Received: 20 July 2023 | Revised: 28 August 2023 | Accepted: 14 September 2023 | Published online: 20 September 2023

Journal of Data Science and Intelligent Systems
2024, Vol. 2(3) 161–173
DOI: 10.47852/bonviewJDSIS32021386

RESEARCH ARTICLE

Identifying Risk Factors for Heart Failure: A Case Study Employing Data Mining Algorithms

Vitória S. Souza[1],* and **Danielli A. Lima**[1]

[1]*Laboratory of Intelligent Computing and Robotics, Federal Institute of Triangulo Mineiro Campus Patrocínio, Brazil*

Abstract: Heart diseases are increasingly present in the lives of human beings and are diseases that affect the heart and blood vessels and can lead the person who develops to death. In this article, we analyzed an open and public database on heart failure composed of a sample of 299 people and 12 attributes. This article presents a preprocessing technique using area under the curve (AUC) filters, which increases the efficiency of the algorithms by decreasing the parameters, leading to better memory usage and computational processing. To enhance our results, we used a methodology involving 10^2 simultaneous validations. This approach allowed us to obtain more robust and reliable results. In addition, we used the receiver operating characteristic curve to evaluate the overall performance of each attribute. We trained a set of nine classification algorithms, among which the random forest learner stood out with an accuracy of 87.21% when using a filter that considered attributes with AUC greater than 0.4, considering values of AUC. Additionally, the fuzzy rules learner demonstrated its effectiveness by achieving an accuracy of 84.45% with a filter limit of 0.6, focusing on ejection fraction, serum sodium, time attributes, and class for death events. This analysis demonstrated the ability of these algorithms to effectively use a reduced number of attributes for accurate predictions.

Keywords: data mining, machine learning, cardiology, heart failure, receive operating characteristic curve, artificial intelligence, random forest learner

1. Introduction

The human body is made up of several organs, and although the exact number may vary depending on how the organs are defined, it is estimated that there are approximately around 78 major organs, including internal and external organs [1]. But, as science advances, this number may change. Among human organs, the heart stands out for its vital importance, being responsible for pumping blood and supplying essential oxygen and nutrients to cells and organs [1]. Given its importance, it requires careful attention due to various diseases that, if left untreated, can lead to death. These diseases may result from poor habits or heredity. Hence, reliable and accurate diagnostic techniques are crucial for their treatment and risk factor identification. A healthy diet, and avoiding obesity, smoking, and excessive alcohol consumption are essential to prevent heart disease [2, 3]. Managing stress and engaging in regular physical activity also support heart health. Medical conditions like hypertension, diabetes, and high cholesterol can contribute to cardiac insufficiency [4].

There are several heart diseases, but some of them are high blood pressure, which represents high levels of blood pressure in the arteries [5], heart failure where the heart becomes unable to meet the body's needs causing restriction of blood flow [6], and acute myocardial infarction, which is known as myocardial necrosis resulting from the obstruction of a coronary artery, among others. Because of this, it is important that the cardiologist is visited regularly. This allows for early detection and management of any potential risk factors or underlying conditions that could contribute to cardiac insufficiency. When we carry out the application of artificial intelligence (AI) in medicine, together with the analysis of medical specialists, we can discover new means of treatments and discover new diseases [7, 8]. By analyzing a database of patients with heart failure and together applying machine learning (ML) algorithms to identify the possibility of survival of patients with heart failure, we can discover which factors are most significant for the diagnosis, using data science and supported by experts. Data mining (DM) has been widely used to identify patterns and build models to predict heart disease.

Classification is among the most commonly used techniques to predict the occurrence of heart disease [9, 10]. There are several classification algorithms that can be applied in this circumstance, but choosing the best one can be challenging because of the complexity of the data and the nature of the problem. In this work, we apply nine [11] classification algorithms, namely decision tree (DT), support vector machine (SVM), probabilistic neural networks (PNNs), multilayer perceptron (MLP), naive Bayes learner (NBL), gradient boosting learner (GBL), K-nearest neighbors (KNNs), random forest learner (RFL), and fuzzy rules (FR) [12, 13], being different supervised machine learning techniques emphasizing the classification of heart disease in relation to survival or death.

---

*Corresponding author: Vitória S. Souza, Laboratory of Intelligent Computing and Robotics, Federal Institute of Triangulo Mineiro Campus Patrocínio, Brazil. Email: vitoria.stefane@estudante.iftm.edu.br

Herein, we aim that with the use of classification algorithms and various ML techniques, together with the correlation of the applicability of the receiver operating characteristic (ROC) curve, we can identify which attributes are most present in predicting the survival of people with heart failure. The area under the ROC curve (AUC) is often used as a metric to quantify the overall performance of the model, with values closer to 1 indicating better discrimination ability. The main innovation of this article lies in our unique approach to preprocessing using filters based on area under curve metric. This approach enhances algorithm performance in terms of memory and computational processing by allowing the utilization of a reduced number of parameters. With the choice of the best techniques and the best algorithm, we can help reduce costs with unnecessary exams and we can contribute to the advancement of DM in health. We aim to (i) compare the results of the nine algorithms applied to the open and public database, (ii) analyze the other metrics identifying the best one, (iii) apply data visualizations so that we can have a better understanding of the behavior of the attributes of the base, (iv) apply the ROC curve to identify the samples that may be more relevant to help in the diagnosis of heart failure survival, and (v) compare and discuss which of the attributes and metrics were best performed in the database.

2. Related Works

Ahmad et al. [14] studied a segment of Pakistan's heart failure population, determining survival and mortality rates. They used 200 bootstrap replications, an ROC curve, a Cox model, and a nomogram. Age, ejection fraction, sodium, anemia, high blood pressure, and creatine were found significant. ROC analysis revealed 81% identification for death events in longer follow-ups and 77% in shorter periods, highlighting risk stratification's importance using clinical and lab factors for higher-risk heart failure patient identification.

Chicco and Jurman [9] already cover only two clinical factors to predict the survival of patients with heart failure, using the same database of the previous study and of this same study, namely, serum creatine and ejection fraction, where the authors based it on the construction of the ML model, where they had the employability of 10 different methods. Hasan et al. [15] concluded that the two attributes of ejection fraction and serum creatine were also more relevant data for predicting heart failure survival. Among the five classifiers used, the DT was the one that provided the best results, with 80%. In Ishaq et al. [16], several classification models were evaluated, including DT, adaptive boost classifier, and logistic regression, among others. The additive tree (ETC) was accurate to 92.62% in predicting the survival of cardiac patients using SMOTE. However, our study seeks to improve the prediction using the ROC curve and filters to reduce attributes and improve relevance.

In the study conducted by Gürfidan and Ersoy [17], they focused on investigating mortality related to heart failure using ML algorithms. The algorithms employed included SVM, logistic regression, DT, KNN, linear discriminant analysis, and Gaussian naive Bayes. With SVM, they had a higher success rate with 83% in the standings. Muntasir Nishat et al. [18] analyze a total of six ML algorithms in a set of 299 individuals with heart failure. The random forest classifier (RFC) in combination with techniques such as SMOTE-ENN, along with balancing and scaling techniques, achieved a test accuracy of 90%, highlighting its effectiveness in predicting survival. Emphasizing the importance of SMOTE-ENN and hyperparameter optimization to improve algorithm performance.

3. Theoretical Foundation

In this section, the definition of heart failure will be presented, along with the description of the database used, its attributes, and the respective related works.

3.1. Heart anatomy

As a vital organ for all of us, the heart is a hollow muscular organ located in the central region of the chest responsible for distributing blood throughout our body [19]. This organ is divided into four main chambers: two right atria located in the upper part and two ventricles located in the lower part [20]. Its main function is to receive venous blood rich in carbon dioxide and pump the blood to the lungs, where oxygenation takes place.

Its anatomy is composed of some structures as shown in Figure 1 [21], such as the superior vena cava, the main vein that transports blood back to the heart [22], the Aorta, which is the largest and most important artery in the human body, responsible for distributing oxygenated blood throughout the body [22], the pulmonary arteries, which are responsible for transporting deoxygenated blood to the lungs where it is oxygenated [22], pulmonary veins are blood vessels that carry oxygen-rich blood from the lungs to the left atrium of the heart [22], the left atrium, on the other hand, will receive the oxygenated blood that leaves the lungs [22], the left ventricle is one of the lower chambers of the heart, which receives blood from the left atrium and pumps this blood to the arteries, specifically to the aorta that distributes blood to the body [22], cardiac muscle or myocardium is the muscle tissue that makes up the wall of the heart, responsible for pumping blood through the circulatory system [22], right ventricle is another lower chamber that receives deoxygenated blood and pumps it to the lungs through the pulmonary arteries [22], the inferior vena cava collects deoxygenated blood from the lower body; the right atrium receives deoxygenated blood from systemic veins, such as the inferior and superior vena cava and pumps it into the right ventricle [22]. The heart also has an intrinsic

Figure 1
Heart anatomy

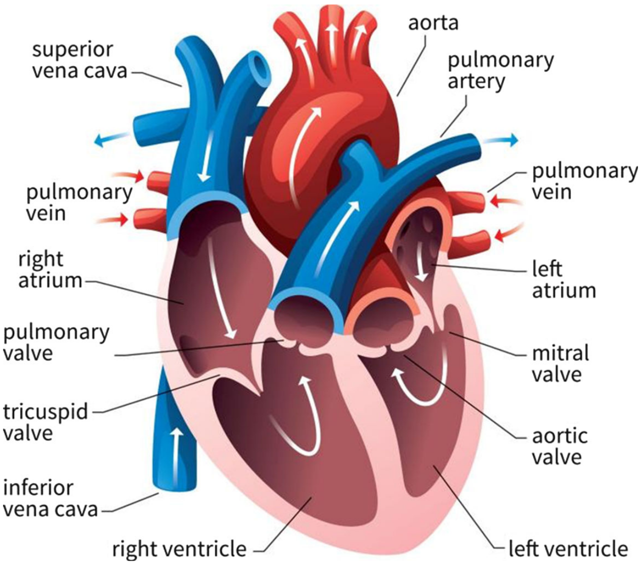


Table 1
Explanation of feature and measurement range

Resources	Explanation	Measurement	Range	Median	Mean
Age	Patient's age	Years	[40, 95]	60	64.43
Anemia	Decreased red blood cells or hemoglobin	Boolean	[0, 1]	0	0.43
High pressure	If the patient has hypertension	Boolean	[0, 1]	0	0.43
Creatine phosphokinase	CPK enzyme level in the blood	mcg/L	[23, 7861]	250	1438.29
Diabetes	If the patient has diabetes	Boolean	[0, 1]	0	0.43
Ejection fraction	Blood ejection from the heart	Percentage	[14, 80]	38	40.86
Sex	Woman or man	Binary	[0, 1]	1	0.57
Platelets	Blood platelets	kiloplatelets/ml	[25.01, 850.00]	262000	309585.7
Serum creatinine	Blood creatinine level	mg/dL	[0.50, 9.40]	1.1	2.27
Serum sodium	Blood sodium level	mEq/L	[114, 148]	137	135
Smoker	If the patient smokes	Boolean	[0, 1]	0	0.43
Time	Follow-up period	Days	[4,285]	113	138.43
Death event (class)	Patient's mortality during follow-up	Boolean	[0, 1]	0	0.32

electrical system that coordinates heartbeats, the sinoatrial node better known as the heart's natural pacemaker. Regarding blood circulation, there are two main circuits: pulmonary circulation and systemic circulation [22].

Heart failure is a disease that affects the heart and is characterized by a condition in which the heart is unable to pump blood effectively to meet the body's needs. As a result, fluid can accumulate in the legs, lungs, and other tissues throughout the body. Heart failure can occur due to several factors, some of which include other diseases such as coronary artery disease, acute myocardial infarction, high blood pressure, and diabetes, among others. Its signs are not always noticed in the initial phase, and some people decide to ignore it or confuse it with just tiredness. But sometimes the symptoms are more evident, such as shortness of breath, swelling in the legs and feet, tiredness, confusion, and impaired memory, among others. The disease may develop more in certain people in particular due to some risk factors such as high blood pressure, heart attack, diabetes, family history of heart disease, cardiomyopathy, and abnormal heart valve [15, 16].

In order to verify the presence of heart problems and monitor heart failure, different parameters and tests are used: (a) cardiac function tests: Electrocardiogram, echocardiogram, stress test (treadmill or exercise bike), and cardiac scintigraphy are performed to assess heart function, heart structure, blood flow, and the heart's ability to respond to physical exertion. (b) Biomarkers: Blood tests, such as measuring levels of biomarkers, for example, brain natriuretic peptide, are performed to aid in the detection and monitoring of heart failure. (c) Monitoring signs and symptoms: Monitoring the patient's symptoms, such as shortness of breath, fatigue, leg swelling, and sudden weight gain, can help identify potential heart problems [15, 16]. Therefore, seeking adequate medical follow-up to carry out an early treatment is efficient to prevent the disease or treat it as soon as possible if it has already developed.

3.2. Database

Aiming at a better understanding of heart disease, specifically heart failure, and how data science can influence it, together with the analysis of common datasets and techniques, can enable a more effective analysis of which factors or attributes may be most important to diagnose survival of patients with heart failure. We use an open and public database taken from the UCI Machine Learning Repository called "heart failure clinical records dataset", consisting of a sample of 299 patients admitted to the Institute of Cardiology and Allied hospital Faisalabad-Pakistan in 2015 [14]. All patients had previous heart failure, were between 40 and 95 years old, 104 women and 195 men, and composed of 12 attributes and a classification class, namely DEATH_EVENT.

In our study, we used a dataset of 299 patients and 13 attributes, where both patients had left ventricular systolic dysfunction and previous heart failure. Among the attributes present, some may or may not be relevant to classify the survival of patients with heart failure, but all of them are shown in Table 1 [9].

Among them, we have the **class** DEATH_EVENT which is the class that classifies whether the patient died – (1) or survived – (0). The other attributes are as follows: **sex**, which indicates the patient's gender (female – 0, male – 1), and **anemia**, which refers to a decrease in red blood cells or hemoglobin. According to Chicco and Jurman [9], patients with hematocrit levels below 36% were considered anemic by the doctors.

Creatine phosphokinase (CPK) is an enzyme that can indicate heart disease when it is found in the blood, as it leaks into the bloodstream when there is damage to muscle tissue. **Serum creatine** with the level of serum creatine in the blood doctors assess kidney function [9]. **Ejection fraction** is the amount of blood the heart releases with each contraction. **Age** is the age of the patients; **Diabetes** whether or not the patient has diabetes caused by insufficient insulin production. High blood pressure represents whether or not the patient has hypertension; **platelets,** also known as thrombocytes, depicts blood platelets. **Smoking** shows whether the patient is a smoker or not. **Time** is the follow-up period of the research, from 4 to 285 days, and **serum sodium** mineral that helps in the good performance of the muscles and nerves and its test indicates if the patient has different levels of sodium in the blood. With supervised learning, we can train different algorithms to achieve better results through well-labeled data and predefined results.

4. Methodology

Our methodology for this article is quantitative, analyzing clinical records of heart failure from the UCI Machine Learning

Repository platform. It is an applied study, employing classification algorithms on a specific database. The methods used involved 10^2 experiments to improve accuracy. Explanatory in nature, we used column filters to identify essential attributes for heart disease classification, particularly heart failure.

4.1. Flowchart for ML proposal

Figure 2 presents a flowchart for the development of the proposal used in this study. First, we date the use of the KNIME Analytics Platform, which is an open-source DM platform that allows data analysis and visualization. First, data were collected from the UCI Machine Learning Repository platform containing 299 participants and 13 attributes and transformed into Excel.xlsx. In the preprocessing of the data, some numeric data were transformed into a string where the string replacer was performed to make the substitutions. Later, age lines with inconsistent values were excluded. Then, we normalized each of the data so that the values did not stand out from the others, and then we applied column filtering so that we could apply the filters to the columns of the database. Furthermore, we apply the X-Partitioner, which represents a cross-validation loop, where all algorithms applied to it are executed as many times as stipulated; in this study, we apply a total of ($X = 10^2$) iterations, and at the end of the loop there will be an X-Aggregator to collect the results of each iteration. Therefore, we chose DM algorithms for learning and prediction and we used nine algorithms, namely:

Decision tree: supervised learning algorithm that classifies data by recursively dividing attributes into a subset, forming a DT. Each leaf node represents class labels, representing the instance classification process [23].

Gradient boosting learner: learning approach that sequentially builds weak learners, (like DTs), correcting the mistakes of previous ones to form an accurate and robust model. It combines the individual predictions of weak learners to generate the final prediction [24].

K-nearest neighbors: a simple and effective algorithm used for classification and regression tasks. It finds the K nearest data points to a new input instance and classifies the instance based on the majority class of its neighbors. KNN is non-parametric and instance-based [11].

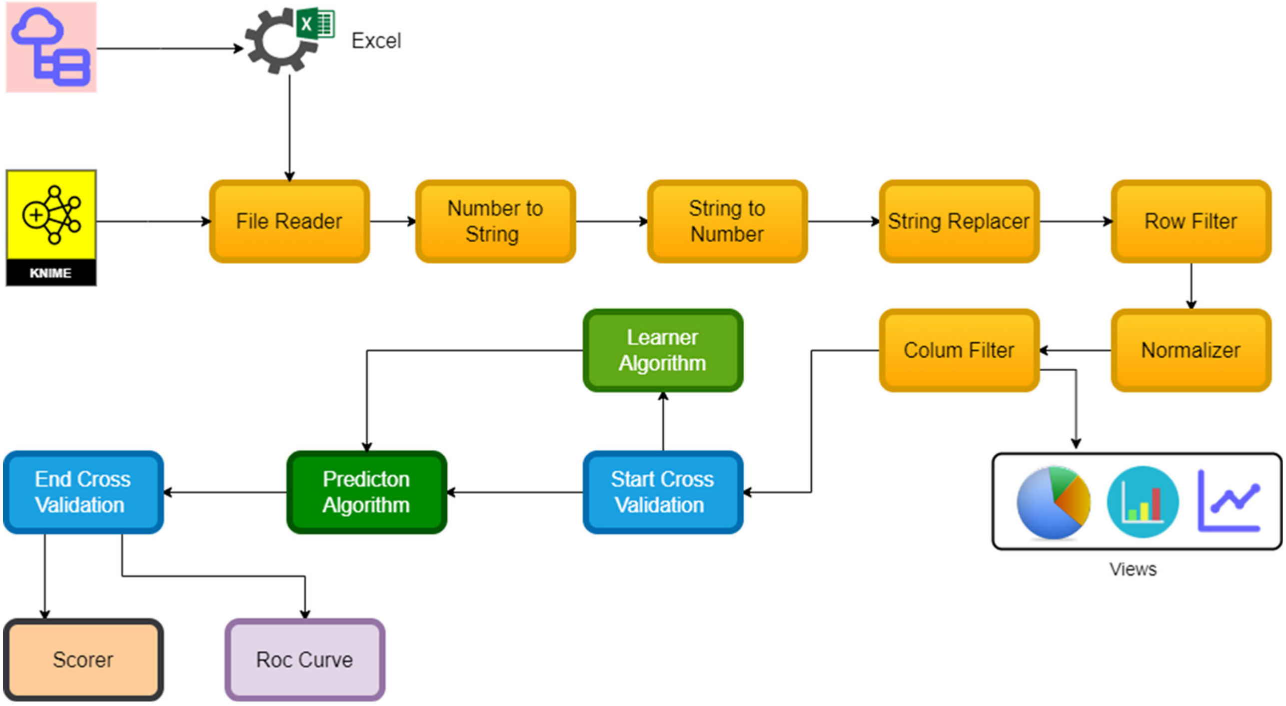
Multilayer perceptron: a feedforward artificial neural network with multiple layers of inter-connected neurons, including an input layer, one or more hidden layers, and an output layer. Neurons apply nonlinear activation functions to approximate complex functions and perform well in various tasks [11].

Naive Bayes learner: a probabilistic classifier based on Bayes' theorem, assuming conditional independence of features given the class label, simplifying posterior probability computation. Despite its naive assumption, it performs well in practice, especially in text classification tasks [8].

Probabilistic neural network: a feedforward neural network using Gaussian radial basis functions in the hidden layer, ideal for pattern recognition and classification tasks with high-dimensional data [8].

Random forest learner: ensemble learning with multiple DTs, aggregating predictions to enhance accuracy and reduce overfitting. It randomly selects feature subsets for each tree [23].

Figure 2
Proposed flowchart for machine learning application using ROC curve and AUC techniques



Support vector machine: powerful supervised learning for classification and regression. Finds optimal hyperplane to separate different classes in feature space. Effective in high-dimensional spaces, handling linear and nonlinear classification [12].

Fuzzy rules: used in fuzzy logic systems to model complex relationships between input and output variables. Expressed in linguistic terms and handling imprecise data, suitable for systems with human-like reasoning [4].

We employ the ROC curve to optimize the pre-filtering of DM algorithms by analyzing the performance of the classifier with changes in the discrimination threshold. In order to select the most relevant attributes for each respective value of AUC [25].

4.2. Steps for learning

The methodology proposed in this study involves the applicability of nine ML algorithms in which both are executed for 100 simulations and with variance filtering columns so that the accuracy of each one of them can be evaluated. The methodology to be proposed is the following: (1) perform data preprocessing, such as cleaning and normalization; (2) divide them into a training set corresponding to 70% and test 30% using cross-validation techniques with $X = 10^2$; (3) apply the nine ML algorithms (KNN, DT, PNN, SVM, random forest (RF), naive Bayes (NB), MLP, gradient boosting (GB), and FR); (4) employ the ROC curve along with the AUC analysis and evaluate the accuracy of the models; (5) calculate the means and standard deviations of the evaluation metrics (accuracy, sensitivity, Cohen's kappa, and *f* measure) for each algorithm over the 100 simulations; and (6) compare the results to obtain the best performing algorithm for classification.

4.2.1. Receiver operating characteristic curve

The ROC curve is characterized by being a tool that is widely used to evaluate and compare the performance of classification models for ML [26]. The ROC curve can also be used in various fields of knowledge, such as the health area [25]. The ROC curve is characterized by a graph that is constructed by plotting the true positive (TP) rate on the *y* axis (sensitivity) and relating the false positive (FP) rate (1-specificity) on the *x* axis, for different principles classification. The graph that is created allows viewing the specific effectiveness of the model, showing the variance of the rates as the cuts are performed. A good model has a curve closer to the upper left corner with high sensitivity and low specificity, i.e., FP rate.

4.2.2. Area under the curve

The AUC can summarize the ROC in only one value, gathering all the ROC thresholds in a single value and performing the calculation of the AUC [27]. The AUC has a variance between 0.0 and 1.0 with a threshold of 0.5 [28]. A value greater than the threshold may represent a better model for one class and a value below the threshold for another. The higher the AUC, the better the importance of the model for the classification. Models close to 0.0 are considered to be wrong predictions, and a model that has an AUC close to 1.0 represents 100% correct predictions. Each database attribute has a different AUC value for the two classes, respectively.

4.3. Evaluation metrics

We will present here, which are the fundamental measures to understand and carry out the identification of algorithms in relation to better performance.

4.3.1. Accuracy

This metric can be measured using four factors [29], namely FP when the result is negative but classified as positive, false negative (FN) when the result is positive but classified as negative, TP when the result is really true, and true negative (TN) when the result is really negative. It is possible to calculate the accuracy for the classification using Equation (1), with a final value between [0, 1].

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

Already the errors are the cases where the algorithm could not get it right. The error can be calculated taking into account the difference of the total minus the accuracy value by Equation (2).

$$E = 1.0 - A \tag{2}$$

4.3.2. Cohen's kappa

Statistical metric is used to evaluate the agreement between two or more notes or classifications [10]. This metric is very useful when you have more than two possible labels and when you do not have a uniform class distribution. Cohen's kappa (k) is calculated as the proportion of observed agreement minus expected agreement divided by the difference between 1 and expected agreement. Ranging from $(-1 \leq \kappa \leq 1)$, where (1) represents a perfect agreement, (0) a random agreement, and (-1) a perfect disagreement. This metric can be calculated by Equation (3).

$$\kappa = \frac{P_o - P_e}{1 - P_e} \tag{3}$$

4.3.3. Sensitivity and specificity

Sensitivity represents the proportion of positive examples that were correctly identified by the Sidey-Gibbons and Sidey-Gibbons [30] model. Specificity measures the proportion of negative examples that were correctly identified by the Sidey-Gibbons and Sidey-Gibbons [30] model. To calculate the sensitivity, use Equation (4).

$$Se = \frac{TP}{(TP + FN)} \tag{4}$$

Specificity can be given by Equation (5).

$$Sp = \frac{TN}{(TN + FP)} \tag{5}$$

When evaluating binary classification algorithms, it is very important to consider these two metrics.

4.3.4. Recall and precision

Recall (TP rate) measures the proportion of TPs that were correctly identified by the model in relation to the total number of positive samples (TP+FN) [31]. It then measures the model's ability to correctly identify all positive cases and can be calculated by Equation (6).

$$Rec = \frac{TP}{(TP + FN)} \tag{6}$$

Precision (positive predictive value) measures the proportion of TPs that were correctly identified by the model in relation to the total number of samples that were classified as positive (TP+FP). The formula for precision can be given by Equation (7).

$$Prc = \frac{TP}{(TP + FP)} \tag{7}$$

All of them vary with values between 0 and 1; the results that are closer to 1 indicate a more accurate and effective model.

4.3.5. F-measure

Also known as $F1$ Score, it combines precision and recall into a single metric, thus providing an overall assessment of the effectiveness of the binary classification model [32]. Its formula is represented by Equation (8). This measure varies from 0 to 1 where the higher the value, the better the performance of the model.

$$F1 = \frac{2 \times (Prc \times Rec)}{(Prc + Rec)} \tag{8}$$

5. Proposal Development

This section presents the proposed proposal and the KNIME Analytics platform workflow. The workflow represents the modules used for data processing and analysis, along with the classification algorithms employed. DM is a valuable tool for early heart disease detection and treatment discovery. Hence, selecting appropriate ML algorithms is crucial for accurate results. The nine algorithms used in this work are known for producing precise predictive models in heart disease prediction. Figure 3

Figure 3
Workflow for transformation and comparison analysis between nine classification models

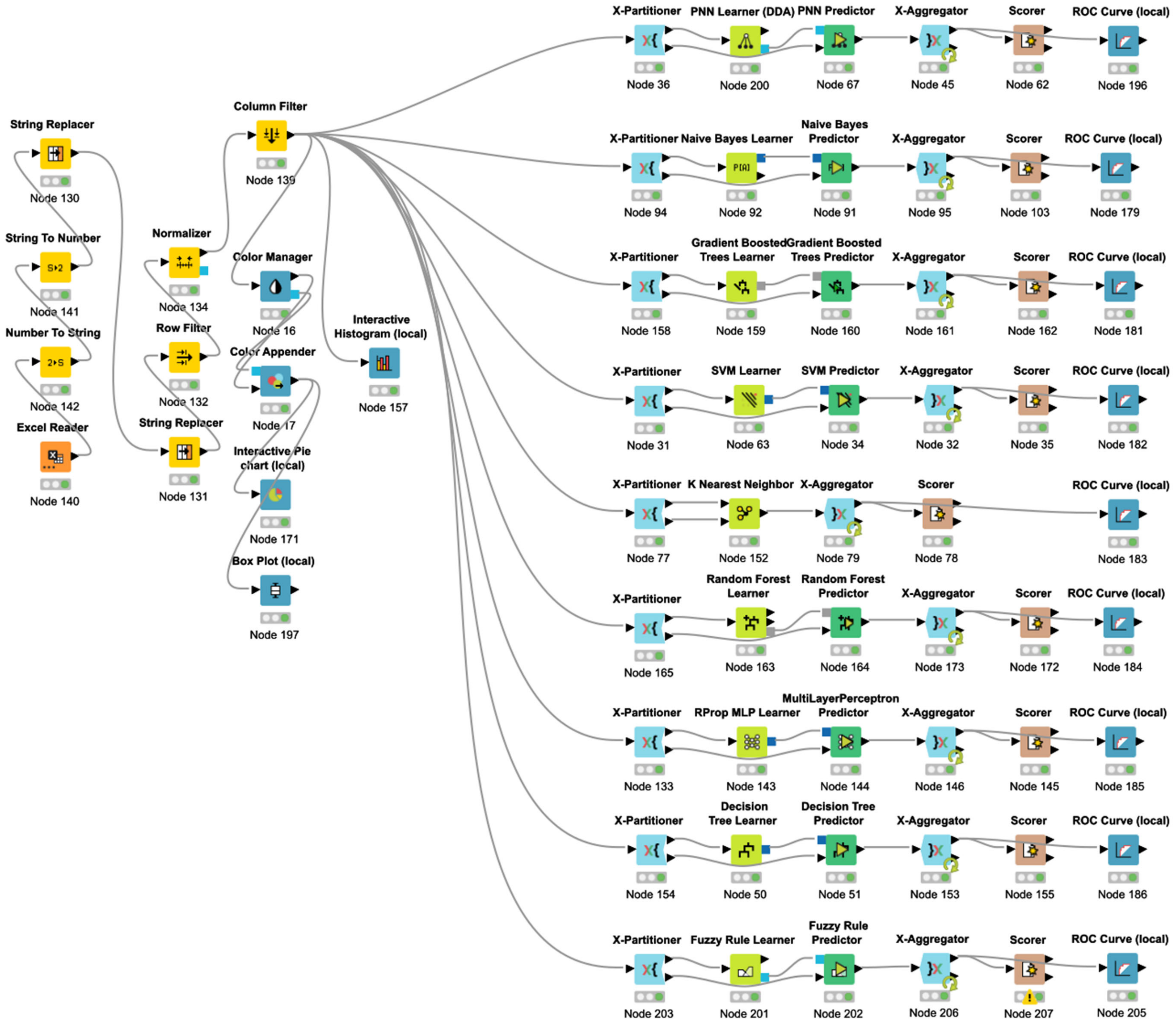
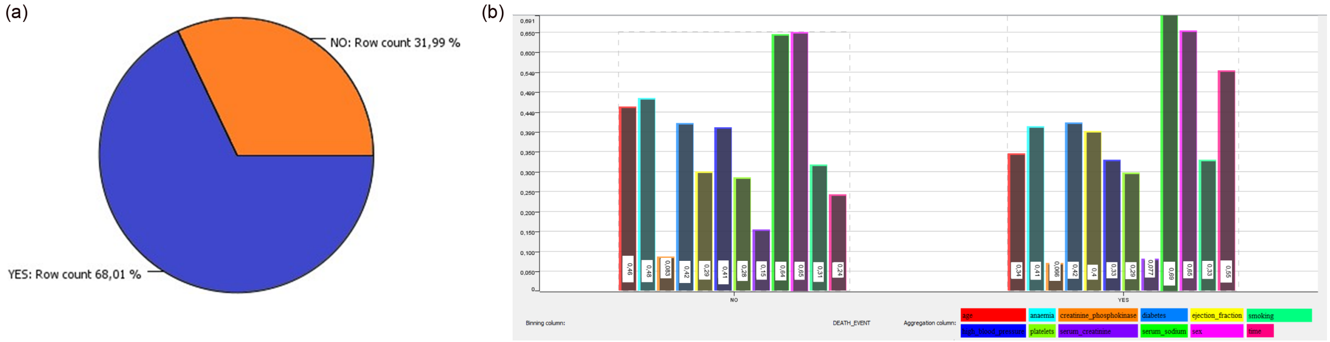


Figure 4
Receiver operating characteristic curve for both classes: $\{0, 1\}$



displays the workflows for DT, SVM, KNN, PNN, MLP, RF, NB, GB, and FR [9, 17].

Additionally, the preprocessing and cross-validation steps are shown. The data visualization step is also presented in the flowchart developed in the KNIME Analytics Platform workflow. Nodes are dragged, connected, and configured in the editor, with arrows illustrating the processing flow.

6. Results

Here, the results obtained from the data analysis will be presented together with the employability of the ROC curve as a whole, in order to identify the attributes that may be more relevant to classify people with heart failure, along with some data visualizations.

6.1. Data visualizations

This section presents data views to gain valuable insights into the dataset, including patterns, trends, and relationships. The database used is for classification, where data are categorized based on its characteristics. The dataset was obtained from the UCI Machine Learning Repository platform, with two classes represented as survivors (YES) and non-survivors (NO). Figure 4(a) shows that 31.99% of patients did not survive, while 68.01% survived heart failure.

The histogram in Figure 4(b) illustrates data attribute distribution for the survivor (YES) and non-survivor (NO) classes. Significant attributes for survivors include serum sodium (average 0.69), sex (average 0.65), and time (average 0.55). For non-survivors, important attributes are sex (average 0.65), serum sodium (average 0.64), and anemia (average 0.48). Notably, sex and serum sodium are important for both classes. The age attribute may be significant for class (NO) with an average of 0.46, while CPK showed relatively low averages for both classes, with values of 0.066 for survivors and 0.083 for non-survivors. Some attributes may be significant for both survival and death rankings, while others may not contribute significantly to the rankings.

6.2. ROC curve analysis

The ROC curve is a widely used tool for evaluating and comparing the performance of classification models in ML, including applications in the health sector [26]. The AUC is a single value that summarizes the ROC curve by calculating the area under it [27]. The AUC ranges from 0.0 to 1.0, with a threshold value of 0.5. Models with an AUC close to the maximum value are considered well-classified, while those below the threshold may not be suitable for the analysis [28]. In Figure 5, the graphical representation of the ROC curve can be seen together with the respective AUC values for each attribute, both for the class of survivors represented in Figure 6(a) and for the class of non-survivors in Figure 6(b). In this way, we can observe which attributes were most relevant, both for the classification of survivors and non-survivors due to heart failure based on the data sample studied.

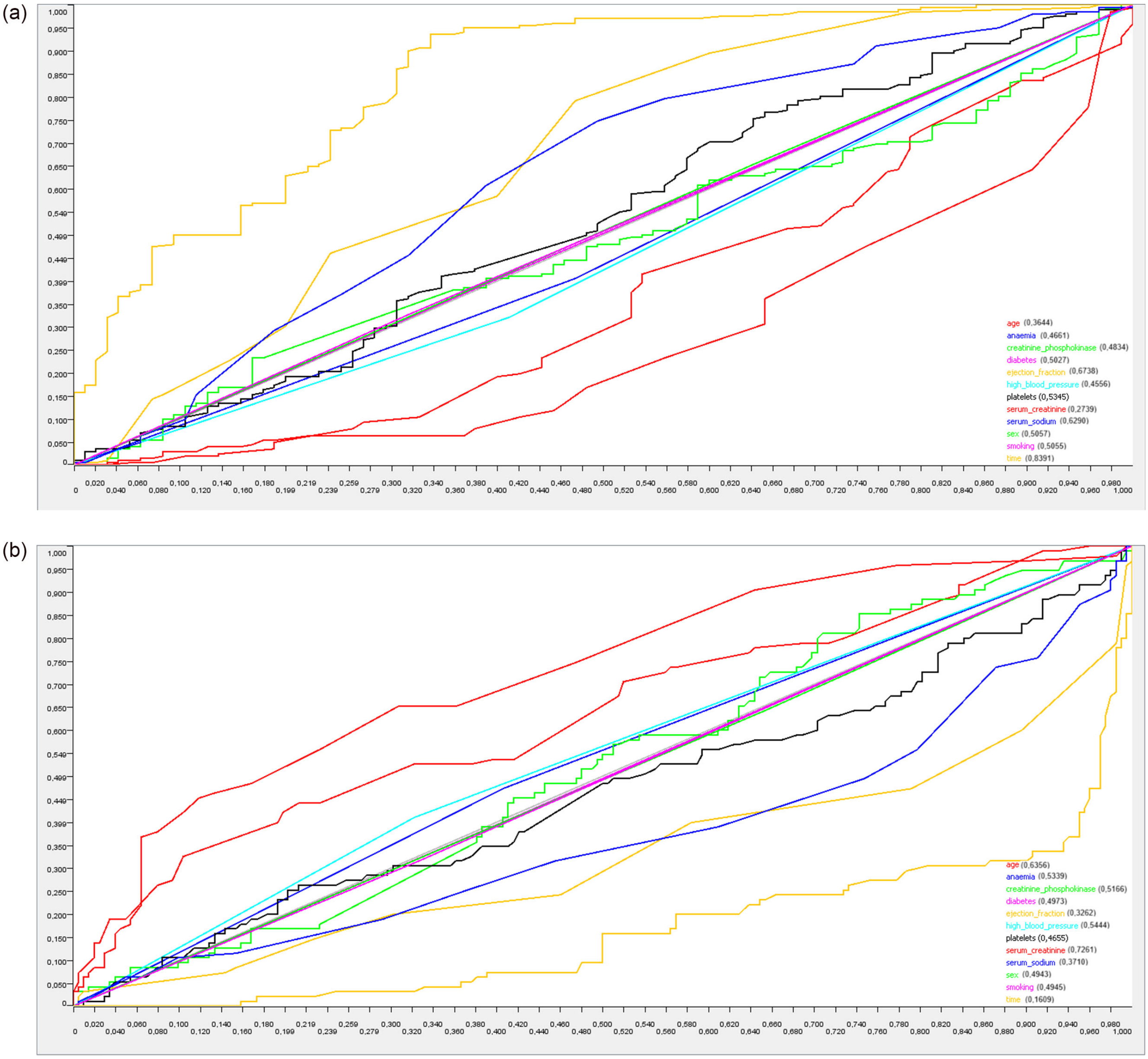
Table 2 represents in detail all the attributes studied together with their respective AUC values for the classification of survival and death of people with heart failure. Among the 13 attributes analyzed by the ROC curve, those that resulted in a value above 0.5 can be considered ideal attributes to be able to carry out the fact. Among all the attributes, the ones that obtained a better value for the AUC in the classification were the time with AUC of 0.833957 together with the attributes of ejection fraction with AUC of 0.673841 and serum sodium with AUC of 0.629026. In addition, the models that did not obtain a good result for classifying the survival of patients were serum creatinine with an AUC of 0.273893.

Table 2
Areas under curve for two classes YES and NO

Attributes	Class YES	Class NO
Age	0.364382	0.635618
Anemia	0.466128	0.533872
Creatinine phosphokinase	0.483377	0.516623
Diabetes	0.502658	0.497342
Ejection fraction	0.673841	0.326159
High blood pressure	0.455628	0.544372
Platelets	0.534549	0.465451
Serum creatinine	0.273893	0.726107
Serum sodium	0.629026	0.370974
Sex	0.505680	0.494320
Smoking	0.505472	0.494528
Time	0.839057	0.160943
Mean	0.51947412	0.48052588

So that we could analyze the class of people who did not survive, three other attributes stand out when we analyze the corresponding AUC values, namely serum creatinine with an AUC of 0.726107, age with an AUC of 0.635618, and high blood pressure with an AUC of 0.544372, respectively. The attribute that

Figure 5
Graphics for visualization of the representation for the classification class



did not obtain a better AUC result for the non-survivor class was the ejection fraction with a value of 0.326159. Therefore, values below 0.5 may not serve as reliable classifiers for predicting the survival possibility of patients with heart failure. However, they can still be valuable models for analyzing the classification of non-surviving patients in the dataset.

6.2.1. Column filtering based on AUC scores

In this work, we utilized column filters to identify relevant attributes for data analysis. The column filter allows us to select specific columns based on name, type, or pattern. It is valuable when aiming to reduce the number of attributes used for analyzing the possibility of patient survival in heart failure data. Table 3 presents four values for column filtering (CF_{ROC}), namely $CF_{ROC} = [0.3, 0.4, 0.5, 0.6]$, along with the no filter analysis, which includes all attributes of the database. The chosen filters were based on the corresponding AUC results. The table displays the number of included and excluded columns with each filter, along with the attributes used for each filter.

6.2.2. Analysis with the proposed algorithms

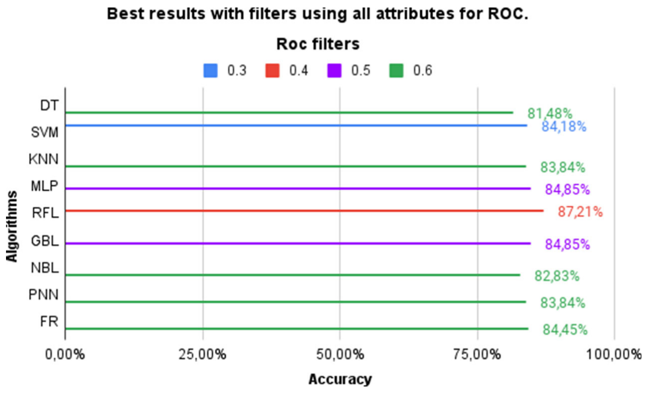
We employed nine selected DM algorithms to work with this database and achieve the best possible results in classifying the chances of survival or death for patients with heart failure. The algorithms used were DT, SVM, KNN, MLP, RFL, GBL, NBL, PNN, and FR, thus obtaining a series of results that were significant for our research. In this initial analysis, we utilized all 12 attributes of the dataset, along with the classification class, applying the proposed filters to identify the most relevant attributes for data analysis.

In Table 4, the results obtained with the first analysis using all the attributes are shown together with the application of the proposed filters. Figure 6 graphically represents the table of the best results

Table 3
Attribute inclusion/exclusion with filters on the ROC curve

Filters	Included	Excluded	Attributes that remained Column filter considering attribute time
No filter	13	0	Age, anemia, creatinine phosphokinase, diabetes, ejection fraction, high blood pressure, platelets, serum creatinine, serum sodium, sex, smoking, time, DEATH EVENT
0.3	12	1	Age, anemia, creatine phosphokinase, diabetes, ejection fraction, high blood pressure, platelets, serum sodium, sex, smoking, time, DEATH EVENT
0.4	11	2	Anemia, creatine phosphokinase, diabetes, ejection fraction, high blood pressure, platelets, serum sodium, sex, smoking, time, DEATH EVENT
0.5	8	5	Diabetes, ejection fraction, platelets, serum sodium, sex, smoking, time, DEATH EVENT
0.6	4	9	Ejection fraction, serum sodium, time, DEATH EVENT
			Column filter without considering attribute time
No filter	12	0	Age, anemia, creatinine phosphokinase, diabetes, ejection fraction, high blood pressure, platelets, serum creatinine, serum sodium, sex, smoking, DEATH EVENT
0.3	11	1	Age, anemia, creatine phosphokinase, diabetes, ejection fraction, high blood pressure, platelets, serum sodium, sex, smoking, DEATH EVENT
0.4	10	2	Anemia, creatine phosphokinase, diabetes, ejection fraction, high blood pressure, platelets, serum sodium, sex, smoking, DEATH EVENT.
0.5	7	5	Diabetes, ejection fraction, platelets, serum sodium, sex, smoking, DEATH EVENT
0.6	3	9	Ejection fraction, serum sodium, DEATH EVENT

Figure 6
Graph representing the best results for the ROC curve with all attributes



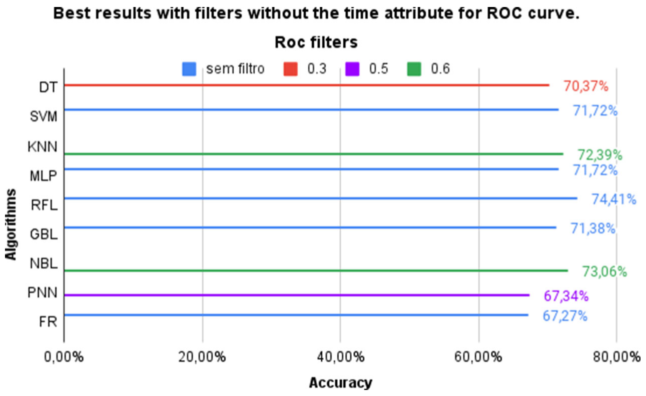
obtained with the filters mentioned above. Highlighted in green in the table is one of the best accuracies achieved, attributed to the RF algorithm with an accuracy of 87.21% and filter 0.4. Additionally, it holds the highest average accuracy among the nine algorithms, represented in magenta color, with a value of 84.51%. The RF algorithm utilizes 11 attributes for patient classification.

The algorithm that did not perform well in classifying the data was the PNN with an accuracy of 68.35% and $CF_{ROC} = 0.3$ which also had a worse average highlighted in red with 74.28%. Even though the RF obtained the best accuracy, it still uses many attributes to analyze the survival of patients with heart failure, but the FR obtained a good result considering a much smaller number of attributes when applying the $CF_{ROC} = 0.6$ resulting in 84.45% accuracy. This filter uses only four attributes to perform a better classification (ejection fraction, serum sodium, time, and the DEATH EVENT class).

But one cannot only evaluate the result obtained with accuracy to assess whether a DM algorithm had a better performance in fact to perform the diagnosis, we must consider other metrics that will be analyzed later.

Aiming as future work for this research to develop an application where ordinary people could see whether or not they could have a worsening of heart failure, we carried out an analysis without using the time attribute of the database, as this attribute refers to time follow-up of patients at the hospital [14], so these users would not have that time to report in the application. The results of this second analysis were represented in Table 5 and, likewise, Figure 7 graphically represents the best results obtained from the table for each algorithm discussed.

Figure 7
Graph representing the best results for the ROC curve without the time attribute



In green, the best accuracy found for the RFL algorithm corresponding to 74.41% is represented, but this time without applying a filter, which suggests that the algorithm made use of all the attributes of the database in the analysis. But this same algorithm stands out with another value, this time with $CF_{ROC} = 0.3$ and an accuracy of 73.06%. The worst performance was with the FR algorithm highlighted in red with 50.34% with a filter of $CF_{ROC} = 0.6$, that is, despite this algorithm using a much smaller number of

Table 4
Result with all attributes based on nine algorithms

Filters	DT	SVM	KNN	MLP	RFL	GBL	NBL	PNN	FR
No filter	81.15%	83.84%	71.04%	80.47%	83.84%	83.17%	76.43%	68.69%	75.54%
0.3	75.42%	84.18%	70.71%	82.49%	83.17%	83.17%	79.80%	68.35%	75.80%
0.4	80.14%	82.49%	71.04%	84.18%	87.21%	83.84%	78.45%	72.39%	73.85%
0.5	78.79%	83.17%	79.80%	84.85%	84.18%	84.85%	81.48%	78.11%	80.50%
0.6	81.48%	83.50%	83.84%	84.18%	84.18%	82.83%	82.83%	83.84%	84.45%
Mean	79.39%	83.43%	75.29%	83.23%	84.51%	83.57%	79.80%	74.28%	78.03%
StdDev.	0.024557	0.006474	0.061329	0.0177205	0.015611	0.008037	0.025082	0.066346	0.043574

Table 5
Results without the time attribute considering nine algorithms

Filters	DT	SVM	KNN	MLP	RFL	GBL	NBL	PNN	FR
No filter	63.30%	71.72%	67.34%	71.72%	74.41%	71.38%	70.03%	67.00%	67.27%
0.3	70.37%	70.03%	65.66%	69.02%	73.06%	67.68%	69.02%	66.33%	61.43%
0.4	67.34%	68.01%	65.32%	70.37%	67.00%	66.67%	68.01%	67.34%	64.03%
0.5	65.32%	68.01%	70.37%	71.38%	68.69%	65.99%	71.38%	67.34%	64.91%
0.6	68.69%	68.01%	72.39%	70.71%	69.02%	67.34%	73.06%	66.67%	50.34%
Mean	67.00%	69.16%	68.22%	70.64%	70.44%	67.81%	70.30%	66.94%	61.60%
StdDev.	0.027763	0.016770	0.030730	0.010484	0.031442	0.020971	0.019834	0.004389	0.066316

attributes, it was unable to overcome the accuracy of the others mentioned above without using the set time attribute.

The best average acquired was with the MLP algorithm highlighted by the magenta color with 70.64% and the worst in dark red with 61.60% belonging to the FR. Based on a lower use of attributes for classification, the NBL proved to be quite effective with 73.06% using four attributes with $CF_{ROC} = 0.6$, allowing unnecessary cost reduction with exams. Further metrics will be analyzed for better performance.

6.2.3. Analysis of ROC results with other metrics

Combining the five best results found in the tables together with the proposed filters, we created Table 6 to analyze the other evaluation metrics mentioned above. For the analysis of the results of the application of the AUC together with the related filters, the first part of the table represents the use of all the attributes of the database. The algorithm that performed best was the RF highlighted in green with an accuracy of 0.872; it also obtained a better Cohen's kappa in blue with 0.694, having applied the filter of 0.4 where the analysis of 11 attributes is done.

The RF also provided a good result for the sensitivity metric responsible for identifying observations that are really positive, marked in darker green with a value of 0.941. Despite not being the best, the specificity that represents how much the model was able to predict as negative the occurrences among all those that are negative did not get a bad value with 0.726. The second part of the table shows the results of the analysis without the time attribute, where the five best values were also chosen to compare the other evaluation metrics.

The RF again provided one of the best accuracies in green with a value of 0.744; it also obtained a good Cohen's kappa in blue with 0.377 but without the application of a filter dating from the use of 12 attributes. But this algorithm carries another good result with a filter of 0.3 and accuracy of 0.731, using much less attributes to perform the classification. The DT was the worst performer for accuracy in red with 0.704. The NBL obtained a better result for sensitivity in dark green with 0.95 for the YES class, while for the NO class, it obtained a bad result, only 0.263 in orange. For specificity, the result is bad for the class of survivors YES highlighted in red with 0.263, but for the class of non-survivors NO in blue a result of 0.95 was obtained. NBL also provided one of the best results for the F-measure metric for class YES, in dark green with 0.828. The RF algorithm without the "time" attribute demonstrated favorable results in sensitivity and specificity for both classes. This suggests that, by applying column filters, it is possible to successfully classify the survival of post-heart failure patients using selected attributes. This approach reduces operational costs by highlighting critical data for a specific disease.

6.3. Discussion

The results of this study are analyzed based on a publicly available dataset on heart disease, specifically heart failure. DM algorithms were used to determine the risk factors associated with the disease, along with the employability of the receiver operating characteristic or ROC curve to estimate the dynamism between the models discussed here. After employing nine classification algorithms and evaluating the (AUC) resulting from the application of the curve (ROC) with all the attributes of the database, the RF algorithm was identified as the most adequate to classify the survival or death of patients with the disease. With an accuracy of 87.21%, using a column filter of $CF_{ROC} = 0.4$ on the data and considering 11 attributes important such as anemia, CPK, diabetes, ejection fraction, high blood pressure, platelets, serum sodium, gender, smoking, time, and DEATH EVENT.

With the parallel analysis, without using one of the attributes of the database, namely the team, the RF algorithm again managed to provide a better result, totaling 74.41% of accuracy, but this result occurred without the employability of a related filter, that is, it used all the attributes of the database. On the other hand, the NB made use of much less clinical data from the set with the employability of $CF_{ROC} = 0.6$ resulting in an accuracy of 73.06% judging as the most relevant data, ejection fraction, serum sodium, and EVENT OF DEATH.

This emphasizes the importance of using DM algorithms when applied together with a tool commonly used in several areas, one of

Table 6
Filters with the best results with filters applied from the ROC curve

Best accuracies with the time attribute													
Algorithm	Filter	Accuracy	Cohen's kappa	Class	TP	FP	TN	FN	Recall	Precision	Sensitivity	Specificity	***F*-measure**
FR				**NO**	66	22	173	22	0.75	0.75	0.75	0.887	0.75
	0.6	0.845	0.637	**YES**	173	22	66	22	0.887	0.887	0.887	0.75	0.887
MLP				**NO**	64	16	186	31	0.674	0.8	0.674	0.921	0.731
	0.6	0.842	0.62	**YES**	186	31	64	16	0.921	0.857	0.921	0.674	0.888
GBL				**NO**	68	18	184	27	0.716	0.791	0.716	0.911	0.751
	0.5	0.848	0.643	**YES**	184	27	68	18	0.911	0.872	0.911	0.716	0.891
RFL				**NO**	69	12	190	26	0.726	0.852	0.726	0.941	0.784
	0.4	0.872	0.694	**YES**	190	26	69	12	0.941	0.88	0.941	0.726	0.909
SVM				**NO**	63	15	187	32	0.663	0.808	0.663	0.926	0.728
	0.3	0.842	0.618	**YES**	187	32	63	15	0.926	0.854	0.926	0.663	0.888
Best accuracies without the time attribute													
RFL				**NO**	47	28	174	48	0.495	0.627	0.495	0.861	0.553
	No filter	0.744	0.377	**YES**	174	48	47	28	0.861	0.784	0.861	0.495	0.821
NBL				**NO**	25	10	192	70	0.263	0.714	0.263	0.95	0.385
	0.6	0.731	0.257	**YES**	192	70	25	10	0.95	0.733	0.95	0.263	0.828
KNN				**NO**	41	28	174	54	0.432	0.594	0.432	0.861	0.5
	0.6	0.724	0.316	**YES**	174	54	41	28	0.861	0.763	0.861	0.432	0.809
DT				**NO**	51	44	158	44	0.537	0.537	0.537	0.782	0.537
	0.3	0.704	0.319	**YES**	158	44	51	44	0.782	0.782	0.782	0.537	0.782
RFL				**NO**	38	23	179	57	0.4	0.623	0.4	0.886	0.487
	0.3	0.731	0.316	**YES**	179	57	38	23	0.886	0.758	0.886	0.4	0.817

which is the health area, where we can evaluate the effectiveness of tests and possible diagnoses to predict the risk of diseases such as heart failure. The ROC curve played a crucial role in evaluating the importance of each clinical data attribute, offering valuable insights into the disease's prognostic factors. In related works, different studies also utilized DM techniques and the same dataset to identify relevant attributes for predicting heart failure survival. Ahmad et al. [14] explored various statistical methods and identified significant attributes such as age, ejection fraction, sodium, anemia, high blood pressure, and creatine, and they achieved in best result 81%. Chicco et al. [10] focused on only two clinical factors, serum creatine, and ejection fraction, while Hasan et al. [15] highlighted ejection fraction and serum creatine as important predictors for survival, achieving 80%. Ishaq et al. [16] employed several categorization models and achieved high accuracy using all attributes of the database, achieving 92.62%, but in this case, all attributes were used for classification.

The study by Gürfidan and Ersoy [17] highlights SVM with 83% accuracy using all attributes, while our work surpassed this result with RF, achieving 87.21% accuracy. This was possible by employing only 11 attributes and a 0.4 filter, evidencing the efficiency of the RF in the classification of patients with heart failure. Muntasir Nishat et al. [18] achieve even higher accuracy than Gürfidan and Ersoy [17], highlighting the effectiveness of the RFC in conjunction with advanced preprocessing techniques. Evidencing a continuous trend of improvement in ML applications in healthcare. However, in our study, we utilized the ROC curve alongside filters to reduce the number of attributes used in prediction and we achieved 87.21% of accuracy, considering attribute time and 74.41% without considering this attribute. This allowed us to identify the most relevant attributes, making our model more interpretable and efficient in classifying heart failure patients. The attributes more relevant were ejection fraction, high blood pressure, platelets, serum sodium, sex, smoking, and class DEATH EVENT.

Although the "time" attribute results in an average accuracy of about ≈ 74% for the RFL algorithm, its consideration requires caution due to different hospital protocols. Previous studies that included it achieved higher accuracies; however, due to its global variation, it is advisable to avoid using it exclusively for forecasting. By relying solely on collected patient data, we achieved 74.41% accuracy for RFL and an average of 70.64% for MLP, excluding the "time" attribute. The exclusion of this attribute is vital to guarantee the generalization of the model. Finally, when considering the "time" attribute, similar to previous studies, our approach achieved an accuracy of 87.21% while using only four attributes. This computational efficiency makes it more attractive for public health systems worldwide, as it is more cost-effective to classify cardiac issues in patients using fewer attributes.

7. Conclusion

The heart, a vital organ, is prone to diseases like hypertension, acute myocardial infarction, and heart failure, stemming from poor habits or heredity [16]. Our research collaborates with data science experts to predict heart disease patients' survival chances. This study identifies significant risk factors and demonstrates the efficacy of DM algorithms, especially when combined with the ROC curve and attribute filtering. The results aid healthcare professionals in risk stratification, enhancing heart failure management and patient care [16]. Analyzing an open and public database, we identified key attributes relevant to classifying survival in heart failure patients. Employing DM and AI techniques, like the ROC curve and AUC, we discovered significant models for predicting patient outcomes.

After performing ML tests and applying AUC-based filters on the ROC curve, we collected detailed results in the tables. The RF algorithm stood out, achieving 87.21% accuracy with a filter of 0.4, and had the best average of 84.51% among the tested algorithms using 11 attributes for patient classification. In the second analysis, without the time attribute, the RF algorithm still performed well with 74.41% accuracy, using all attributes. However, for attribute reduction, the NBL algorithm showed good results with 73.06% accuracy, using only three attributes: ejection fraction, serum sodium, and DEATH EVENT with a filter of 0.6.

Future work includes developing an application for common users to assess their predisposition to heart failure complications using clinical, social, and behavioral data. Exploring other supervised algorithms for diagnosis will also be part of the research. The application will contribute to an updated database for a better understanding of aggravating factors and improved treatment. Additional studies are needed to evaluate and compare algorithms in various clinical contexts and diverse populations. Employing the ROC curve enables in-depth analysis of attributes and models important for both survivors and non-survivors. This reduces data collection to focus only on the most relevant factors for specific diseases.

Funding Support

This research was funded by the Foundation for Supporting Research in the state of Minas Gerais (Fapemig).

Ethical Statement

This study does not contain any studies with human or animal subjects performed by any of the authors.

Conflicts of Interest

The authors declare that they have no conflicts of interest to this work.

Data Availability Statement

The data that support the findings of this study are openly available in [UC Irvine Machine Learning Repository] at https://doi.org/10.24432/C5Z89R.

References

[1] Standring, S. (2021). *Gray's anatomy e-book: The anatomical basis of clinical practice*. Netherlands: Elsevier Health Sciences.

[2] Day, E., & Rudd, J. H. F. (2019). Alcohol use disorders and the heart. *Addiction*, *114*(9), 1670–1678. https://doi.org/10.1111/add.14703

[3] Kim, H., Caulfield, L. E., Garcia-Larsen, V., Steffen, L. M., Coresh, J., & Rebholz, C. M. (2019). Plant-based diets are associated with a lower risk of incident cardiovascular disease, cardiovascular disease mortality, and all-cause mortality in a general population of middle-aged adults. *Journal of the American Heart Association*, *8*(16), e012865. https://doi.org/10.1161/JAHA.119.012865

[4] Ohkuma, T., Komorita, Y., Peters, S. A. E., & Woodward, M. (2019). Diabetes as a risk factor for heart failure in women and men: A systematic review and meta-analysis of 47 cohorts including 12 million individuals. *Diabetologia*, *62*, 1550–1560. https://doi.org/10.1007/s00125-019-4926-x

[5] Brouwers, S., Sudano, I., Kokubo, Y., & Sulaica, E. M. (2021). Arterial hypertension. *The Lancet*, *398*(10296), 249–261. https://doi.org/10.1016/S0140-6736(21)00221-X

[6] Santos, I. D. S., & Bittencourt, M. S. (2008). Insuficiência cardaca Failure cardiac. *Revista de Medicina*, 224–231.

[7] Duarte, D., Puerari, I., Dal Bianco, G., & Lima, J. F. (2020). Exploratory analysis of electronic health records using topic modeling. *Journal of Information and Data Management*, *11*(2), 131–147.

[8] Dornelas, R. S., & Lima, D. A. (2023). Correlation filters in machine learning algorithms to select demographic and individual features for autism spectrum disorder diagnosis. *Journal of Data Science and Intelligent Systems*, *1*(2), 105–107. https://doi.org/10.47852/bonviewJDSIS32021027

[9] Chicco, D., & Jurman, G. (2020). Machine learning can predict survival of patients with heart failure from serum creatinine and ejection fraction alone. *BMC Medical Informatics and Decision Making*, *20*(1), 16. https://doi.org/10.1186/s12911-020-1023-5

[10] Chicco, D., Warrens, M. J., & Jurman, G. (2021). The Matthews correlation coefficient (MCC) is more informative than Cohen's Kappa and Brier score in binary classification assessment. *IEEE Access*, *9*, 78368–78381. https://doi.org/10.1109/ACCESS.2021.3084050

[11] Osisanwo, F. Y., Akinsola, J. E. T., Awodele, O., Hinmikaiye, J. O., Olakanmi, O., & Akinjobi, J. (2017). Supervised machine learning algorithms: Classification and comparison. *International Journal of Computer Trends and Technology*, *48*(3), 128–138.

[12] Lima, D. A., Ferreira, M. E. A., & Silva, A. F. F. (2021). Machine learning and data visualization to evaluate a robotics and programming project targeted for women. *Journal of Intelligent & Robotic Systems*, *103*(1), 4. https://doi.org/10.1007/s10846-021-01443-w

[13] Lima Filho, S. P., da Silva, M. F., Sampaio, J. O., & Rodrigues, L. C. R. (2022). A study about gathering features in depression detection'problem with health professionals community. *iSys-Brazilian Journal of Information Systems*, *15*(1), 1–26.

[14] Ahmad, T., Munir, A., Bhatti, S. H., Aftab, M., & Raza, M. A. (2017). Survival analysis of heart failure patients: A case study. *PLOS ONE*, *12*(7), e0181001. https://doi.org/10.1371/journal.pone.0181001

[15] Hasan, M. A. M., Shin, J., Das, U., & Srizon, A. Y. (2021). Identifying prognostic features for predicting heart failure by using machine learning algorithm. In *Proceedings of the 11th International Conference on Biomedical Engineering and Technology*, 40–46. https://doi.org/10.1145/3460238.3460245

[16] Ishaq, A., Sadiq, S., Umer, M., Ullah, S., Mirjalili, S., Rupapara, V., & Nappi, M. (2021). Improving the prediction of heart failure patients' survival using smote and effective data mining techniques. *IEEE Access*, *9*, 39707–39716. https://doi.org/10.1109/ACCESS.2021.3064084

[17] Gürfidan, R., & Ersoy, M. (2021). Classification of death related to heart failure by machine learning algorithms. *Advances in Artificial Intelligence Research*, *1*(1), 13–18.

[18] Muntasir Nishat, M., Faisal, F., Jahan Ratul, I., Al-Monsur, A., Ar-Rafi, A. M., Nasrullah, S. M., . . . , & Khan, M. R. H. (2022). A comprehensive investigation of the performances of different machine learning classifiers with SMOTE-ENN oversampling technique and hyperparameter optimization for imbalanced heart failure dataset. *Scientific Programming*, *2022*(1), 3649406. https://doi.org/10.1155/2022/3649406

[19] Lotufo, P. A. (1998). Premature mortality from heart diseases in Brazil. An international comparison. *Arquivos Brasileiros de Cardiologia*, *70*(5), 321–325. https://doi.org/10.1590/S0066-782X1998000500003

[20] Buckberg, G. D., Nanda, N. C., Nguyen, C., & Kocica, M. J. (2018). What is the heart? Anatomy, function, pathophysiology, and misconceptions. *Journal of Cardiovascular Development and Disease*, *5*(2), 33. https://doi.org/10.3390/jcdd5020033

[21] Cardiology Associates of Michigan. (2019). *Basic anatomy of the human heart*. Retrieved from: https://www.cardofmich.com/anatomy-human-heart-fun-facts/

[22] Berto, K. A., & Berto, B. A. (2015). *Aspectos Anatômicos do Tórax: Anatomia Torácica* [Anatomical aspects of the Thorax: Thoracic Anatomy]. Spanish: Alción Editora.

[23] Alves, F. B., & Lima, D. A. (2018). *Uso de la clasificaci ón para el análisis y la minera de datos en la herramienta de enseñanza aprendizaje google classroom* [Use of classification to data analysis and mining in the Google classroom teaching-learning tool]. *Nuevas Ideas en Informática Educativa*, *14*, 589–594.

[24] Lu, H., Uddin, S., Hajati, F., Moni, M. A., & Khushi, M. (2022). A patient network-based machine learning model for disease prediction: The case of type 2 diabetes mellitus. *Applied Intelligence*, *52*(3), 2411–2422. https://doi.org/10.1007/s10489-021-02533-w

[25] Braga, A. C. S. (2001). *Curvas ROC: Aspectos funcionais e aplicações* [ROC Curves: Functional aspects and applications]. PhD Thesis, University of Minho.

[26] Kumar, R., & Indrayan, A. (2011). Receiver operating characteristic (ROC) curve for medical researchers. *Indian Pediatrics*, *48*, 277–287. https://doi.org/10.1007/s13312-011-0055-4

[27] Walter, S. D. (2005). The partial area under the summary ROC curve. *Statistics in Medicine*, *24*(13), 2025–2040. https://doi.org/10.1002/sim.2103

[28] Marzban, C. (2004). The ROC curve and the area under it as performance measures. *Weather and Forecasting*, *19*(6), 1106–1114.

[29] Japkowicz, N. (2006). Why question machine learning evaluation methods (An illustrative review of the shortcomings of current methods)? In *AAAI Workshop on Evaluation Methods for Machine Learning, 6*(11), 1–6.

[30] Sidey-Gibbons, J. A. M., & Sidey-Gibbons, C. J. (2019). Machine learning in medicine: A practical introduction. *BMC Medical Research Methodology*, *19*(1), 64. https://doi.org/10.1186/s12874-019-0681-4

[31] Alvarez, S. A. (2002). *An exact analytical relation among recall, precision, and classification accuracy in information retrieval.* Technical Report BCCS-02-01, Boston College, 1–22.

[32] Hand, D. J., Christen, P., & Kirielle, N. (2021). F*: An interpretable transformation of the F-measure. *Machine Learning*, *110*(3), 451–456. https://doi.org/10.1007/s10994-021-05964-1

How to Cite: Souza, V. S., & Lima, D. A. (2024). Identifying Risk Factors for Heart Failure: A Case Study Employing Data Mining Algorithms. *Journal of Data Science and Intelligent Systems*, *2*(3), 161–173. https://doi.org/10.47852/bonviewJDSIS32021386