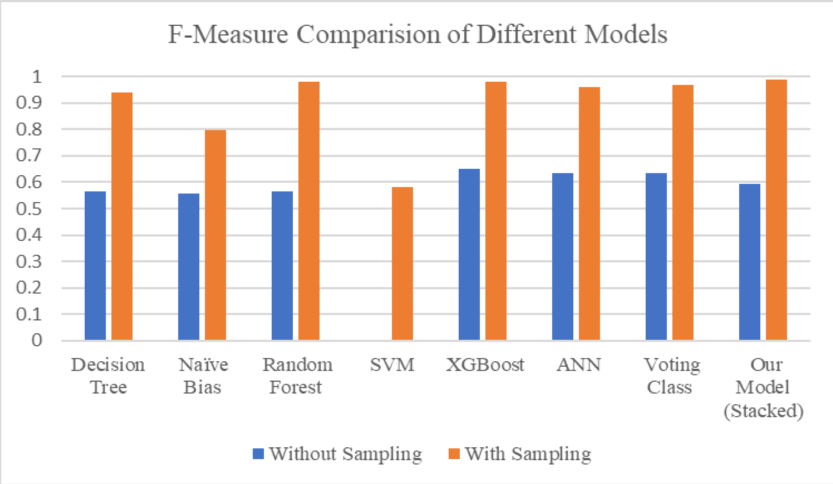

Fig.7. F-Measure comparison of different models

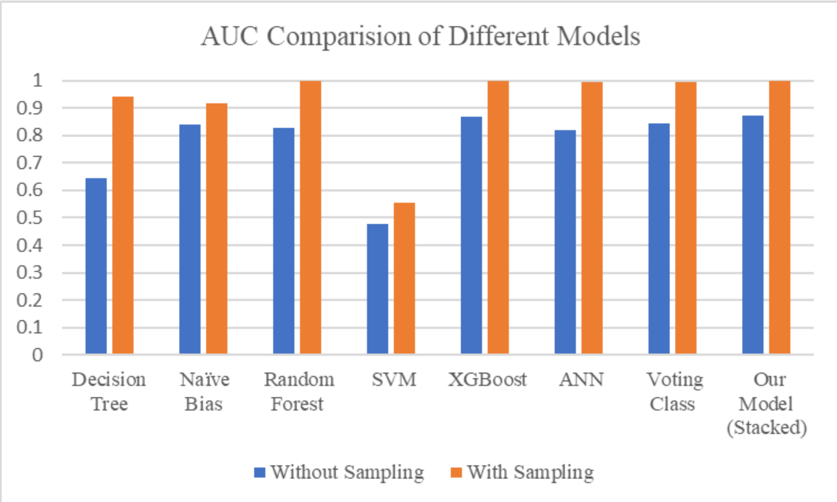

Fig.8. AUC comparison of different models

On the above we compare different models using various model such as decision tree, Naïve Bayes, Random forests, SVM, XGBoost, ANN, Voting Classifier and Stacked model on two datasets, one is dataset with sampling and other one is dataset without sampling. Diverse result can be seen in the graphs for different types of comparison.

The concept of explainable AI is that artificial intelligence systems should give concise, comprehensible justifications for their decisions and behaviors. Increased transparency, trustworthiness, compliance, bias detection, debugging and improvement, human-AI collaboration, domain-specific understanding, error detection, education and knowledge transfer, and ethical considerations are just a few of the reasons why this is important. Users can check the justification for a decision with explainable AI, boosting responsibility and trust. Additionally, it aids organizations in fulfilling legal obligations for the fairness and transparency of automated decision-making processes. Additionally, explanations help developers find and correct mistakes or defects, improving performance and dependability. Furthermore, Explainable AI encourages human-AI collaboration by enabling users to offer suggestions and feedback to enhance the system's judgment. We introduced experimental AI in response to RQ: 03, and we found the reason for heart failure using explainable AI.

From the fig.9, it can be seen that the feature value depends on the SHAP value for each of the individual parameters. Here it can be seen that, for the time attribute, when the SHAP value is negative and the range is -6 ≥ SHAP value ≤ -1, the feature value tends to be high. For serum_creatinine, when the SHAP value is positive and the range is 1 ≥ SHAP value ≤ 4, the feature value tends to be high. When the SHAP value is negative and the range is -2.5 ≥ SHAP value ≤ -1, the feature value is high for ejection_fraction. The feature value is high for age, in the range of -1 ≥ SHAP value ≤ 4. For creatinine_phosphokinase, when the SHAP value is positive and the range is 0 ≥ SHAP value ≤ 3 & 3.9 ≥ SHAP value ≤ 4.1, the feature value tends to be high. The feature value of platelets is high for the range of -2 ≥ SHAP value ≤ 1. Like platelets, the feature value of serum_sodium, sex is high in the range of -3 ≥ SHAP value ≤ 1 and -1 ≥ SHAP value ≤ 0 respectively. The feature value of diabetes is high for the range of -1 ≥ SHAP value ≤ 1. For smoking, when the SHAP value range is -1 ≥ SHAP value ≤ 1, the feature value of smoking is high. The feature values of anaemia, high_blood_pressure is high for the range of -1 ≥ SHAP value ≤ 1 and 0 ≥ SHAP value ≤ 0.2 respectively.

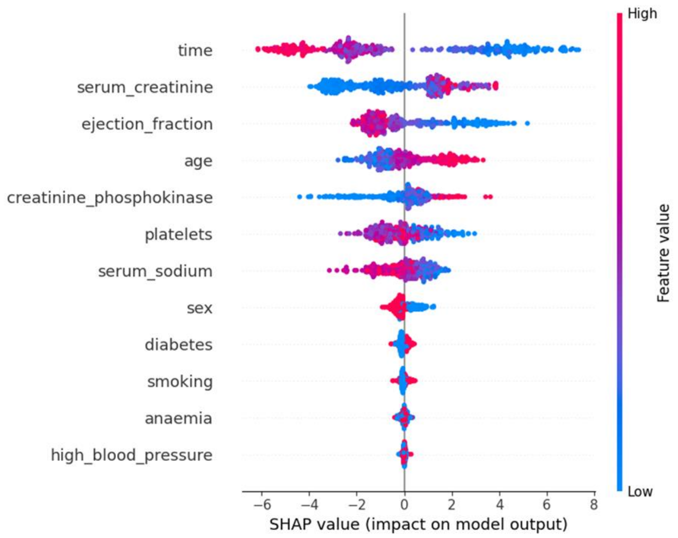

Fig.9. Findings of explainable AI

5. Conclusions

In this paper a stacked machine learning model using scikit-learn using Decision Tree, Naive Bias, Random Forest, Linear Regression, SVM, XGBoost, and ANN was introduced. This study used the stacked model, which provided higher accuracy than all the other existing solutions. Our proposed explainable AI helped to find out the actual reason for heart failure, which will help the doctors diagnose. According to the results of the study, doctors could use this method in the future to correctly predict a patient's risk of having a heart attack which put positive impact in the medical sector of a country. Despite of that, there are some scopes of this study that will need to be covered in the future work, such as implementing our model on comparatively high-quality dataset, build a system using our model, further refine the explainable AI etc.

References

[1] Moreno-Sanchez, P. A., "Development of an Explainable Prediction Model of Heart Failure Survival by Using Ensemble Trees," 2020 IEEE International Conference on Big Data (Big Data), Atlanta, GA, USA, 2020, pp. 4902-4910, doi: 10.1109/BigData50022.2020.9378460

[2] Mbotwa, J., Kamps, M. D., Baxter, P. D., Gilthorpe, M., "Application of Cox Model to predict the survival of patients with Chronic Heart Failure: A latent class regression approach." arXiv preprint arXiv:1907.07957 (2019).

[3] Haque, M. E., Uddin, S., Islam, M. A., Khanom, A., Suman, A., Paul, M., "Analysis and prediction of heart stroke from ejection fraction and serum creatinine using LSTM deep learning approach." arXiv preprint arXiv:2209.13799 (2022).

[4] David, H. B. F., Belcy. S. A., "HEART DISEASE PREDICTION USING DATA MINING TECHNIQUES." ICTACT Journal on Soft Computing 9, no. 1 (2018).

[5] Yang, X., Gong, Y., Waheed, N., March, K., Bian, J., Hogan, W. R., Wu, Y., "Identifying cancer patients at risk for heart failure using machine learning methods." In AMIA Annual Symposium Proceedings, vol. 2019, p. 933. American Medical Informatics Association, 2019.

[6] Dong, S., Mutharasan, R., Jonnalagadda, S. R., "Using Natural Language Processing to Screen Patients with Active Heart Failure: An Exploration for Hospital-wide Surveillance." arXiv preprint arXiv:1609.01580 (2016).

[7] Jabbar, M. A., Deekshatulu, B. L., Chandra, P., "Heart disease prediction using lazy associative classification." In 2013 International Mutli-Conference on Automation, Computing, Communication, Control and Compressed Sensing (iMac4s), pp. 40-46. IEEE, 2013.

[8] Roadknight, C., Suryanarayanan, D., Aickelin, U., Scholefield, J., Durrant, L., "An ensemble of machine learning and anti-learning methods for predicting tumour patient survival rates." In 2015 IEEE International Conference on Data Science and Advanced Analytics (DSAA), pp. 1-8. IEEE, 2015.

[9] Dev, S., Wang, H., Nwosu, C. S., Jain, N., Veeravalli, B., John, D., "A predictive analytics approach for stroke prediction using machine learning and neural networks." Healthcare Analytics 2 (2022): 100032.

[10] Yin, Y., Chou, C., "Early ICU Mortality Prediction and Survival Analysis for Respiratory Failure." arXiv preprint arXiv:2109.03048 (2021).

[11] 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, no. 1 (2020): 1-16.

[12] Chicco, D., Jurman, G., "The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation." BMC genomics 21, no. 1 (2020): 1-13.

[13] Chicco, D., Rovelli, C., "Computational prediction of diagnosis and feature selection on mesothelioma patient health records." PloS one 14, no. 1 (2019): e0208737.

[14] Patel, J., TejalUpadhyay., Patel, S., "Heart disease prediction using machine learning and data mining technique." Heart Disease 7, no. 1 (2015): 129-137.

[15] Erdaş, Ç. B., Ölçer. D., "A machine learning-based approach to detect survival of heart failure patients." In 2020 Medical Technologies Congress (TIPTEKNO), pp. 1-4. IEEE, 2020.

[16] Ishaq, A., Sadiq, S., Umer, M., Ullah, S., Mirjalili, S., Rupapara, V., Nappi, M., "Improving the prediction of heart failure patients' survival using SMOTE and effective data mining techniques." IEEE access 9 (2021): 39707-39716.

[17] Alotaibi, F. S. "Implementation of machine learning model to predict heart failure disease." International Journal of Advanced Computer Science and Applications 10, no. 6 (2019).

Authors' Profiles

Md. Raihan Mahmud is currently pursuing a master's degree in computer science at American International University-Bangladesh. He obtained a B.Sc. in Computer Science and Engineering from the same university. Machine-learning, health and biomedical analytics, and software engineering are among Raihan's research interests.

Prof. Dr. Dip Nandi received his PhD in Computer Science from RMIT University's School of Computer Science and Information Technology in Melbourne, Australia. He is an associate dean in the Department of Computer Science at American International University-Bangladesh. Online Learning, E-Learning, Online Education, Technology Enhanced Learning, Distance Education, Theory of e-Learning, ICT in Education are his research area.

Dr. Md. Shamsur Rahim received his PhD from the University of Technology Sydney, Australia. Before that he obtained his Bachelor of Science and Master of Science in Computer Science degrees in 2014 and 2016, respectively, from American International University-Bangladesh. Data Science, and Software Engineering are among Rahim's research interests.

Dr. Christe Antora Chowdhury received her MPH from American International University-Bangladesh and MBBS from Bangladesh Medical College, Dhaka Bangladesh. Her research interest includes Public Health, Disease Demography etc.

How to cite this paper: Md. Raihan Mahmud, Dip Nandi, Md. Shamsur Rahim, Christe Antora Chowdhury, "A Proposed Stacked Machine Learning Model to Predict the Survival of a Patient with Heart Failure", International Journal of Intelligent Systems and Applications(IJISA), Vol.16, No.3, pp.35-46, 2024. DOI:10.5815/ijisa.2024.03.03