



An improved machine learning-based prediction framework for early detection of events in heart failure patients using mHealth

Deepak Kumar¹ · Keerthiveena Balraj² · Sandeep Seth³ · Shivani Vashista³ · Manojkumar Ramteke^{1,2}  · Anurag S. Rathore^{1,2}

Received: 23 July 2023 / Accepted: 25 February 2024 / Published online: 5 March 2024

© The Author(s) under exclusive licence to International Union for Physical and Engineering Sciences in Medicine (IUPESM) 2024

Abstract

Purpose Acute decompensated heart failure (ADHF) is the most prevalent cause of acute respiratory distress worldwide, accounting for the majority of new cases and associated fatalities according to global statistics, making it a serious public health issue at present. The prognosis and probability of survival can be greatly enhanced by an early diagnosis of ADHF since it encourages patients to receive prompt clinical care. Over the past decade, there have been notable advancements in the use of artificial intelligence to interpret cardiovascular data from echocardiograms, and cardiac magnetic resonance imaging to ascertain hazard manifestations or future risks of cardiovascular disorders.

Methods In this paper, a model is devised to forecast events in outpatients with heart failure. This analysis utilized ten classification models to estimate the patient's prognosis. The order of the importance of the features is determined based on the recursive feature elimination technique by using all of the aforementioned approaches as a base model. The top five features such as gender, diabetes mellitus, systolic blood pressure, sodium, and heart rate were selected from the average ranking of the classifier.

Results The experimental results demonstrate that the Gaussian Naive Bayes classifier is superior to other models on an average performance basis, whereas K-Nearest Neighbor provided the best results over other classifiers with precision, recall, and F1 scores of 0.98, 0.91, and 0.94, respectively.

Conclusion Finally, a web-based mHealth application is built to estimate the probability of heart failure depending on the accuracy level.

Keywords Heart failure analysis · Data mining · Class imbalance · Feature extraction · Classification

1 Introduction

According to the World Health Organization (WHO) the top three causes of death worldwide are: (a) cardiovascular diseases, including ischaemic heart disease and stroke; (b)

Respiratory diseases which consist of chronic obstructive pulmonary disease and lower respiratory infections; (c) Neonatal conditions (birth asphyxia and birth trauma, neonatal sepsis and infections, and preterm birth complications) [1]. Of these, cardiovascular diseases are the leading cause of mortality across the globe, accounting for 16% of the total number of fatalities [2]. Heart failure occurs when the heart's ability to pump blood at a pace sufficient for the needs of the metabolizing tissues is compromised or can only do so with a raised diastolic filling pressure due to an anomaly of cardiac function. Figure 1 represents the global statistics and age distribution pattern of the rising prevalence and incidence of heart failure. In 2017, there were 56.5 million deaths worldwide, with slightly under half of them adults over the age of 70, 26% were between the ages of 50 and 69, 13% were between the ages of 15 and 49, 1% were between the ages of 5 and 14 or below, and over 9% were children under the age of 5 [3].

Deepak Kumar and Keerthiveena Balraj have contributed equally.

✉ Manojkumar Ramteke
ramtekemanoj@gmail.com

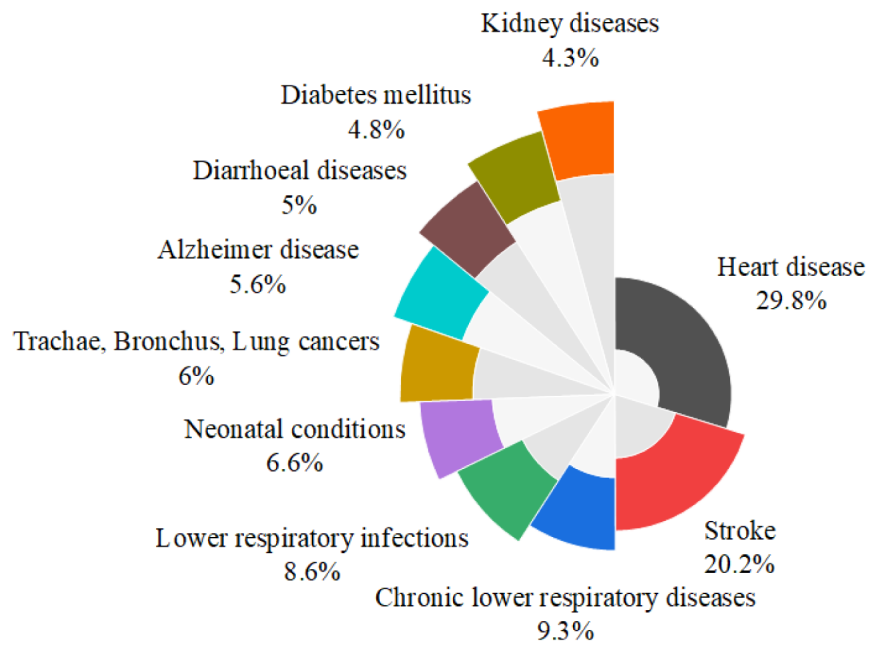
✉ Anurag S. Rathore
asrathore@biotechmz.com

¹ Department of Chemical Engineering, Indian Institute of Technology, Hauz Khas, New Delhi 110016, India

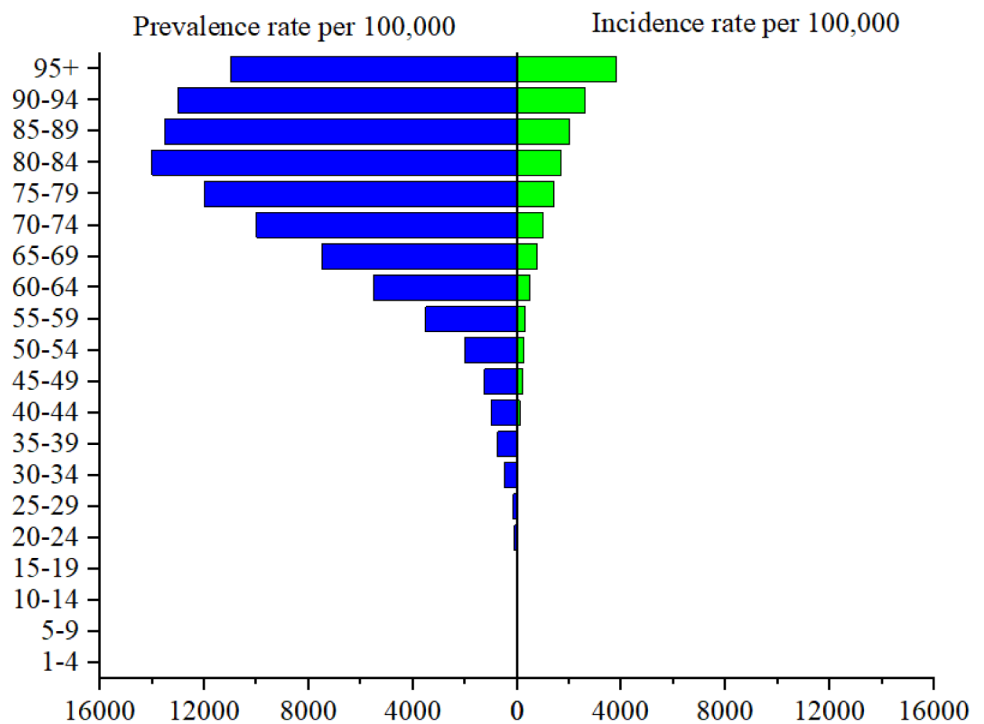
² School of Artificial Intelligence, Indian Institute of Technology, Hauz Khas, New Delhi 110016, India

³ All India Institute of Medical Sciences Delhi, India, Delhi

Fig. 1 (a) Leading causes of death worldwide and (b) Age distribution of heart disease worldwide based on prevalence and incidence



(a)



(b)

In 2020, India, Pakistan, Bangladesh, Sri Lanka, Nepal, Bhutan, and the Maldives will collectively account for 1.8 billion people, thereby accounting for 23% of the world's population [4]. South Asian nations continue to bear a heavy burden of disorders including congenital, nutritional, and

infectious diseases that are typical of earlier phases of the epidemiological transition. Every year, over 17 million people worldwide die from cardiovascular diseases, of which 0.5 to 1.8 million new cases of heart failure per year are in India [5]. Further, heart failure is strongly associated with

age, and the population of India is on average younger than that of other countries [6].

There are several risk factors for heart failure, including gender, family history, and age. These risk factors are categorized as uncontrollable risk factors, whereas high cholesterol, smoking, high blood pressure, and obesity are categorized as controllable risk factors [7]. In the realm of healthcare, electronic health records (EHR) [8] are a systematized collection of patient data that may also be used in clinical and research settings. The EHR can assist care-related tasks, either directly or indirectly, via various interfaces, such as evidence-based decision support, quality monitoring, and outcomes reporting. Currently, mass screening for heart disease is time-consuming and error-prone, and hence expert systems built on artificial intelligence (AI) models are effectively utilized to predict heart failure readmissions and modality rates. These strategies employ patient data to identify the occult patterns that have a big impact on the outcomes. For predicting hospitalization due to heart failure, the combination of clinical, therapeutic data, routine laboratory data, and biomarkers is also beneficial.

Healthcare services, primary care, standardization, cost, data privacy, and security are the challenges faced in the healthcare sector [9]. Although the effectiveness of AI has been established in many domains, the medical sector has encountered significant deployment challenges for probable reasons: (1) the need for enormous patient data to train the model; (2) If the data used to train machine learning (ML) models is biased, AI-powered systems using those models will produce misleading results; (3) When all accessible input data is used, the results might be tainted [6]. To achieve rapid pattern and anomaly detection, the best features for the model are automatically identified by feature engineering leveraging domain knowledge and data mining. Various automated approaches for heart disease identification have been suggested by researchers in recent years using clinical feature-based data modalities [9–14].

Currently, it is still challenging to predict the survival of patients with heart failure both in terms of achieving high prediction accuracy and ascertaining the contributing components. As highlighted by Sakamoto et al., most models created for this purpose only achieve moderate accuracy, with poor interpretability from the predictive parameters, mostly owing to a lack of repeatability [15]. To determine the best features and estimate the rate of modality, researchers have used Cox regression, Kaplan–Meier survival plot, and sex-based mortality prediction models [16]. While multiple researchers have used standard statistical approaches to generate some intriguing outcomes, these techniques are ineffective when working with huge datasets [17], and more efficient ML algorithms are needed. In recent years, mobile health (mHealth) has garnered considerable attention and popularity due to the pervasive use

of mobile devices and the availability of numerous health and wellness applications [18]. A comprehensive evaluation conducted in 2020 [12] also concluded that mobile health interventions were superior to standard treatment in lowering all-cause mortality and all-cause hospitalizations among patients with heart failure.

The use of mHealth applications has the potential to revolutionize the delivery of healthcare by rendering healthcare solutions more practical and readily accessible. It is estimated that over 500 million patients utilize mHealth apps to assist their activities related to self-care [19]. In addition, fitness tracking, stress management, monitoring blood pressure, Chronic disease management, tracking fetal development, and social support are some of the most promising categories of mHealth utilization. The following are the main characteristics of mHealth apps for cardiovascular disease self-management [20]: (1) general health status; (2) parameters like weight, step count, heart rate, and blood pressure monitoring; (3) electronic health records; (4) control and follow-up of medication; and (5) real-time warning and medication reminders. This research project explored the feasibility of using mHealth technology to aid cardiovascular disease patients in disease management.

In view of these gaps, we propose a ML-based approach for accurate survival diagnosis of cardiovascular patients. The primary objective is to estimate the mortality rate of heart failure in India and explore its relationship with other significant risk factors. To address the issue of class imbalance, no sampling, random oversampling (ROS), random under sampling (RUS), and Synthetic Minority Oversampling Technique (SMOTE) were utilized. Ten different classifiers were employed for medical data mining to predict mortality from heart failure. Furthermore, a robust technique based on recursive feature elimination (RFE) is utilized to extract essential features from the dataset that affect the performance of the ML algorithm, and these features are then used to investigate the most significant risk factors. Figure 2 depicts each level of the proposed machine learning-based approach for heart failure prediction. The model that offers the best accuracy is chosen in the subsequent phase and combined with a web-based mHealth application to diagnose heart failure conditions. The Python programming language was used to construct the Flask backend for this web-based application. The proposed mHealth apps have a (1) user-friendly interface, (2) more reliable data, (3) the ability to be used by medical professionals even in the absence of an internet connection, and (4) highly customized applications with distinguished features. The following are the key contributions of the proposed work:

- An efficient decision support system has been developed for early prediction of heart failure patients.

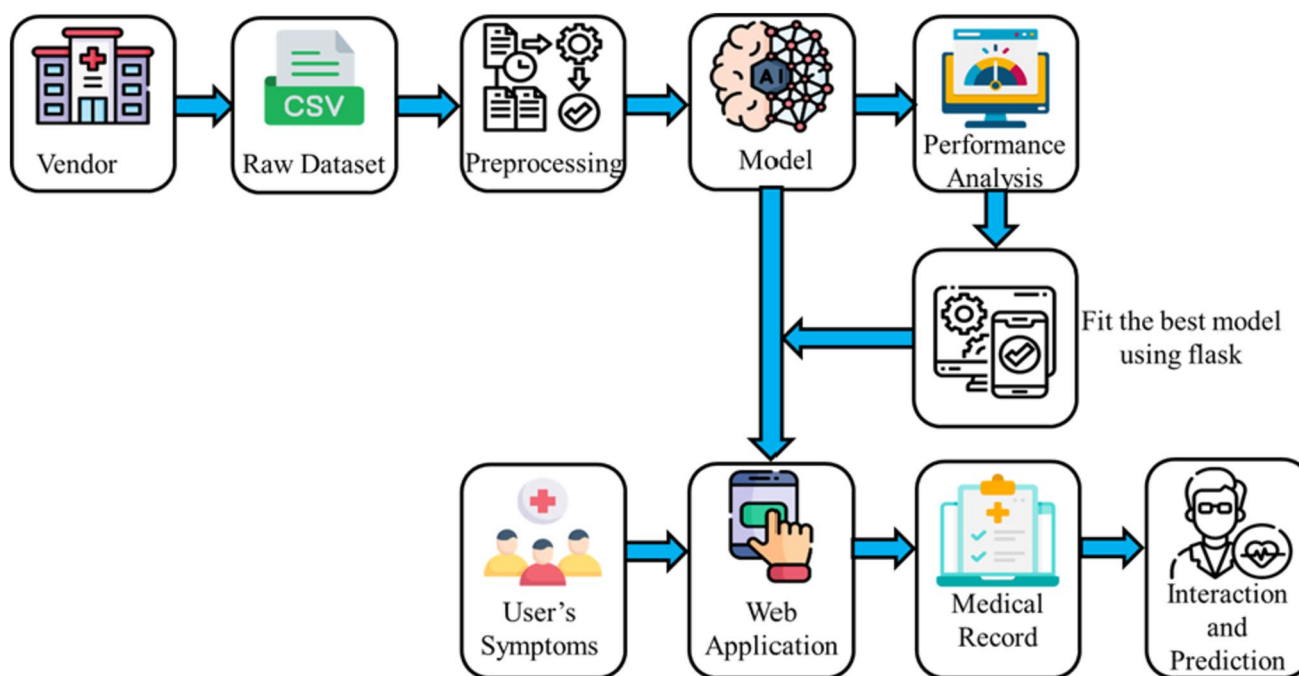


Fig. 2 Overview of the proposed machine learning-based heart failure prediction model

- ROS, RUS, and SMOTE have been used as benchmarks against which other classification models' ability to predict cardiac patient survival is evaluated.
- RFE technique has been used to facilitate a decision support system to automatically identify the highest-ranking optimum features during the training phase in order to improve performance.
- The effectiveness of the machine learning algorithm is evaluated with regard to several criteria, including precision, recall, and F1 scores.
- Furthermore, a web application utilizing the Python Flask web development framework was developed by integrating the proposed model.

2 Related work

Healthcare experts often employ computer-aided diagnostic (CAD) systems to save expenses, assist physicians in disease detection, and distinguish between disease progressions. Many studies in the field of cardiovascular diseases use CAD systems based on echocardiography, cardiac magnetic resonance, cardiac computed tomography, or single photon emission computed tomography [21]. When it comes to the detection of cardiac issues, angiography is among the most reliable of the traditional diagnostic tools. Angiography, on the other hand, has a few limitations, including high cost, computational complexity, and the need for advanced technology [22]. Due to human error, conventional

approaches often result in inaccurate diagnoses and involve time-consuming evaluations. This has motivated researchers to develop a non-invasive smart healthcare system based on predictive machine learning (ML) [23]. These predictions can then be used to identify patients at risk of developing heart failure or other conditions, allowing healthcare providers to intervene early and provide treatment before the condition becomes severe. By identifying patients at risk early and providing appropriate treatment, the mortality rate for heart failure can be reduced.

Machine learning models have been extensively employed in detecting precursor markers at the early stages of cardiac disease. Risk factors for cardiovascular diseases include tobacco use, advanced age, diabetes, and high blood pressure [24, 25]. With 159 patient data, Yang et al. [26] developed a scoring model based on support vector machines and obtained a Youden's index of 69, a sensitivity of 75%, and a specificity of 94%. Twelve different parameters were adopted, including the ratio between low-frequency and high-frequency heart rate variability, sodium, B-type natriuretic peptide, left ventricular ejection fraction, left ventricle end-diastolic diameter, left atrium maximal volume, left ventricular posterior wall, P-R interval, cardiothoracic ratio, six-minute walk distance, standard maximum oxygen consumption, and the ratio between early and late ventricular filling velocity. Overall, the model offered a classification accuracy of 74.4%, with individual accuracy rates of 78.79%, 87.5%, and 65.85% for identifying the healthy group, the heart failure prone group, and the heart failure group. With features

including gender, age, blood pressure, and cigarette usage, Gharehchopogh and Khalifelu [27] were able to predict 85% of test cases accurately using data from the medical records of 40 patients.

Accuracy rates of 79.54%, 61.46%, and 68.96% have been reported for the identification of left anterior descending, left circumflex, and right coronary artery stenoses, respectively, using a bagging-based classification model (Alizadeshani et al. [28]). A two-tier ensemble model has been suggested by Tama et al. [29] which incorporates the predictions of class labels from a Gradient Boosting Machine (GBM), Random Forest (RF), and Extreme Gradient Boosting (XGBoost). The investigation conducted by Parthiban and Srivatsa [30] concentrated on diabetic patients who also had cardiac problems. They utilized a variety of different prognostic factors, including age, blood pressure, and blood sugar levels, among others. They were able to attain an accuracy of 94.60% using the support vector machine (SVM) classifier. The training was performed on an imbalanced dataset which is very common in heart diagnosis because certain medical conditions, such as heart failure, are relatively rare compared to healthy cases. An imbalanced dataset may result in biased training, as the model may be trained to predict the majority class more accurately than the minority class. Additionally, if the model is not able to generalize well and has poor performance on unseen examples, it may misdiagnose healthy patients as having heart failure, which can lead to unnecessary treatment and costs.

Adrian et al. [31] collected features such as blood urea nitrogen, confidence interval, chronic kidney disease epidemiology collaboration, estimated glomerular filtration rate, high-density lipoprotein, hazard ratio, N-terminal pro-B-type natriuretic peptide, systolic blood pressure from multi centers. They identified predictors of mortality, and the predictors of hospitalization due to heart failure were noticeably different. In order to better understand the causes of mortality and cardiac problems, Shah et al. [32] established a research framework that employs ML techniques. Only 14 of the possible 76 features were chosen since the researchers' primary focus was on developing an accurate and effective system with a minimum number of components. K-Nearest Neighbor (KNC) was the most accurate classifier out of a total of four, which included Naive Bayes (NB), Radial Basis Function (RBF), and ensemble technique.

Our literature survey indicated that only a limited number of studies have been performed on heart failure diagnosis. In this study, we aim to develop a decision support system for detecting heart failure that is both accurate and efficient. Dataset acquisition is difficult when building an ML-based system, especially in the medica

l field as issues arise regarding privacy, confidentiality, ethics, and security [33]. Fortunately, multivariate data comprising 43 features were collected from the All India

Institute of Medical Sciences (AIIMS) for 152 patients and deployed for further study. Most significant features were selected from the feature space utilizing the average importance of each feature based on 10 different models. During RFE, the best precision and F1-score of 0.98 and 0.95, respectively were achieved using KNC with 4 features only. While most of the published studies utilize a single model to select the best features, the feature importance varies with respect to the selection of the model. So, averaging out the feature importance over multiple models helped us in identifying the more accurate importance of each feature. The significant properties identified by RFE were gender, diabetes mellitus, Na, heart rate, and systolic blood pressure. On these selected features, GNB was found to be the best among all chosen classifiers, with the best precision value of 0.86 and recall value of 0.71.

3 Methods and materials

In this study, a ML model is proposed to detect heart failure using data collected between 2020 and 2022 at the All India Institute of Medical Sciences (AIIMS), New Delhi. The model is composed of supervised machine learning classification techniques including Decision tree classifier (DTC), Adaptive boosting classifier (ABC), Random forest classifier (RFC), Gaussian process classifier (GPC), Gradient boosting classifier (GBC), k-nearest neighbors classifier (KNC), Extra tree classifier (ETC), Gaussian Naive Bayes classifier (GNB), Multi-layer perceptron (MLPC) and Support vector machine (SVM). Figure 3 illustrates the architectural schematic of the model. It is seen that there are seven steps, including pre-processing, scaling, sampling, training, assessment, feature engineering, and classification. Features having more than 10% missing data are removed from the dataset. Further, patients having missing features greater than 5% are removed. The remaining missing values are filled based on predicted value using a regression model with the top three correlated variables as input. Further, two scaling techniques, standard scaling, and min–max scaling are applied to the whole dataset followed by a random selection of training and testing datasets. Furthermore, random oversampling (ROS), random under-sampling (RUS), and synthetic minority oversampling technique (SMOTE) are used to handle the imbalanced class across the training set. The aforementioned 10 different classifiers are trained on the training dataset and the importance order of the feature set is calculated using recursive feature elimination. A combined average feature ranking is calculated to select important features so as to make better predictions with fewer features helping doctors for accurate and early diagnosis.

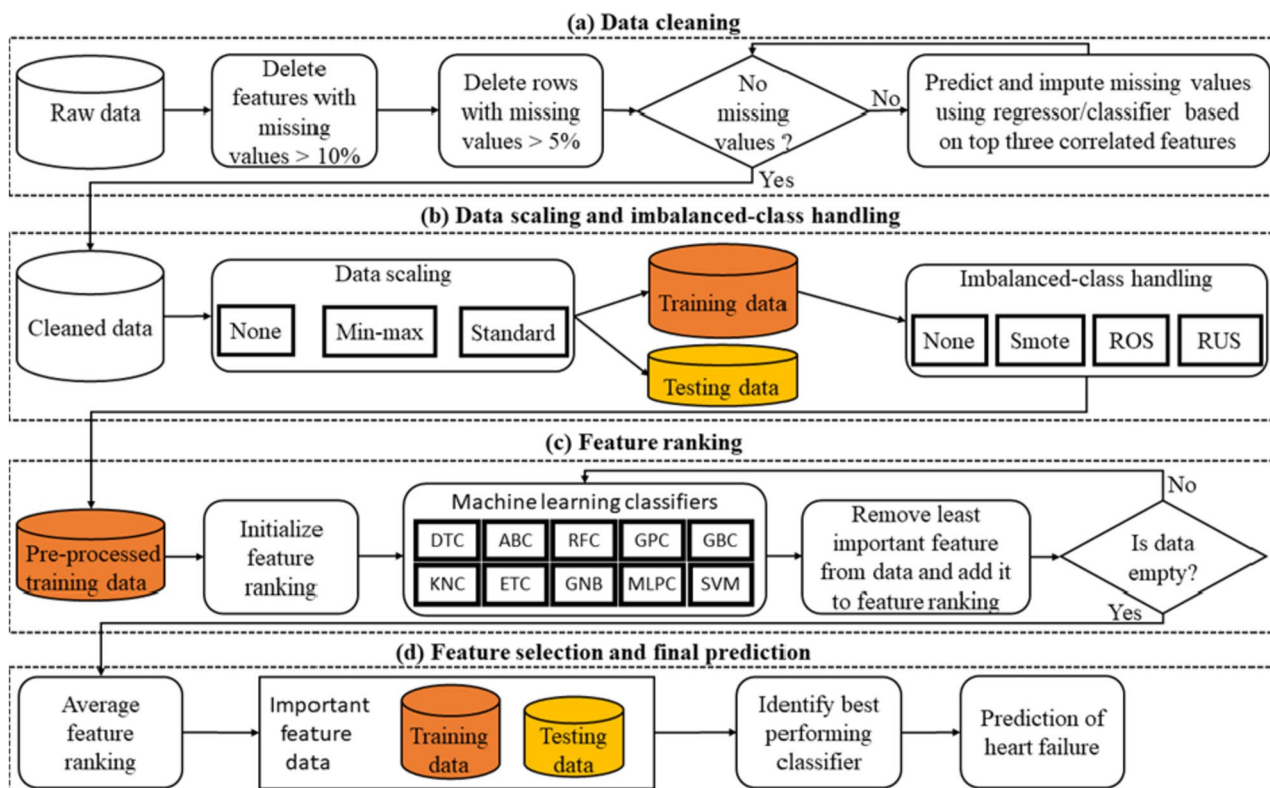


Fig. 3 Flowchart of the proposed framework for predicting patient's survival in heart failure

3.1 Dataset description

The current investigation is carried out prospectively on individuals who were hospitalized over a two-year period (2020–2022) at AIIMS Delhi, India. The patients range in age from 17 to 80, with 30.46% women and 69.54% men. All 151 patients had acute decompensated heart failure and fell into classes II, III, and IV of the New York Heart Association's (NYHA) assessment of the stages of heart failure according to their past heart failures.

Table 1 consists of 33 features, which report information on the body, clinical data, and lifestyle: BNP, stair, age, gender, diabetes mellitus (DM), thyroid, chronic obstructive pulmonary disease (COPD), chronic kidney disease (CKD), cardiac resynchronization therapy (CRT), edema, tobacco usage, systolic blood pressure (SBP), diastolic blood pressure (DBP), heart rate (HR), sodium (Na), potassium (K), blood urea, serum creatinine (SCr), hemoglobin, white blood cell (WBC), albumin, total cholesterol, high-density lipoprotein (HDL), low-density lipoprotein (LDL), triglycerides, serum glutamic oxalacetic transaminase (SGOT/AST), serum glutamic pyruvic transaminase (SGPT/ALT), bilirubin, ejection fraction (EF), dressing yourself, showering bathing, walking 1 block on level ground, and hurrying or jogging.

3.2 Dataset pre-processing

Prior to building the machine learning model, the data must be pre-processed. Key activities include exploratory data analysis, generation of new features from the current features, transformation of data, and partitioning of data into training data, validation data, and testing data. A total of 151 patient records, each of which has 43 attributes, were retrieved from AIIMS Delhi, India. The handling of missing numbers, which are often represented by the notation NaN (Not a Number), is one of the first significant issues that machine learning must tackle. This issue is handled in an iterative way using a two-stage strategy that takes into account the total number of NaN cells. First, the column(s) containing NaN values are identified and then if the proportion of NaN values in those columns is more than 10%, the columns containing those values are removed. The row(s) that included more than 5% NaN values were removed during the second phase of the process. As a result of the preliminary processing, a total of 128 patients containing 33 features were adopted for further processing.

3.3 Scaling

When training on a dataset, it is important that all features get appropriate weightage in the determination of the outcome. As

seen above in Fig. 4, a few parameters such as BNP, SGOT, and SGPT typically fall into a range of thousand, while others, like SBP, DBP, HR, Na, cholesterol, HDL, and LDL fluctuate in the hundreds. Hence, feature scaling techniques are used to normalize the distribution of each feature's magnitude before the dataset is utilized. An overview of data preparation techniques along with feature selection and dimensionality reduction for the transformation of attributes has been offered by Tan et al. [34]. The following transformation approaches were examined: standard scaling (x_{ss}) and min–max scaling (x_{ms}).

$$x_{ss} = \frac{x - \bar{x}}{\sigma} \tag{1}$$

$$x_{ms} = \frac{x - x_{min}}{x_{max} - x_{min}} \tag{2}$$

where, \bar{x} and σ denote the average and standard deviation, respectively, and the output of x_{ss} is of the form $\bar{x} = 0$ and $\sigma = 1$, whereas x_{ms} restricts all values to the range [0, 1].

3.4 Class imbalance

A notable use of data analysis in the field of medicine is in the diagnosis of disorders. However, significant challenges that arise in this endeavor are the uneven distribution of data and the unequal quality of the majority

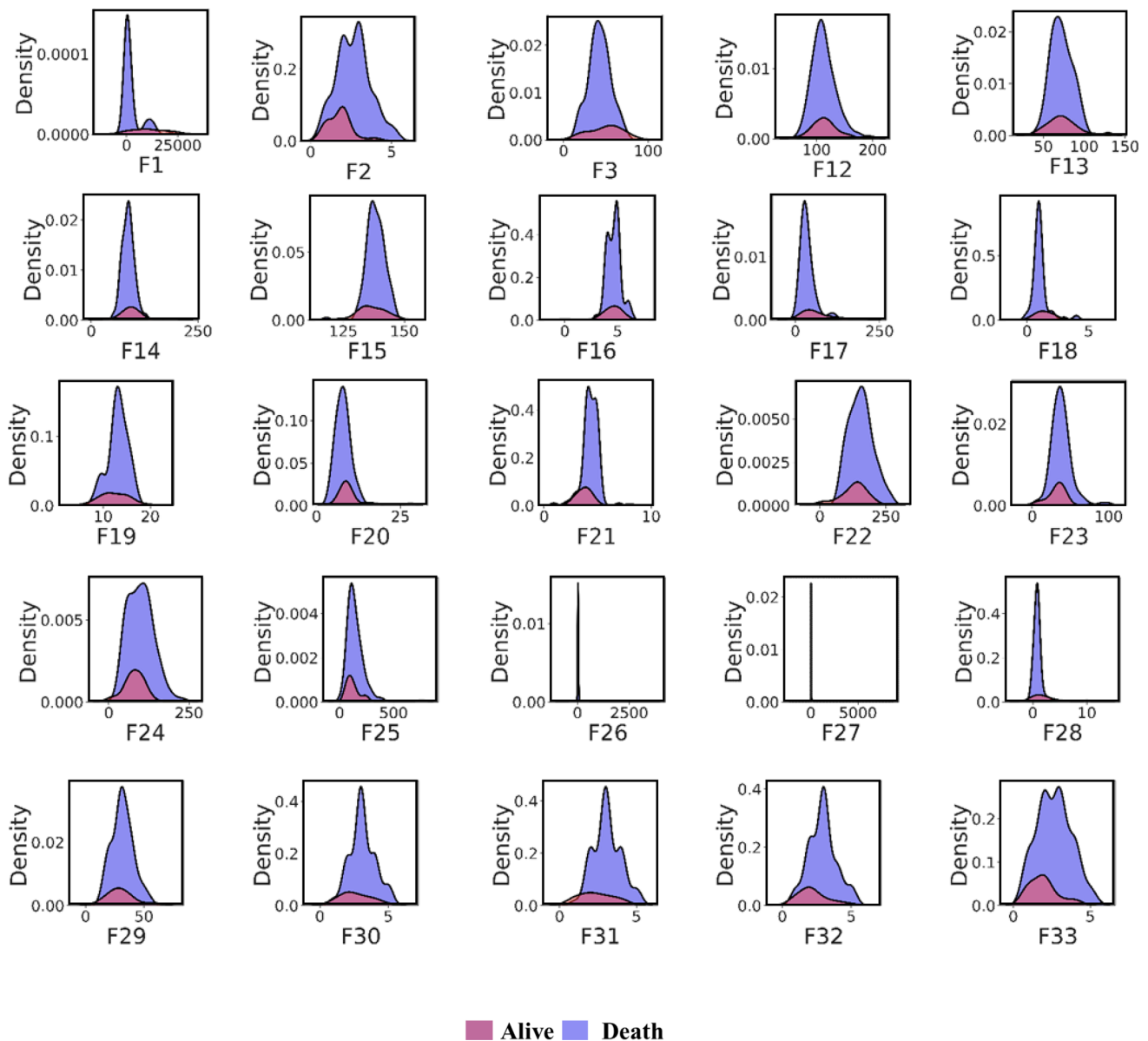


Fig. 4 Probability distribution for all features according to the target variable

Table 1 Characteristics of HF patients in the study cohort

Features	Description	Range	Measured In	
F1	Nt pro BNP	Biomarker	5–24,260	Binary
F2	Stair	Functional class	1–5	Binary
F3	Age	Age of patient	17–80	Binary
F4	Gender	Male or Female	0, 1	Boolean
F5	DM	Co-morbidities i.e. if the patients have a history or not	0, 1	Boolean
F6	Thyroid	Co-morbidities i.e., if the patients have a history or not	0, 1	Boolean
F7	COPD	Comorbidities i.e., if the patients have a history or not	0, 1	Boolean
F8	CKD	Co-morbidities i.e., if the patients have a history or not	0, 1	Boolean
F9	CRT	A device implanted in the chest to make the heart's chambers squeeze(contract) in a more organized and efficient way	0, 1	Boolean
F10	Edema	Edema present in extremities	0, 1	Boolean
F11	Tobacco use	The patient has a deleterious habit of tobacco chewing or not	0, 1	Boolean
F12	SBP	Checking the systolic blood pressure	70–196	Binary
F13	DBP	Checking the diastolic blood pressure	40–129	Binary
F14	HR	Recording of heart rate	42–190	Binary
F15	Na	Level of sodium in the blood	118–147	Binary
F16	K	Level of potassium in the blood	0–6.4	Binary
F17	Blood Urea	Level of blood urea in the blood	5–199	Binary
F18	SCr	Level of serum creatinine in the blood	0–5	Binary
F19	Hemoglobin	Level of hemoglobin in the blood	6.6–19	Binary
F20	WBC	Level of white blood cells in the blood	4–28	Binary
F21	Albumin	Level of albumin in the blood	0.9–8	Binary
F22	Total Cholesterol	Level of total cholesterol in the blood	9–268	Binary
F23	HDL	Level of high-density lipoprotein in the blood	4–100	Binary
F24	LDL	Level of low-density lipoprotein in the blood	3–230	Binary
F25	Triglycerides	Level of triglyceride in the blood	2–760	Binary
F26	SGOT/AST	Level of serum glutamic oxalacetic transaminase in blood	9–2923	Binary
F27	SGPT/ALT	Level of serum glutamic pyruvic transaminase in blood	0–6451	Binary
F28	Bilirubin	Level of bilirubin in blood	0–11	Binary
F29	EF	Percentage of blood leaving the heart at each concentration	10–60	Binary
F30	Dressing yourself	Functional class	1–5	Binary
F31	Showering bathing	Functional class	1–5	Binary
F32	Walking 1 block on level ground	Functional class	1–5	Binary
F33	Hurrying or jogging	Functional class	1–5	Binary

and minority groups, which typically result in misclassification [35]. The minority class samples are used to diagnose mortality rate, despite the fact that the majority class samples and their accurate classification are more crucial to the classifier. Misdiagnosis leads to additional clinical testing for individuals with mild heart failure but may be fatal for those with severe heart failure. Therefore, it is crucial from a clinical perspective to research the class imbalance by performing a detailed assessment of the impacts of the imbalanced data. In this context, three different techniques were employed namely, ROS, RUS, and SMOTE.

- ROS is a non-heuristic technique that includes randomly duplicating instances from the minority class and adding them to the training dataset.
- In RUS, instances from the majority class are randomly chosen and removed from the training dataset. As a result, there are fewer examples in the majority class in the training dataset that has been modified. Until the desired class distribution, such as an equal number of samples for each class is reached, this process can be repeated.
- SMOTE uses existing data to generate new synthetic data in order to strengthen the data for the minority class. As a result, the overfitting issue is avoided, and the decision

limits for the minority class are expanded into the space of the majority class.

Algorithm 1 Recursive feature elimination algorithm

Input: Initial feature subset, $F = \{F_1, F_2, \dots, F_n\}$

Output: Order feature set based on increasing rank score, R

Step 1: Initialize $R = \{ \}$

Step 2: Until F is no longer empty, repeat steps 3-8.

Step 3: Train the DTC, ABC, RFC, GPC, GBC, KNC, ETC, GNB, MLPC, and SVM using feature set F.

Step 4: Compute the ranking of all features using permutation feature importance (PFI).

Step 5: Sort the features according to their rank. $NewR_k = \text{sort}(\text{Rank})$

Step 6: Update $R = R + NewR_k[-1]$, where $NewR_k[-1]$ is the least important feature.

Step 7: Discard features with the lowest rank.

Step 8: Update $F = F - NewR_k[-1]$

4 Feature selection

Following the computation of the missing values, it is necessary to select the critical aspects that significantly correlate with features of relevance for disease diagnosis. The construction of a reliable diagnostic model is hindered by unnecessary and irrelevant features during the vector feature extraction process [36]. In this investigation, we utilized the recursive feature elimination (RFE) method to extract the most important characteristics of cardiovascular disorders of the patients. RFE is an advanced technique applied to methodically evaluate and prioritize features in a dataset by their level of importance [37, 38]. In each iteration, it detects and eliminates the least significant feature according to relevance scores, thereby improving model performance by preserving the most valuable features. As illustrated in Algorithm 1, the features search process employs backward selection, initially considering the complete feature set $F = \{F_1, F_2, \dots, F_n\}$, and iteratively eliminates features that do not enhance the accuracy of the classification. The RFE process iterates until it reaches the optimal number of features, ensuring an effective feature subset for enhancing classification accuracy.

5 Classification

Classification algorithms such as DTC, ABC, RFC, GPC, GBC, KNC, ETC, GNB, MLPC, and SVM have been employed in medical data mining to predict mortality from heart failure. A brief description of the various classification

algorithms that were examined in this study is shown in Table S1 of the supplementary section.

6 Results and discussion

To evaluate the efficacy of the proposed technique, a number of experiments for heart failure patients’ survival predictions were conducted. In order to construct a generic machine learning model predicting heart failure or mortality, we split our cohort into two sets of data: 70% for training, and 30% for testing over unlabelled data. Initially, we trained the model with a complete set of features followed by a significant set of features. ROS, RUS, and SMOTE were employed to address the issue of class imbalance. All the experiments were conducted in a Python environment using different libraries in a server with an Intel(R) Silver(R) 4210 CPU on 2.19 GHz and 128 GB RAM.

6.1 Evaluation metric

Accuracy, precision, recall, and the F-score are the three fundamental metrics that were investigated in this study to identify the extent of difference between different machine learning-based algorithms. These metrics are described in Eqs. (3) - (6) in which true positive (K_{TP}) represents correct predictions, whereas K_{FP} and K_{FN} represent false positive and false negative, respectively.

$$\text{Precision} = \frac{K_{TP}}{K_{TP} + K_{FP}} \times 100 \tag{3}$$

$$\text{Recall} = \frac{K_{TP}}{K_{TP} + K_{FN}} \times 100 \tag{4}$$

$$F1 - \text{score} = 2 \times \frac{\text{Precision} \times \text{recall}}{\text{Precision} + \text{recall}} \tag{5}$$

$$\text{Accuracy} = \frac{(K_{TP} + K_{TN})}{K_{TP} + K_{FN} + K_{FP} + K_{TN}} \tag{6}$$

6.2 Feature correlation

The dataset after deleting rows and columns having more than 5% and 10% missing data consists of 151 instances, 11 NaN values, and 33 features, including patient ID, age, gender, and other important features as indicated in Fig. 5. The dataset’s feature correlation matrix is depicted, which estimates the correlation coefficients between different variables. Each cell statistically represents the Pearson correlation coefficient (Eq. 7) between two variables. The values range from +1 representing a perfect direct

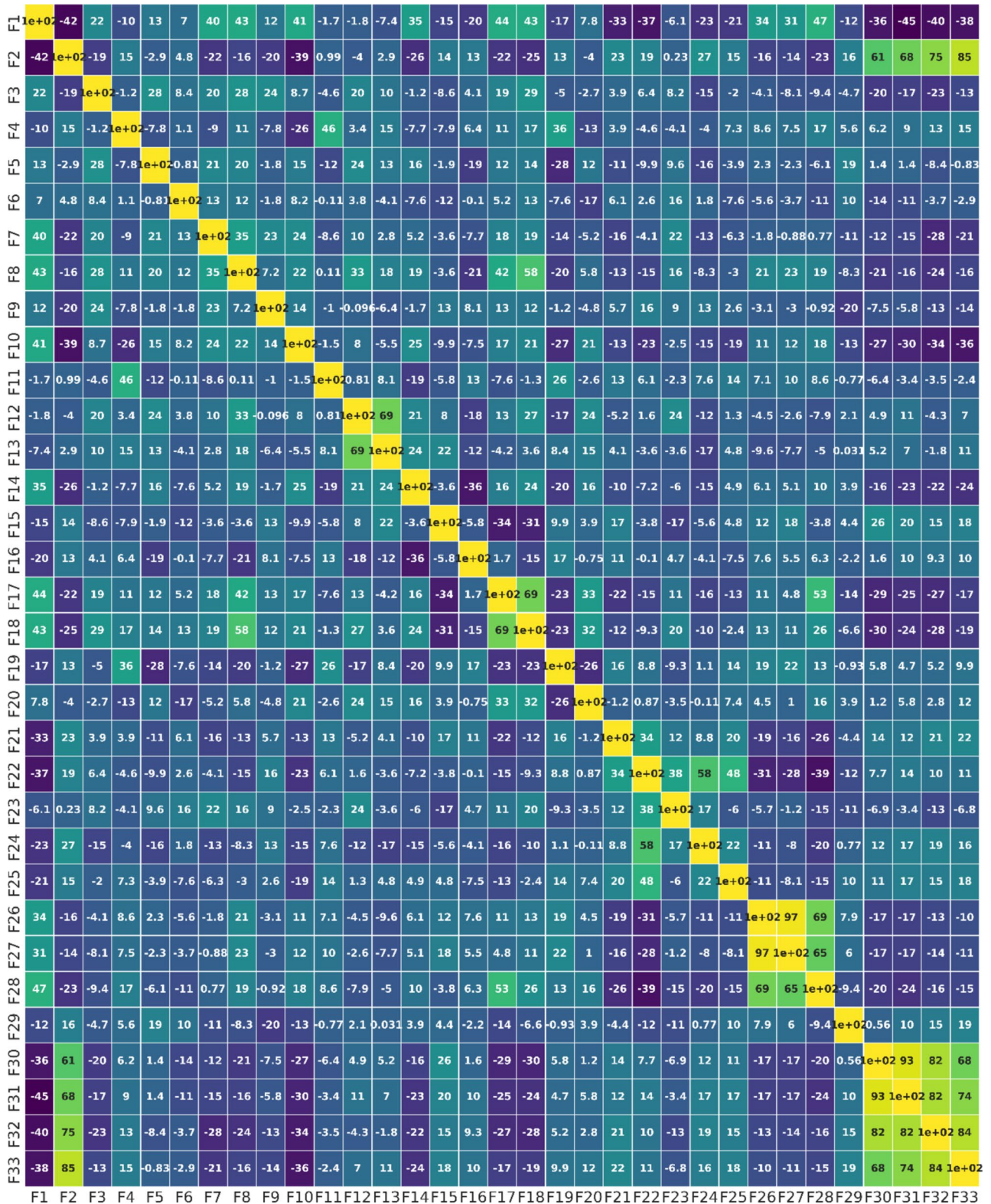


Fig. 5 Pearson's correlation matrix for all features included in the study. Except for features F30, F31, F32, and F33, there was no significant association between input variables

relationship and -1 indicating no direct relationship/perfect inverse relationship. Additionally, a correlation of 0 indicates there is no connection between the movements of the two variables.

$$\text{Pearson correlation coefficient} = \frac{\sum(V_i - \bar{V})(Y_i - \bar{Y})}{\sqrt{\sum(V_i - \bar{V})^2 \sum(Y_i - \bar{Y})^2}} \quad (7)$$

where, V_i and Y_i represents the value of the V and Y variables in a sample, \bar{V} and \bar{Y} represents the mean of the values of the V and Y variables. High correlations were observed between the features representing the best and worst values of the different feature sets. For instance, the stair (F2) is highly correlated with the dressing yourself (F30), showering bathing (F31), walking 1 block on level ground (F32), and hurrying or jogging (F33) features. SGOT (F26), and SGPT (F27) are highly correlated with each other, whereas BNP (F1) represents a poor correlation with total cholesterol (F22), dressing yourself (F30), showering bathing (F31), walking 1 block on level ground (F32), and hurrying or jogging (F33). The 11 missing values are predicted by a regression model developed using the best 3 correlated features as input and the feature with missing value as output.

6.3 Ablation study on the impact of significant features

RFE has been implemented over the 10 classifiers with four different performance criteria to calculate permutation feature importance (PFI) namely, balanced accuracy, precision, recall, and F1-score. The performance among classifiers varies greatly because of their different and unique algorithm. As given in Table 2, all the classifiers show their best precision value greater than 0.95 with different numbers of selected features at the instance. The highest precision value of 0.98 is given by KNC using 4 features namely, F1, F2, F31, and F3 with ROS and MMS preprocessing techniques and recall as PFI performance criteria. GBC also shows its best precision of 0.97 with the same preprocessing techniques, F1-score as PFI performance criteria, and 4 features, namely F1, F2, F25, and F3. Both, the best recall and F1-score of each classifier, are greater than 0.8 with the highest recall of 0.93 for RFC with 11 features and the highest F1-score of 0.95 for KNC with 4 features. Although other classifiers perform better at their best, GNB yields the best performance when averaged over all feature subsets and preprocessing techniques. The best performance of GNB is achieved with RUS and no scaling technique with precision, recall, and F1-score of 0.90 each with 7 features.

The RFE result for ROS and MMS is shown in Fig. 6 and a peak at 4 features for KNC can be noted for precision, recall, and F1-score. From Table 2 and Fig. 6, it can be concluded that there is a lot of variation in the performance of different classifiers with differences in the number of features, preprocessing techniques, and PFI performance criteria. At the same time, the feature ranking also varies greatly. In order to identify the features affecting the patient's condition, an average of all the feature rankings is utilized. Figure 7 shows the violin plot of ranking of all features based on all classifiers and preprocessing techniques. A visual gap can be noted in the average feature rank of 10 with 5 features having an average rank below 10. Finally, gender (F4), diabetes mellitus (F5), Na (F12), HR (F14), and systolic blood pressure (F15) are highlighted as significant properties by RFE.

6.4 Heart failure prediction with selected features

The selected features from the averaged ranking of RFE are further used to classify the modality of the patients. Figure 8 shows the precision, recall, and F1-score of selected 10 classifiers using 5 features with different scaling and sampling techniques. From Fig. 8(a), it is evident that ABC with RUS and MMS yield the best performance amongst all classifiers (recall value of 0.763, precision value of 0.718, and F1 score of 0.649) when compared to the performance of RFC with RUS and MMS (recall value of 0.744, the precision value of 0.692, and F1 score of 0.675) and GBC (recall value of 0.6203, precision of 0.595 and F1 score of 0.5606). For RUS with SS, ABC (recall of 0.763, precision of 0.7187, and F1 score of 0.6491) outperforms other classifiers such as RFC (recall of 0.7180, precision of 0.6726 and F1 score of 0.6405) and GNB (recall of 0.6917, precision of 0.6545 and F1 score of 0.6060). Figure 8(c), it is seen that ABC outperforms other classifiers in the absence of standardization. Classification performance over selected features using different sampling and scaling with ten different classifiers is discussed in Fig. S1 in the supplementary section. The significance of the features selected are.

- Gender—When compared to heart failure in females, the incidence of heart failure is much higher in males.
- Diabetes- it has been observed in the last two decades that the prevalence of heart failure in diabetes is very high and the prognosis is worse [49].
- Sodium- Patients with heart failure are restricted with sodium intake as it may lead to comorbidities like hypertension, chronic kidney disease, stroke, and cardiovascular diseases.

Table 2 Classification results of machine learning models using a selected number of features

Best performance measure	Classifiers	Pre-processing techniques (Sampling, scaling)*	Number of features	Precision	Recall	F1-score	
Precision	DTC	No, MMS	15	0.9583	0.75	0.8115	
	RFC	ROS, SS	6	0.9714	0.8333	0.8852	
	GBC	ROS, MMS	4	0.9714	0.8333	0.8852	
	ABC	S, MMS	9	0.9714	0.8333	0.8852	
	ETC	No, MMS	14	0.9583	0.75	0.8116	
	SVC	No, SS	5	0.9583	0.75	0.8116	
	KNC	ROS, MMS	4	0.9852	0.9167	0.947	
	MLPC	No, MMS	11	0.9583	0.75	0.8116	
	GPC	ROS, SS	19	0.9714	0.8333	0.8853	
	GNB	S, No	10	0.9714	0.8333	0.8853	
	Recall	DTC	S, SS	3	0.803	0.803	0.803
		RFC	RUS, SS	11	0.8	0.9394	0.8427
		GBC	ROS, MMS	1	0.6764	0.8333	0.6608
		ABC	S, No	28	0.7963	0.8712	0.8258
ETC		RUS, MMS	10	0.75	0.9091	0.7833	
SVC		S, SS	7	0.7611	0.856	0.7936	
KNC		ROS, MMS	4	0.9852	0.9167	0.947	
MLPC		ROS, No	1	0.75	0.9091	0.7833	
GPC		RUS, SS	14	0.7963	0.8712	0.8259	
GNB		RUS, No	7	0.9015	0.9015	0.9015	
F1-score		DTC	No, MMS	15	0.9583	0.75	0.8116
		RFC	ROS, SS	6	0.9714	0.8333	0.8852
		GBC	ROS, MMS	4	0.9714	0.8333	0.8853
		ABC	S, MMS	9	0.9714	0.8333	0.8853
	ETC	ROS, MMS	2	0.8415	0.8863	0.8615	
	SVC	ROS, SS	9	0.8705	0.8181	0.8412	
	KNC	ROS, MMS	4	0.9852	0.9167	0.947	
	MLPC	ROS, MMS	16	0.8705	0.8181	0.8412	
	GPC	ROS, SS	19	0.9714	0.8333	0.8852	
	GNB	RUS, No	7	0.9015	0.9015	0.9015	

*Sampling- no sampling (No), random oversampling (ROS), random under sampling (RUS), and Synthetic Minority Oversampling Technique (SMOTE). Scaling- no scaling (No), standard scaling (SS) and min-max scaling (MMS)

- Heart rate- raised heart rate was postulated as a high risk for heart failure [50].
- Systolic blood pressure- was strongly associated with heart failure, its elevation was found to be a major risk factor in middle-aged men [51].

6.5 Prediction results of web application

Self-management mHealth applications are primarily developed to support medical diagnostic and treatment strategies using machine learning techniques by building prediction models. As a result of this trend, the demand for mHealth applications in the market is also increasing, ranging from fitness apps to medical apps for managing chronic conditions such as heart failure [52], diabetes [53],

hypertension [54], breast cancer [55], blood pressure monitoring [56] and lung diseases [57]. The proposed framework allows the mHealth application to develop solutions for improved decision-making based on the selected feature. As a result, cardiovascular disease diagnosis can be improved, the efficiency of cardiovascular disease preventative measures can be increased, and better cardiovascular disease self-management can be promoted. Figure 9 illustrates the fact that the user is required to provide some information in order to get the desired output. In response to a request from the Flask server, an explainable machine learning model calculates a predicted result when data is submitted via a mobile app. The result is then sent from the Flask server to the mobile app, which retrieves it and shows it to the user. Comparison of the best model using

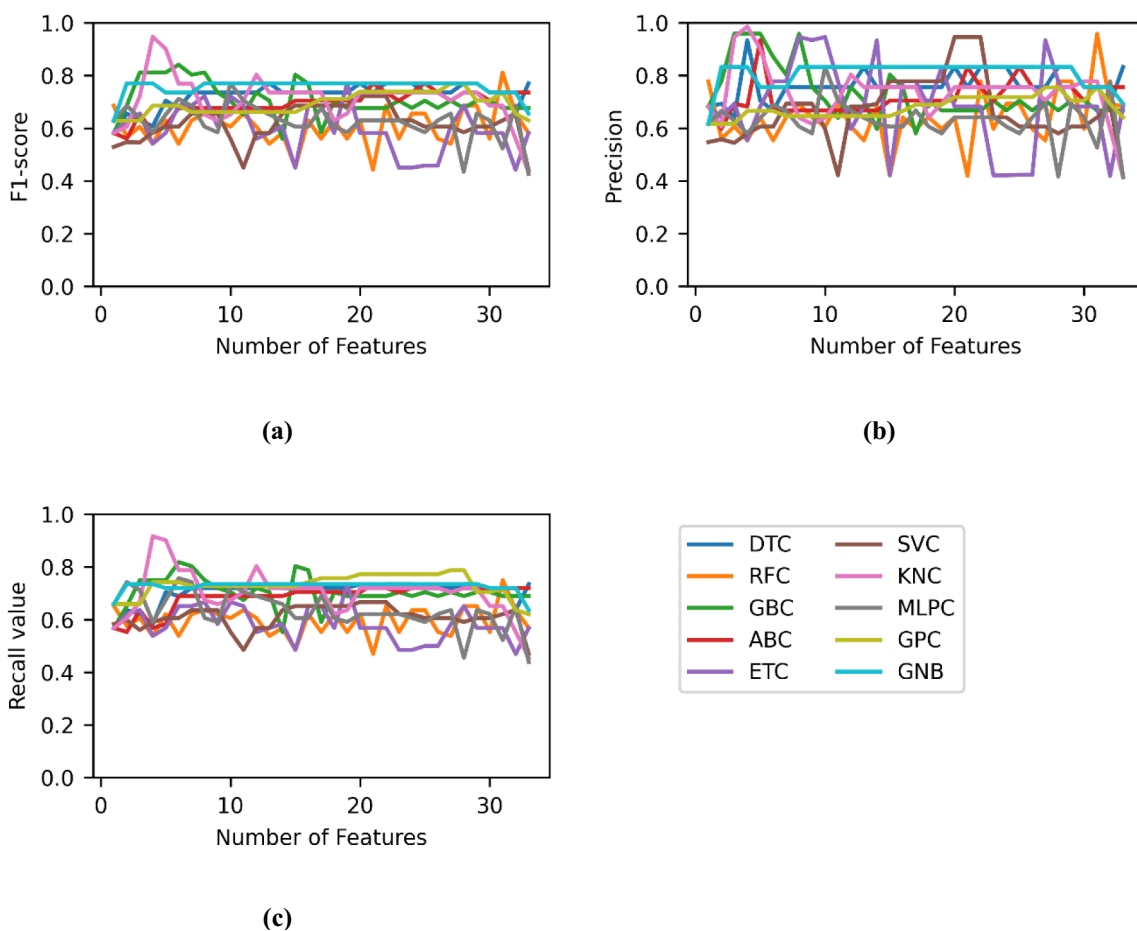


Fig. 6 Performance analysis of recursive feature extraction (a) F1 score, (b) Precision and (c) Recall for ROS and MMS preprocessing steps

the suggested feature elimination approach with various scaling and imbalance class handling, as well as comparison of performance metrics, allows for an accurate assessment of a classification model's efficacy.

The app development has been strongly influenced by factors including gender, diabetes, sodium, heart rate, and systolic blood pressure. Users will see a screen similar to Fig. 9 when the app is launched. The application is able to

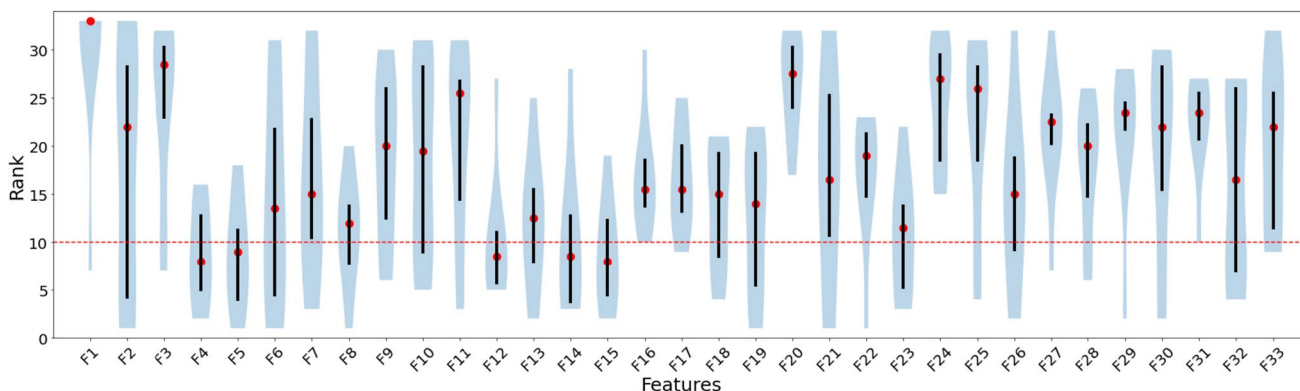
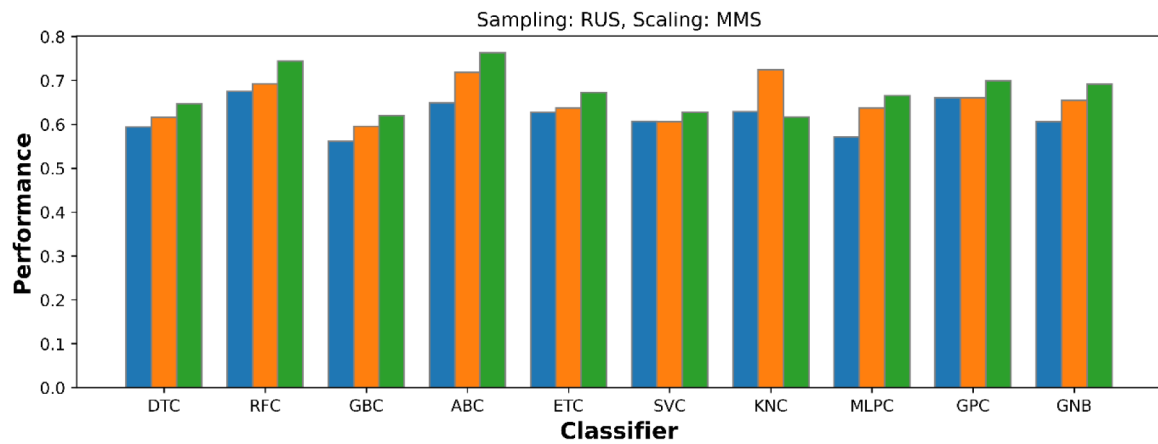
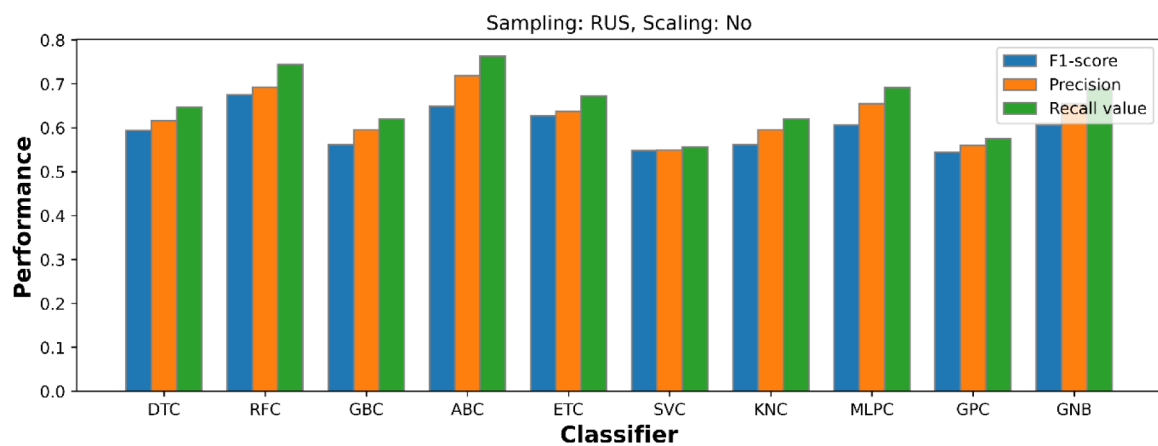


Fig. 7 Variation in feature rankings based on all RFE results. The red dot shows the median rank and the black line spans 25th to 75th percentile rankings. Red dotted lines are used as the threshold rank that

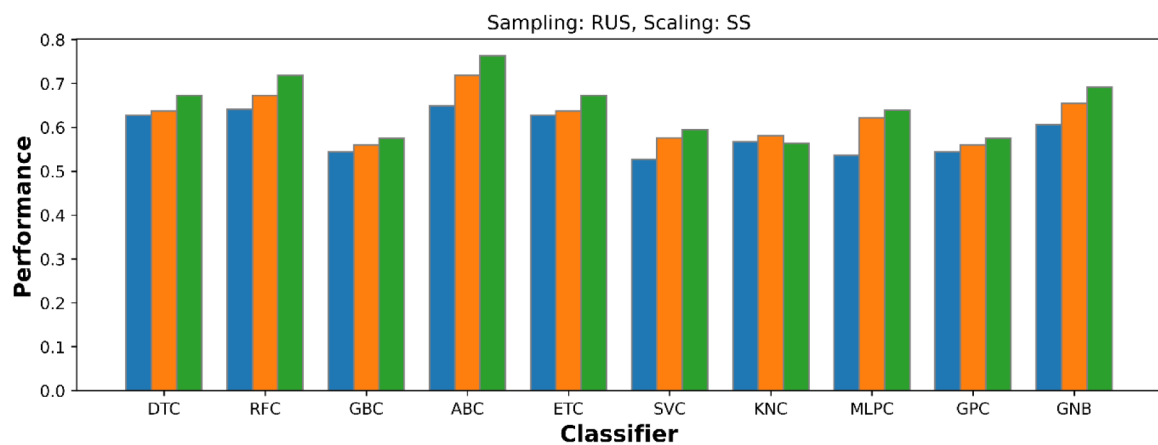
best separates the maximum average rank below the threshold and the minimum average rank above the threshold



(a)



(b)



(c)

Fig. 8 Classifier performance over selected features

verify that the input provided for each field is legitimate. A cautionary notice shown in Fig. 9(b), will appear on the screen for the user to read in the event that they input an

incorrect value for any of the parameters. If the user enters accurate data, the app will estimate the probability of heart failure, as shown in Fig. 9(c). By downloading and using this

Heart Failure Prediction

Gender (0: Male, 1: Female):

Diabetes (0: Yes, 1: No):

Systolic blood pressure (Range 50-200):

Heart rate (Range 40-200):

Sodium level (Range 120-145):

Predict

Heart Failure Prediction

Gender (0: Male, 1: Female):

Diabetes (0: Yes, 1: No):

Systolic blood pressure (Range 50-200):

Heart rate (Range 40-200):

Sodium level (Range 120-145):

Predict

Kindly check the sodium level input again. It should be between 120 and 145 only.

(a)

(b)

Heart Failure Prediction

Gender (0: Male, 1: Female):

Diabetes (0: Yes, 1: No):

Systolic blood pressure (Range 50-200):

Heart rate (Range 40-200):

Sodium level (Range 120-145):

Predict

Prediction result: The probability of heart failure is 0.722

(c)

Fig. 9 Web application for heart failure prediction: (a) user input validation, (b) error screen, and (c) prediction results

application on their mobile devices or desktop PCs, users of any device may get a heart failure risk assessment.

7 Conclusions

Despite recent medical technical advances, the number of deaths caused by heart failure hospitalization is continually rising, which may be related to woefully inadequate diagnostic tools. In order to reduce the risk of developing heart failure, there is a need for a system that is able to either produce rules or categorize the data via the use of machine learning strategies. Unlike the conventional ML approach that assumes an identical risk contribution for all the attributes of heart failure, the proposed method considered each of the attributes and can identify discrete subsets with a high discriminating capability. We tested the proposed technique using the AIIMS clinical dataset, which contains 151 samples from possible heart failure patients and found that the proposed algorithm outperformed the state-of-the-art methods in terms of accuracy, precision, recall, and F1-score. Comparison with different ML techniques including DTC, ABC, RFC, GPC, GBC, KNC, ETC, GNB, MLPC, and SVM has been presented. Further, ROS, RUS, and SMOTE were applied to deal with class imbalance problems, and furthermore, RFE was employed for feature ranking. According to RFE, the most significant features are Gender, Diabetes Mellitus, Na, hr, and systolic blood pressure. This study has the potential to advance medical practice and give clinicians a new resource for gauging heart failure patients' likelihood of survival. The performances of machine learning models have been compared on a full set of features and selected features from the heart failure clinical records dataset. ROS significantly improved the performance of KNC in predicting heart patient survival. ROS with MMS showed the highest result in all evaluation measures and achieved 0.9852 precision, 0.9167 recall, and 0.9470 F1-score. Any user may provide the clinical data to be analyzed by the web-based application, which has the ability to determine whether or not heart failure is present. In the future, this work might be extended to the classification of heart failure using an explainable machine learning technique, with the goal of defining the model's accuracy, transparency, and outcomes.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s12553-024-00832-z>.

Author contributions Deepak Kumar: Data Curation, Formal Analysis, Software, Writing—Original Draft, Keerthiveena B: Data Curation, Formal Analysis, Software, Writing—Original Draft, Sandeep Seth: Conceptualization, Methodology, Investigation, Formal Analysis, Validation, Shivani Vashista: Investigation, Formal analysis, Validation, Manojkumar Ramteke: Conceptualization, Methodology, Investigation, Writing—Review & Editing, Anurag S Rathore: Conceptualization, Methodology, Investigation, Writing—Original Draft, Writing—Review & Editing.

Funding The Department of Cardiology at All India Institute of Medical Sciences Delhi assisted in obtaining images for this project. This study was supported by the Department of Biotechnology (DBT) (BT/COE/34/ SP15097/2015, PI: ASR; BT/PR34224/AI/133/10/2019, PIs: ASR and SS) and Tower project (RP04636N, PI: ASR). This study used the high-performance computing facilities of IIT Delhi. The authors would like to express their gratitude to the Yardi School of Artificial Intelligence for their support.

Data availability The datasets from All India Institute of Medical Sciences (AIIMS) Delhi were approved for the current study and are not publicly available.

Code availability The codes generated during the current study are available from the corresponding author upon reasonable request.

Declarations

Ethical approval This study was approved by the Institute ethics committee of All India Institute of Medical Sciences.

Consent for publication We confirm that our research did not use humans or animals as experimental subjects.

Conflict on interests The authors declare that they have no competing interests.

References

- Kaptoge S, Pennells L, De Bacquer D, Cooney MT, Kavousi M, Stevens G, Riley LM, Savin S, Khan T, Altay S, Amouyel P. World Health Organization cardiovascular disease risk charts: revised models to estimate risk in 21 global regions. *Lancet Glob Health*. 2019;7(10):e1332–45.
- Amini M, Zayeri F, Salehi M. Trend analysis of cardiovascular disease mortality, incidence, and mortality-to-incidence ratio: results from global burden of disease study 2017. *BMC Public Health*. 2021;21(1):1–12.
- Ritchie H, Spooner F, Roser M. Causes of death. *Our world in data*. 2018. Published online. <https://ourworldindata.org/causes-of-death>
- Hosain N, Amin F, Maruf MF, Chowdhury MAQ, Chowdhury MR, Mahmud AU, Akter T, Anisuzzaman M, Rahim A. Global geographical discrepancy in numerical distribution of cardiovascular surgeries and human resource development in South Asia. *JTCVS open*. 2022;11:192–9.
- Kumar AS, Sinha N. Cardiovascular disease in India: a 360 degree overview. *Med J Armed Forces India*. 2020;76(1):1.
- Huffman MD, Prabhakaran D, Osmond C, Fall CH, Tandon N, Lakshmy R, Ramji S, Khalil A, Gera T, Prabhakaran P, Biswas SD. Incidence of cardiovascular risk factors in an Indian urban cohort: results from the New Delhi Birth Cohort. *J Am Coll Cardiol*. 2011;57(17):1765–74.
- Verma M, Jagia P, Roy A, Chaturvedi PK, Kumar S, Seth S, Singh V, Ojha V, Pandey NN. Lung water estimation on cardiac magnetic resonance imaging for predicting adverse cardiovascular outcomes in patients with heart failure. *Br J Radiol*. 2023;96(1146):20220723.
- Sadoughi F, Khodaveisi T, Ahmadi H. The used theories for the adoption of electronic health record: a systematic literature review. *Heal Technol*. 2019;9:383–400.
- Liaw ST, Georgiou A, Marin H. Evaluation of Digital Health & Information Technology in Primary Care. *Int J Med Informatics*. 2020;144:104285–104285.

10. Katarya R, Meena SK. Machine learning techniques for heart disease prediction: a comparative study and analysis. *Heal Technol*. 2021;11:87–97.
11. Mohan S, Thirumalai C, Srivastava G. Effective heart disease prediction using hybrid machine learning techniques. *IEEE access*. 2019;7:81542–54.
12. Jagadeeswari V, Subramaniaswamy V, Logesh R, Vijayakumar V. A study on medical Internet of Things and Big Data in personalized healthcare system. *Health Inf Sci Syst*. 2018;6:1–20.
13. Anuragi A, Sisodia DS, Pachori RB. Classification of focal and non-focal EEG signals using optimal geometrical features derived from a second-order difference plot of FBSE-EWT rhythms. *Artif Intell Med*. 2023;139: 102542.
14. Nourmohammadi-Khiarak J, Feizi-Derakhshi MR, Behrouzi K, Mazaheri S, Zamani-Harghalani Y, Tayebi RM. New hybrid method for heart disease diagnosis utilizing optimization algorithm in feature selection. *Heal Technol*. 2020;10:667–78.
15. Sakamoto T, TEI, C., MURAYAMA, M., ICHIYASU, H., HADA, Y., HAYASHI, T. and AMANO, K. Giant T wave inversion as a manifestation of asymmetrical apical hypertrophy (AAH) of the left ventricle echocardiographic and ultrasono-cardiotomographic study. *Jpn Heart J*. 1976;17(5):611–29.
16. George B, Seals S, Aban I. Survival analysis and regression models. *J Nucl Cardiol*. 2014;21(4):686–94.
17. Marx V. The big challenges of big data. *Nature*. 2013;498(7453):255–60.
18. Saheb T. An empirical investigation of the adoption of mobile health applications: integrating big data and social media services. *Heal Technol*. 2020;10(5):1063–77.
19. Athilingam P, Jenkins B. Mobile phone apps to support heart failure self-care management: integrative review. *JMIR cardio*. 2018;2(1): e10057.
20. Cruz-Ramos NA, Alor-Hernández G, Colombo-Mendoza LO, Sánchez-Cervantes JL, Rodríguez-Mazahua L, Guarneros-Nolasco LR. February MHealth apps for self-management of cardiovascular diseases: A scoping review. In *Healthcare*, vol. 10, No. 2. MDPI; 2022. p. 322.
21. Nikolaou K, Alkadhi H, Bamberg F, Leschka S, Wintersperger BJ. MRI and CT in the diagnosis of coronary artery disease: indications and applications. *Insights Imaging*. 2011;2(1):9–24.
22. Muhammad Y, Tahir M, Hayat M, Chong KT. Early and accurate detection and diagnosis of heart disease using intelligent computational model. *Sci Rep*. 2020;10(1):1–17.
23. Javaid M, Haleem A, Singh RP, Suman R, Rab S. Significance of machine learning in healthcare: features, pillars and applications. *Int J Intell Netw*. 2022;9:58–73. <https://doi.org/10.1016/j.ijin.2022.05.002>.
24. Shah A, Ahirrao S, Pandya S, Kotecha K, Rathod S. Smart cardiac framework for an early detection of cardiac arrest condition and risk. *Front Public Health*. 2021;9:762303. <https://doi.org/10.3389/fpubh.2021.762303>.
25. Buttar HS, Li T, Ravi N. Prevention of cardiovascular diseases: role of exercise, dietary interventions, obesity and smoking cessation. *Exp Clin Cardiol*. 2005;10(4):229–49.
26. Yang G, Ren Y, Pan Q, Ning G, Gong S, Cai G, Zhang Z, Li L, Yan J. October. A heart failure diagnosis model based on support vector machine. In 2010 3rd international conference on biomedical engineering and informatics. *IEEE*. 2010;3:1105–1108. <https://doi.org/10.1109/BMEI.2010.5639619>.
27. Gharehchopogh FS, Khalilifelu ZA. July Neural network application in diagnosis of patient: a case study. In *International Conference on Computer Networks and Information Technology*. *IEEE*; 2011. p. 245–9.
28. Alizadehsani R, Habibi J, Sani ZA, Mashayekhi H, Boghrati R, Ghandeharioun A, Khozeimeh F, Alizadeh-Sani F. Diagnosing coronary artery disease via data mining algorithms by considering laboratory and echocardiography features. *Research in Cardiovascular Medicine*. 2013;2(3):133.
29. Tama BA, Im S, Lee S. Improving an intelligent detection system for coronary heart disease using a two-tier classifier ensemble. *Biomed Res Int*. 2020; 2020; Article ID 9816142. <https://doi.org/10.1155/2020/9816142>.
30. Parthiban G, Srivatsa SK. Applying machine learning methods in diagnosing heart disease for diabetic patients. *Int J Appl Inf Syst*. 2012;3(7):25–30.
31. Hernandez AF, Greiner MA, Fonarow GC, Hammill BG, Heidenreich PA, Yancy CW, Peterson ED, Curtis LH. Relationship between early physician follow-up and 30-day readmission among Medicare beneficiaries hospitalized for heart failure. *JAMA*. 2010;303(17):1716–22.
32. Shah NH, Milstein A, Bagley SC. Making machine learning models clinically useful. *JAMA*. 2019;322(14):1351–2.
33. Gerke S, Minssen T, Cohen G. Ethical and legal challenges of artificial intelligence-driven healthcare. In *Artificial intelligence in healthcare*. Academic Press; 2020. pp. 295–336. <https://doi.org/10.1016/B978-0-12-818438-7.00012-5>.
34. Tan M, Tsang IW, Wang L. Towards ultrahigh dimensional feature selection for big data. *J Mach Learn Res*. 2014;15:1371–429.
35. Leevy JL, Khoshgoftaar TM, Bauder RA, Seliya N. A survey on addressing high-class imbalance in big data. *J Big Data*. 2018;5(1):1–30.
36. Richhariya B, Tanveer M, Rashid AH, Alzheimer's Disease Neuroimaging Initiative. Diagnosis of Alzheimer's disease using universum support vector machine based recursive feature elimination (USVM-RFE). *Biomed Signal Process Control*. 2020;59:101903. <https://doi.org/10.1016/j.bspc.2020.101903>.
37. Guyon I, Weston J, Barnhill S, Vapnik V. Gene selection for cancer classification using support vector machines. *Mach Learn*. 2002;46:389–422. <https://doi.org/10.1023/A:1012487302797>.
38. Yin Z, Zhang J. Operator functional state classification using least-square support vector machine based recursive feature elimination technique. *Comput Methods Programs Biomed*. 2014;113(1):101–15.
39. Mayerich DM, Walsh M, Kadjacsy-Balla A, Mittal S, Bhargava R. March Breast histopathology using random decision forests-based classification of infrared spectroscopic imaging data. In: *In Medical Imaging 2014: Digital Pathology*. 2014;9041:38–44. <https://doi.org/10.1117/12.2043783>.
40. Maniruzzaman M, Rahman MJ, Ahammed B, Abedin MM. Classification and prediction of diabetes disease using machine learning paradigm. *Health Inf Sci Syst*. 2020;8(7):1–14.
41. Masetic Z, Subasi A. Congestive heart failure detection using random forest classifier. *Comput Methods Programs Biomed*. 2016;130:54–64.
42. Kim HC, Ghahramani Z. Bayesian Gaussian process classification with the EM-EP algorithm. *IEEE Trans Pattern Anal Mach Intell*. 2006;28(12):1948–59.
43. Son J, Jung I, Park K, Han B. Trackingmentation with online gradient boosting decision tree. In: *In Proceedings of the IEEE international conference on computer vision*. 2015. p. 3056–64.
44. Abu Alfeilat HA, Hassanat AB, Lasassmeh O, Tarawneh AS, Alhasan MB, Eyal Salman HS, Prasath VS. Effects of distance measure choice on k-nearest neighbor classifier performance: a review. *Big data*. 2019;7(4):221–48.
45. Sharma D, Kumar R, Jain A. Breast cancer prediction based on neural networks and extra tree classifier using feature ensemble learning. In: *Measurement: Sensors*. 24. 2022. p. 100560.
46. Jahromi AH, Taheri M. October. A non-parametric mixture of Gaussian naive Bayes classifiers based on local independent features. In: *In 2017 Artificial intelligence and signal processing conference (AISP)*. *IEEE*; 2017. p. 209–12.

47. Wan S, Liang Y, Zhang Y, Guizani M. Deep multi-layer perceptron classifier for behavior analysis to estimate Parkinson's disease severity using smartphones. *IEEE Access*. 2018;6:36825–33.
48. Keerthiveena B, Esakkirajan S, Subudhi BN, Veerakumar T. A hybrid BPSO-SVM for feature selection and classification of ocular health. *IET Image Proc*. 2021;15(2):542–55.
49. Lehrke M, Marx N. Diabetes Mellitus and Heart Failure. In: *The American Journal of Cardiology*. 2017. p. S37–47.
50. Böhm M, Swedberg K, Komajda M, Borer JS, Ford I, Dubost-Brama A, Lerebours G, Tavazzi L. Heart rate as a risk factor in chronic heart failure (SHIFT): the association between heart rate and outcomes in a randomised placebo-controlled trial. *The Lancet*. 2010;376(9744):886–94.
51. Stamler J, Neaton JD, Wentworth DN. Blood pressure (systolic and diastolic) and risk of fatal coronary heart disease. *Hypertension*. 1989;13(5_supplement):I2.
52. Oronti IB, Lopez-Perez L, Piaggio D, Fico G, Pecchia L. Effects of home telemonitoring using non-invasive versus conventional approaches on patients affected by congestive heart failure: a systematic literature review and meta-analysis. In: *Health and Technology*. 2023. p. 1–21.
53. Krmptic K, Gallant JR, Zufelt K, Zuijdwijk C. User-centred development of an mHealth app for youth with type 1 diabetes: The challenge of operationalizing desired features and feasibility of offering financial incentives. *Heal Technol*. 2022;12(2):499–513.
54. Albini F, Liu X, Torlasco C, Soranna D, Faini A, Ciminaghi R, Celsi A, Benedetti M, Zambon A, Di Rienzo M, Parati G. August. An ICT and mobile health integrated approach to optimize patients' education on hypertension and its management by physicians: The Patients Optimal Strategy of Treatment (POST) pilot study. In: *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE. 2016;517–20. <https://doi.org/10.1109/EMBC.2016.7590753>.
55. Andrzejewski D, Breitschwerdt R, Fellmann M, Beck E. Supporting breast cancer decisions using formalized guidelines and experts decision patterns: initial prototype and evaluation. *Health Information Science and Systems*. 2017;5:1–10.
56. Kaplan AL, Cohen ER, Zimlichman E. Improving patient engagement in self-measured blood pressure monitoring using a mobile health technology. *Health Inf Sci Syst*. 2017;5:1–9.
57. Amiri P, Ahmadian L, Khajouei R. The applications and the effectiveness of mHealth interventions to manage lung cancer patients: a systematic review. *Heal Technol*. 2023;13(2):171–80.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.