Research Article – Vol. 02, Issue 04, No. 10

Advanced in Engineering and Intelligence Systems

Journal Web Page: https://aeis.bilijipub.com

The use of neural networks in the diagnosis of heart failure via the analysis of medical data

Valentino Blanco[1, *], Aitana Iglesias[2]

[1] *Univ Buenos Aires, Fac Agron, 4453 San Martin Ave, Buenos Aires, DF, Argentina*
[2] *Fac Reg San Nicolas FRSN UTN San Nicolas, Dept Ingn Elect, Buenos Aires, DF, Argentina*

Highlights

- Emphasis on the critical role of effective decision-making in management, relying on information availability and communication.
- Utilization of hospital managers' outcomes from data mining to develop an intelligent model using machine learning techniques.
- Objective to enhance prediction accuracy and facilitate more effective decision-making in patient treatment.
- Analysis of a dataset comprising demographic and clinical information of 297 individuals, applying "k-means, support vector machine, and neural network" models for classification.
- Demonstration of the neural network model achieving a classification accuracy of 89.9%, while the support vector machine with radial basis function kernel achieves a higher accuracy of 93%.

Article Info

Received: 29 November 2023
Received in revised: 25 December 2023
Accepted: 30 December 2023
Available online: 31 December 2023

Keywords

data mining;
k-means;
neural network;
SVM;
heart disease

Abstract

The crux of effective management is in the process of decision-making, which is contingent upon the availability of information and effective communication. The fundamental responsibility of executives is to provide the necessary information to facilitate sound management decisions. This study seeks to utilize hospital managers' outcomes from data mining of hospital information systems to develop an intelligent model using machine learning techniques. The objective is to enhance the accuracy of predictions and facilitate more effective decision-making in patient treatment, recognizing the significance of hospital managers' decision-making approaches in advancing hospital goals and addressing patients' treatment challenges. The dataset used in this research pertains to the demographic and clinical information of 297 individuals. This data was obtained from the UCI website's data warehouse and encompasses 14 distinct variables. The three models, namely "k-means, support vector machine, and neural network," are extensively used classification methods in the domains of data mining and machine learning. These models have been applied to forecast cardiac disease, and their predictive performance has been evaluated and compared. The findings demonstrate that the neural network model, characterized by a multi-layered perceptron architecture, achieved a classification accuracy of 89.9% when applied to the test dataset. However, the support vector machine using the radial basis function kernel demonstrates enhanced accuracy, achieving a level of 93%.

Nomenclature

Indices			
CVDs	*Cardiovascular diseases*	*PCA*	*Principal component analysis*
CHD	*Coronary heart diseases*	*LR*	*Logistic regression*
IHD	*İschemic heart diseases*	*TN*	*Treatment not nrcessary*
CHF	*Chronic heart failure*	*FP*	*False positive result*
GPLv3	*General public licevse version 3*	*FN*	*False negative diagnosis*
DSS	*Decision support system*	*RBT*	*Radial basis function*
DNN	*Deep neural network*		

1. Introduction

Acquiring a precise diagnosis for a patient's medical condition is a prominent concern within medical science, a highly consequential discipline. The process of medical diagnosis is often seen as a substantial and demanding endeavor that must be conducted with prudence and

thoroughness [1], [2]. Machine learning modeling and data mining technologies are gaining popularity in the healthcare industry due to their potential to improve patient care through early disease detection, reduce the burden on care professionals associated with treatment plans, and decrease overall healthcare expenses. In the last several years, there has been a proliferation of diverse machine learning methodologies and algorithms used for the purpose of predicting medical diagnoses [3]– [5]. Clustering and classification are two essential techniques used in the data mining process. Classification is a machine learning technique that falls under the category of supervised learning since it requires labeled training data to make predictions. On the other hand, clustering is an unsupervised learning strategy that does not rely on labeled data to group similar instances together.[6]

Cardiovascular diseases (CVDs), often known as cardiovascular ailments, have surpassed all other contributing factors to mortality and have emerged as the primary cause of death globally [7], [8]. Based on data from the World Health Organization, it has been determined that cardiovascular disease is responsible for the deaths of around 12 million individuals worldwide [9]. Based on a comprehensive study conducted by the Ministry of Health and Medicine in China, it has been determined that cardiovascular diseases account for 39.9% of the total mortality rate in the country. Given that they are the primary cause of mortality in China, this statistical data indicates that one-third of all fatalities may be attributed to this factor [10]

Coronary heart disease (CHD), ischemic heart disease (IHD), stroke, chronic heart failure (CHF), vascular and brain disorders, and other conditions are included under the classification of heart disease. Therefore, it is essential to identify more effective methods for diagnosing and evaluating the prognosis of these patients [11]

Given the notable rise in the occurrence of these ailments, along with the subsequent ramifications and difficulties they present, as well as the economic strain they impose on society, the healthcare sector has been actively pursuing endeavors aimed at promoting additional investigation, prevention, timely identification, and patient diagnosis. This phenomenon may be attributed to the notable rise in various ailments. Furthermore, the medical community has aggressively pursued programs to facilitate further study. This phenomenon may be attributed to the consistent upward trend in the prevalence of various diseases in recent decades. Furthermore, there has been a collective effort within the scientific and medical sectors to identify novel initiatives to facilitate research. This answer addresses the impact of the financial burden associated with these illnesses on the overall economic system of society. [12]

The economic burden of the expenses related to managing chronic illnesses significantly impacts the healthcare industry, resulting in adverse consequences. The user's text does not provide any information or context. Therefore, it cannot be rewritten. Within the realm of contemporary medical research, a substantial multitude of medical institutions are now engaged in the active acquisition of such data to fulfill various distinct objectives. This phenomenon might be attributed to the prevailing characteristic of contemporary medical research, which involves the systematic accumulation of substantial volumes of data about many areas. The interested person can investigate several methods to get this information. The user's text consists of a numerical range, specifically [13], [14]

In recent years, there has been an observable inclination in several domains of medicine towards the use of data mining methodologies for the purpose of analyzing extensive datasets with the aim of constructing predictive models and identifying patterns [15]. Currently, many studies are emphasizing the use of data mining and machine learning methodologies as valuable tools for identifying important health patterns embedded within medical records. The following paragraphs will include an analysis of two instances of prior research conducted within this domain [16]–[18]

The establishment of a robust methodology for predicting the incidence of cardiovascular disease.

This study presents an automated framework for investigating treatment methodologies. The current framework has the potential to decrease expenses, and it further incorporates techniques for assessing a patient's risk level for a particular health indicator [19]. This research makes a valuable contribution to the area by offering important recommendations to general practitioners. These recommendations will assist healthcare professionals in effectively assessing the likelihood of cardiovascular disease in their patients. The researchers conducted the implementation of this study with the use of knowledge extraction technology rooted in evolutionary learning, known as KEEL. The KEEL programming engine is open-source software written in Java and is distributed under the GNU General Public License version 3 (GPLv3) [20], [21].

Furthermore, in the data preparation phase, the "all possible" approach was used to address the challenge of imputing the missing values. This procedure was conducted as a component of "imputing the missing values." The

implementation of this method was undertaken in order to address the situation at hand.

Performance Comparison

The results obtained from the application of several established algorithms to the dataset, including information about cardiovascular diseases, are summarized in Table 1, which presents the findings. Based on the research results, the predictive accuracy of the technique for cardiovascular disease surpasses that of all previously disclosed classification systems.[22]

Table 1. The performance of different known algorithms in cardiac disease datasets.

The Algorithm used	**SVM**	**C4.5**	**1-NN**	**PART**	**MLP**	**RBF**	**TSEAFS**	**Efficient heart disease prediction system**
Accuracy(%)	70.59	73.53	76.47	73.53	74.85	78.53	77.45	86.7

2- Analyzing the characteristics of the coronary heart data set

After applying the C4.5 tree and the Fast Decision Tree algorithms to the Cleveland heart data set, the resulting outcomes are then confirmed and compared. In this particular scenario, the selection of the most desired features is based on the available options, which are evaluated using two trees, as shown in Table 2.

Table 2. uses datasets.

Dataset	**Decision Tree**	**Best selected feature**
Cleveland	C4.5 FDT	Cp, Thal, Ca, Thal Cp, Age, Ca, Thal, Thalach
Hungarain	C4.5 FDT	Exang, Oldpeak, Sex, Cp, Slop, Thalach Cp,Oldpeak
V.A	C4.5 FDT	Cp, Age, Exang, Fbs, Sex Cp, Age, Chol
Statlog	C4.5 FDT	Thal, Cp, Ca, Exang, Age Ca, Thal, Cp, Thalach, Oldpeak

As seen by the data presented in Table 3, the accuracy of the two trees in relation to the best-selected features surpasses that of the original dataset. The data shown in this table illustrates the potential of extracting knowledge via data integration to enhance the accuracy of the diagnostic procedure. The methodology presented effectively addresses and resolves the inherent ambiguity found in various sources of information.

Table 3. Comparison results of the accuracy of the best features and the average of the collected data set

Decision Tree	**Average of each separate dataset accuracy %**	**Best selected featured collected dataset accuracy %**
C4.5 FDT	76.30 75.48	77.50 78.06

The authors in [23]present a decision support system (DSS) based on neural networks and statistical process control charts for diagnosis and control of myocardial infarction (MI) and continuous patient blood pressure monitoring. [24] proposed a new approach for heart disease prediction that uses a pre-trained deep neural network (DNN) for feature extraction, principal component analysis (PCA) for dimensionality reduction, and logistic regression (LR) for prediction.

In recent years, there has been an increasing use of data mining techniques in medical research, especially in the field of medical diagnosis. This approach involves extracting valuable insights from large and growing medical databases. By employing data mining methods, researchers can uncover patterns and relationships within vast datasets, leading to advancements in patient care. The growing availability of information within these databases is a key factor contributing to this trend. The primary objective of this research is to develop a methodology that effectively forecasts the occurrence of coronary disease while simultaneously reducing the incidence of false-positive results. This objective will be achieved by using a diverse range of data mining and machine learning methodologies, such as the K-means algorithm and the

support vector machine with neural network architecture, among many others. Several techniques will be incorporated.

The rest of this paper can be categorized as follows: In Section 2, the method is presented. The result is discussed in Section 3. In Section 4, the discussion is expressed, and in Section 5, the conclusion is stated.

2. Method

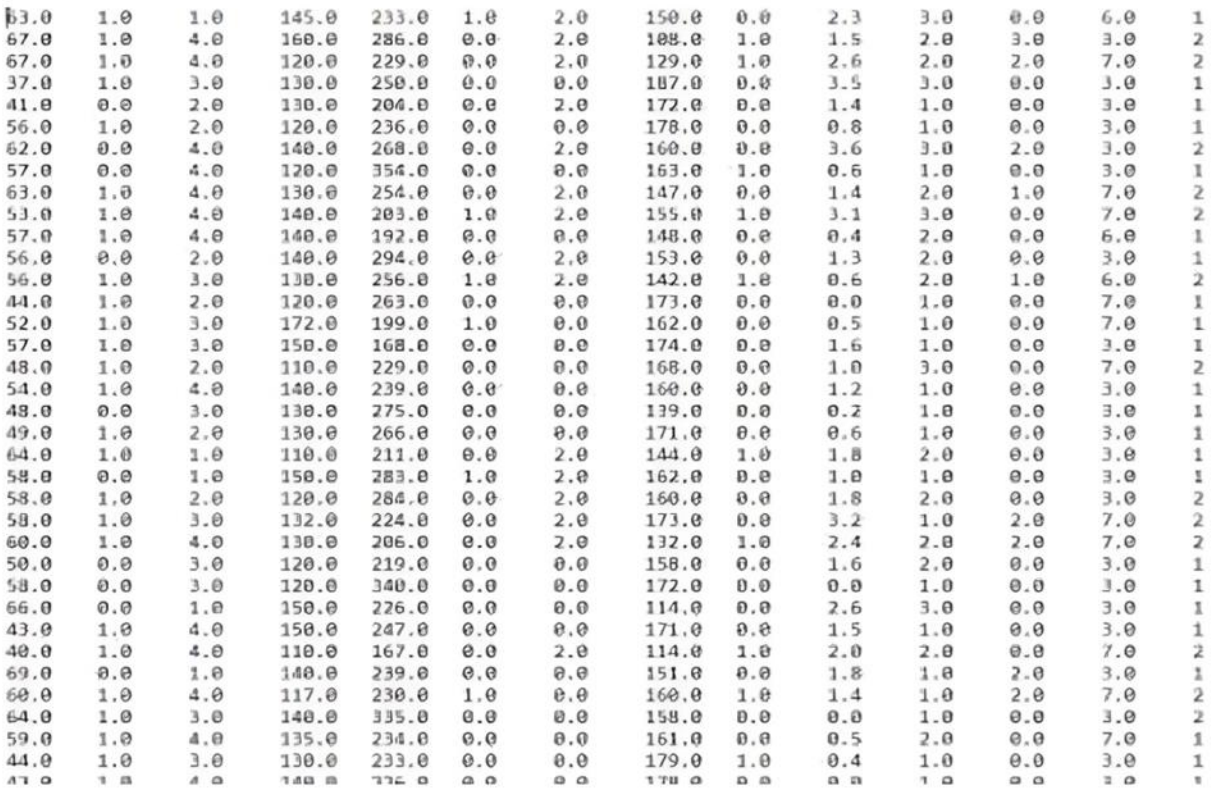
In this research, the data were collected and standardized according to the desired models. Subsequently, the data were used inside the Matlab programming environment, and the resulting results were evaluated at the completion of the inquiry.

The heart disease dataset used in this research effort was obtained from the Cleveland Foundation and sourced from the University of California Machine Learning Information Repository (UCI). The dataset has 297 distinct samples and encompasses 14 distinct variables. Variable 14 is designated as the predictive response variable, while the other variables are classified as predictive explanatory variables. The correlations are presented in a concise manner in Table 4, which is accessible for viewing at this location.

As seen in Fig. 1, each column inside the figure corresponds to an individual value pertaining to one of the variables outlined in Table 4. As previously said, the last column presents a comprehensive analysis of the numerical data pertaining to each class. Class, one signifies those in good health, whereas class two denotes those who are ill.

Table 4. Abbreviated title and type of used variables

variable type	**Variable name**
quantitative	Age
qualitative	Sex
qualitative	Chest pain
quantitative	Blood pressure
quantitative	Serum cholesterol
qualitative	Blood Sugar
qualitative	ECG static
quantitative	Heart rate
qualitative	Angina
quantitative	ST decrease
qualitative	ST Peak
quantitative	Number of large vessels
qualitative	Defect
qualitative	Heart Disease



63.0	1.0	1.0	145.0	233.0	1.0	2.0	150.0	0.0	2.3	3.0	0.0	6.0	1
67.0	1.0	4.0	160.0	286.0	0.0	2.0	108.0	1.0	1.5	2.0	3.0	3.0	2
67.0	1.0	4.0	120.0	229.0	0.0	2.0	129.0	1.0	2.6	2.0	2.0	7.0	2
37.0	1.0	3.0	130.0	250.0	0.0	0.0	187.0	0.0	3.5	3.0	0.0	3.0	1
41.0	0.0	2.0	130.0	204.0	0.0	2.0	172.0	0.0	1.4	1.0	0.0	3.0	1
56.0	1.0	2.0	120.0	236.0	0.0	0.0	178.0	0.0	0.8	1.0	0.0	3.0	1
62.0	0.0	4.0	140.0	268.0	0.0	2.0	160.0	0.0	3.6	3.0	2.0	3.0	2
57.0	0.0	4.0	120.0	354.0	0.0	0.0	163.0	1.0	0.6	1.0	0.0	3.0	1
63.0	1.0	4.0	130.0	254.0	0.0	2.0	147.0	0.0	1.4	2.0	1.0	7.0	2
53.0	1.0	4.0	140.0	203.0	1.0	2.0	155.0	1.0	3.1	3.0	0.0	7.0	2
57.0	1.0	4.0	140.0	192.0	0.0	0.0	148.0	0.0	0.4	2.0	0.0	6.0	1
56.0	0.0	2.0	140.0	294.0	0.0	2.0	153.0	0.0	1.3	2.0	0.0	3.0	1
56.0	1.0	3.0	130.0	256.0	1.0	2.0	142.0	1.0	0.6	2.0	1.0	6.0	2
44.0	1.0	2.0	120.0	263.0	0.0	0.0	173.0	0.0	0.0	1.0	0.0	7.0	1
52.0	1.0	3.0	172.0	199.0	1.0	0.0	162.0	0.0	0.5	1.0	0.0	7.0	1
57.0	1.0	3.0	150.0	168.0	0.0	0.0	174.0	0.0	1.6	1.0	0.0	3.0	1
48.0	1.0	2.0	110.0	229.0	0.0	0.0	168.0	0.0	1.0	3.0	0.0	7.0	2
54.0	1.0	4.0	140.0	239.0	0.0	0.0	160.0	0.0	1.2	1.0	0.0	3.0	1
48.0	0.0	3.0	130.0	275.0	0.0	0.0	139.0	0.0	0.2	1.0	0.0	3.0	1
49.0	1.0	2.0	130.0	266.0	0.0	0.0	171.0	0.0	0.6	1.0	0.0	3.0	1
64.0	1.0	1.0	110.0	211.0	0.0	2.0	144.0	1.0	1.8	2.0	0.0	3.0	1
58.0	0.0	1.0	150.0	283.0	1.0	2.0	162.0	0.0	1.0	1.0	0.0	3.0	1
58.0	1.0	2.0	120.0	284.0	0.0	2.0	160.0	0.0	1.8	2.0	0.0	3.0	2
58.0	1.0	3.0	132.0	224.0	0.0	2.0	173.0	0.0	3.2	1.0	2.0	7.0	2
60.0	1.0	4.0	130.0	206.0	0.0	2.0	132.0	1.0	2.4	2.0	2.0	7.0	2
50.0	0.0	3.0	120.0	219.0	0.0	0.0	158.0	0.0	1.6	2.0	0.0	3.0	1
58.0	0.0	3.0	120.0	340.0	0.0	0.0	172.0	0.0	0.0	1.0	0.0	3.0	1
66.0	0.0	1.0	150.0	226.0	0.0	0.0	114.0	0.0	2.6	3.0	0.0	3.0	1
43.0	1.0	4.0	150.0	247.0	0.0	0.0	171.0	0.0	1.5	1.0	0.0	3.0	1
40.0	1.0	4.0	110.0	167.0	0.0	2.0	114.0	1.0	2.0	2.0	0.0	7.0	2
69.0	0.0	1.0	140.0	239.0	0.0	0.0	151.0	0.0	1.8	1.0	2.0	3.0	1
60.0	1.0	4.0	117.0	230.0	1.0	0.0	160.0	1.0	1.4	1.0	2.0	7.0	2
64.0	1.0	3.0	140.0	335.0	0.0	0.0	158.0	0.0	0.0	1.0	0.0	3.0	2
59.0	1.0	4.0	135.0	234.0	0.0	0.0	161.0	0.0	0.5	2.0	0.0	7.0	1
44.0	1.0	3.0	130.0	233.0	0.0	0.0	179.0	1.0	0.4	1.0	0.0	3.0	1
47.0	1.0	4.0	140.0	226.0	0.0	0.0	178.0	0.0	0.0	1.0	0.0	3.0	1

Fig. 1. Initial dataset

3. Result

After normalization, the desired data were analyzed with three methods:

1- k-means algorithm

2- Multilayer neural network

3- Svm

1- Steps of the K-means algorithm

[1] First, we randomly select k cluster centers.

[2] We assign the samples to the closest cluster center. Here, we have used the rectangular distance to calculate the distance of each data point from the cluster centers.

[3] We update the cluster centers. For this, we consider the average of the data in each cluster as the center of the cluster.

[4] We repeat steps 2 and 3 until the new centers of the clusters are equal to the centers of the previous step.

performance evaluation

The efficacy of the proposed methodology has been assessed via the use of several statistical indicators, including sensitivity, specificity, and precision, with the aim of establishing the dependability and accuracy of the examination. These metrics have been used to evaluate the efficacy of the methodology. The assessment of a diagnostic test's sensitivity is conducted to ascertain its efficacy in accurately detecting individuals who exhibit a positive manifestation of an ailment. The evaluation of individuals' features is conducted in order to assess the extent to which healthy patients may be differentiated from those with the disease.

The accuracy of a diagnostic test may be determined by computing the ratio of correctly identified cases to the total number of samples used in the research. The acronyms TP, TN, and FN correspond to the concepts of sensitivity, specificity, and accuracy, respectively. When it has been ascertained that the patient really has the condition and the diagnosis is deemed to be precise. The concept of Treatment Not Necessary (TN) refers to situations where there is no need for medical intervention due to the absence of any identifiable sickness. Both categories TP and TN are considered legitimate choices. Nevertheless, it is important to note that no medical test can provide absolute assurance of a diagnosis.

For example, a false positive result (FP) refers to a diagnostic outcome that implies the existence of a disease in a patient who does not really possess the said condition. Conversely, a false negative diagnosis (FN) denotes the absence of a disease in a patient who does, in fact, have the disease. Both of these diagnostic outcomes may arise in

cases where the patient really presents with the condition. Both the categories FP and FN are inaccurate.

$$\text{Sensitivity} = TP/(TP + FN) \quad (1)$$

$$\text{Specificity} = TN/(TN + FP) \quad (2)$$

$$\text{Accuracy} = (TN + TP)/(TN + TP + FN + FP) \quad (3)$$

A portion of the outcomes obtained from executing the k-means algorithm on the specified dataset is shown in Fig. 2. The numerical values are shown in the first column, followed by the classification values in the second column. The third column displays the distance measurements between the data points and the centroid of their corresponding classes. The first column comprises numerical data, the subsequent column has categorization values, and the last column comprises distance measurements. The first column presents the numerical values, the subsequent column provides the categorization values, and the last column presents the distance measurements. Within the framework of the class number section, it becomes evident that the numeral 1 denotes an individual with a state of optimal well-being, while the numeral 2 signifies an individual plagued by a pathological condition. In contrast, the numeral 2 signifies an individual experiencing a state of affliction. Conversely, an individual whose general health is assessed as a three is seen to be in a satisfactory state.

points	cluster	Distance
1.0000	2.0000	3.8500
2.0000	2.0000	2.6269
3.0000	2.0000	1.8291
4.0000	1.0000	2.7019
5.0000	1.0000	2.3489
6.0000	1.0000	1.8941
7.0000	1.0000	3.5245
8.0000	1.0000	3.0304
9.0000	2.0000	1.9376
10.0000	2.0000	3.0362
11.0000	1.0000	2.2609
12.0000	1.0000	2.2948
13.0000	2.0000	2.1823
14.0000	1.0000	2.3504
15.0000	1.0000	2.8906
16.0000	1.0000	2.0167
17.0000	1.0000	2.6902
18.0000	1.0000	1.9180
19.0000	1.0000	1.8686
20.0000	1.0000	1.8057
21.0000	2.0000	2.8246
22.0000	1.0000	3.2257

Fig. 2. is a part of the results of the implementation of the K-means algorithm on the desired data.

Table 5 shows the results of this clustering method. The first column shows the classification accuracy of the model.

Table 5. Results of K-means clustering method

Method	**Accuracy**	**Specificity**	**Sensitivity**
k-means	81, 0 %	90, 0 %	70, 0 %

2- Multilayer neural network

This section employs a neural network model to forecast the occurrence of heart disease among hospital visits, using information pertaining to the distribution factors. Initially, the primary dataset, consisting of 279 visits, is partitioned into three distinct categories:

1. 80% of train data
2. 10% test data

3. 10% data validation

Subsequently, the neural network undergoes training using the training data. Ultimately, the network's performance is assessed by using the test data.

The structure of the neural network

Here, a map and pattern recognition neural network is used, which uses the posterior neural network and transit transfer function to train the network. The internal structure of the neural network is shown in Fig. 3.

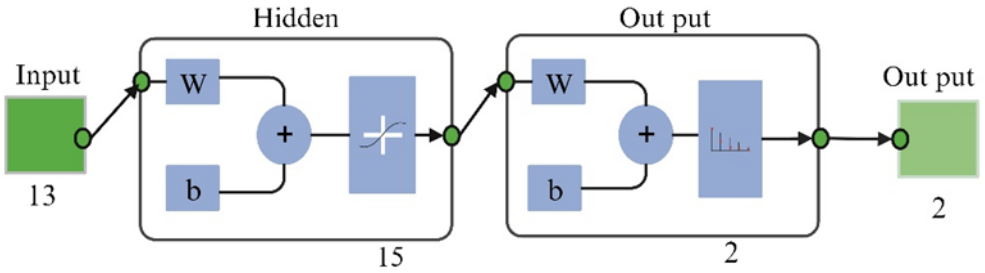


Fig. 3. The internal structure of the neural network

The matrix shown in Fig. 4 displays the first two diagonal cells, which indicate the quantity and proportion of accurate classifications made by the network that underwent training. An instance of sample 148 is accurately categorized as an individual in good health. The matrix in question represents about 49.8% of the total population of 297 people. In a similar vein, a total of 119 instances were accurately classified as persons with a disease. This figure represents 40.1% of the total population.

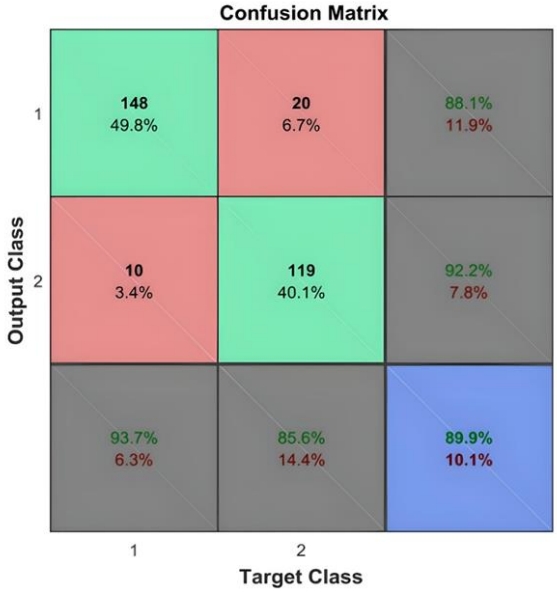


Fig. 4. Confusion matrix resulting from the implementation of the neural network

A total of 20 people who are inappropriately unwell have been misclassified as healthy individuals, accounting for 6.7% of the whole dataset consisting of 297 observations. In a similar vein, it was observed that 10 samples from the healthy group were erroneously labeled as sick, accounting for around 3.4% of the whole dataset.

Among a total of 168 forecasts pertaining to the state of health, it has been seen that 88.1% of these predictions are accurate, while the remaining 11.9% are deemed to be incorrect. Among the 129 individual patient forecasts, a notable 92.2% were accurately classified, while the remaining 7.8% were classified incorrectly. Among a sample of 158 persons who were deemed to be in good health, it was observed that 93.7% of them were accurately classified as healthy individuals, while the remaining 6.3% were incorrectly classified as sick. Among the total sample size of 139 patients, it was observed that 85.6% of the

people were accurately identified as patients, while the remaining 14.4% were classified as healthy persons.

Overall, the accuracy rate of the forecasts is 89.9%, with the remaining 10.1% falling into the category of inaccuracies. As seen in the aforementioned image, the neural network demonstrates a classification accuracy of 89.9% when applied to the specified dataset.

Based on the observations presented in Fig. 5, there is a clear decrease in the error rate of the validation, test, and training data during the initial stages. However, from epoch 14 onwards, a significant decrease is observed in the training error, while the test and validation errors show an increase. This pattern indicates that the system is experiencing overfitting when the error rate of the training data decreases while the error rate of the validation data increases. Consequently, if the training operation encounters a validation failure and the number of consecutive failures surpasses a predetermined threshold (in this instance, 6 epochs), the operation is halted.

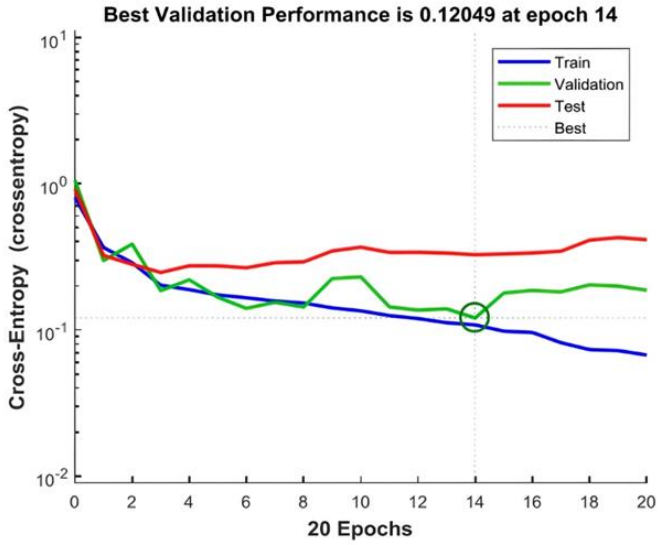


Fig. 5. Data error variation diagram

Best validation: the best efficiency and accuracy for validation operations.

Fig. 6 displays the variations in gradient during each era. The values of the changes in the gradient of the training data, which is equivalent to the gradient of the error and the objective function, are indicative of the underlying principles. As previously discussed in the validation review section, the occurrence of a validation error six consecutive times prompts the termination of the training process to avoid overfitting.

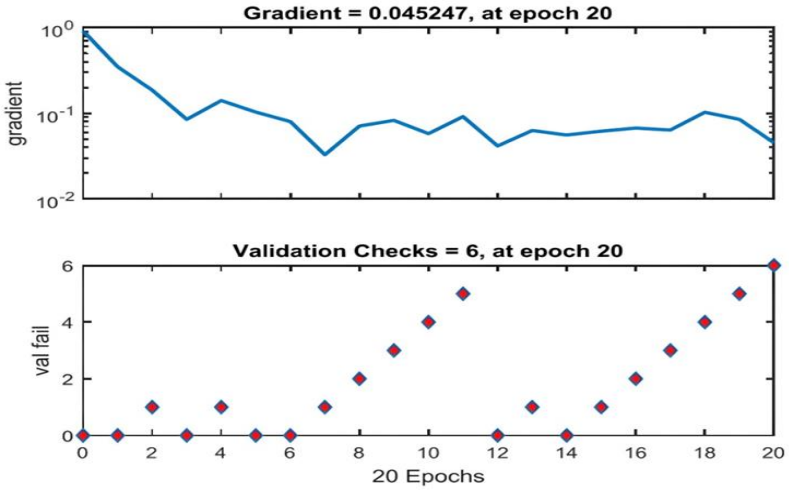


Fig. 6. Graph of gradient changes and validation failure

The error values, denoted as e, are divided into 20 intervals to construct a histogram, as seen in Fig. 7. A substantial portion of the data is concentrated in the interval representing zero error, indicating that the samples conform to a normal distribution and the network has successfully converged.

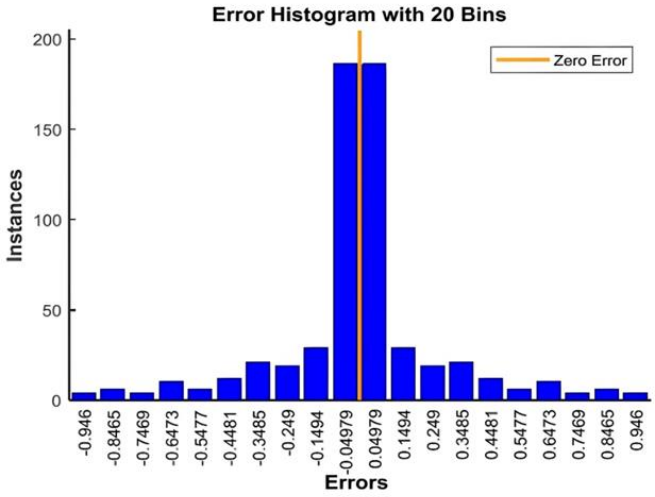


Fig. 7. Error histogram diagram

The ROC curve shown in Fig. 8 illustrates that the rate of accurate two-class classification significantly surpasses that of incorrect classification. In this context, the greater the displacement of the vectors representing the classes from the upward 45-degree line, the more favorable the outcome, as indicated by a perpendicular angle with respect to this line.

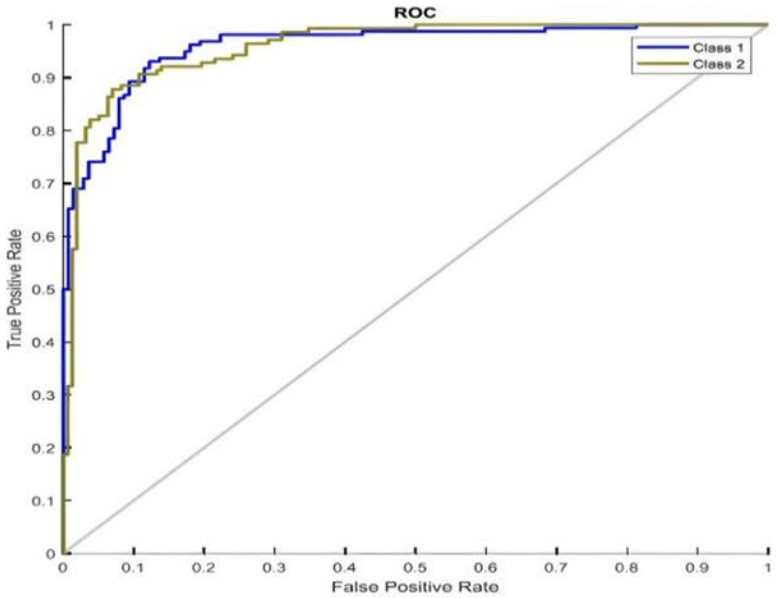


Fig. 8. ROC diagram

3- The results of SVM implementation

In this study, the necessary data were subjected to modeling using support vector machines (SVM) with linear and radial basis function (RBF) kernels. The confusion matrix is computed for each of the two kernels, and subsequently, the accuracy is derived from both techniques. The findings obtained indicate that the SVM-Linear Kernel exhibits worse accuracy compared to both the neural network and the RBF-SVM Kernel, as seen in Table 6.

Table 6. ACCURACY resulting from the implementation of three neural network and K-means algorithms

Method	**Accuracy**
Svm- Linear Kernel	84%
Svm- RBF Kernel	93%
Neural network	89.9%
K-means	81%

4. Discussion

In light of the worldwide importance of cardiac disorders as a leading cause of death, this research aimed to construct a model using medical data from patients with comparable traits. This involved employing data mining and information extraction techniques, followed by applying machine learning algorithms to analyze the gathered information. Due to the considerable worldwide impact of cardiac problems as a leading cause of death, the present research was undertaken to achieve this goal. Given the global prevalence of cardiac diseases as a prominent cause of mortality, the objective of this study was to develop a model. The significant impact of cardiovascular diseases on worldwide mortality rates has underscored the need to prioritize efforts in this area. The primary objective of this research was to address the significant global impact of cardiovascular diseases as a leading cause of mortality. The study aimed to collect pertinent information in order to contribute towards the attainment of this objective. It would be beneficial to support healthcare professionals in their endeavors to anticipate the early diagnosis of this disease proactively. This might be achieved by offering them support or aid. Providing help would serve as the mechanism through which this aim would be achieved. This discovery could mitigate the loss of human lives across diverse populations. As a result, it is plausible to begin treatment at an earlier stage, ultimately leading to improved outcomes.

The RBF-SVM kernel approach demonstrated superior accuracy when compared to the other three distinct strategies that were assessed for their potential applicability. Currently, the overall accuracy is around 93%. The part positioned at the top of the figure, Fig. 6, compiles results from previous research initiatives undertaken in this location. The trials were done at the same place throughout their duration. The experiments mentioned above were conducted at the specified location, should there be any inquiries about this matter. The guidelines proposed in this composition suggest that healthcare establishments should enhance their engagement with researchers to protect patient confidentiality while effectively gathering precise medical information. Given the need for hospitals to enhance collaboration with researchers in the field, the abovementioned proposal has been put up. Ultimately, the objective of this joint endeavor is to maximize the potential advantages derived from the findings obtained via the use of a wide array of research approaches.

5. Conclusion

In this paper, the use of hospital managers' results from data mining of hospital information systems to develop an intelligent model using ML techniques is discussed. The goal was to increase the accuracy of predictions and facilitate more effective decision-making in patient treatment and to recognize the importance of hospital managers' decision-making approaches in advancing the hospital's goals and addressing patients' treatment challenges. Three models, namely "k-means, SVM, and neural network," are widely used classification methods in the fields of data mining and ML. These models were applied to predict cardiac disease, and their predictive performance was evaluated and compared. The findings showed that the neural network model, characterized by a multi-layered perceptron architecture, achieved a classification accuracy of 89.9% when applied to the test dataset. However, the SVM using the radial basis function kernel showed increased accuracy and achieved a level of 93%.

For future work, the detection of heart failure can be investigated by utilizing new bio-signals and a knowledge-enhanced neural network and comparing its performance with existing methods.

REFRENCES

[1] K. Saxena and R. Sharma, “Efficient heart disease prediction system,” *Procedia Comput Sci*, vol. 85, pp. 962–969, 2016.

[2] J. Alcalá-Fdez *et al.*, “KEEL: a software tool to assess evolutionary algorithms for data mining problems,” *Soft comput*, vol. 13, pp. 307–318, 2009.

[3] J. Derrac, J. Luengo, J. Alcalá-Fdez, A. Fernández, and S. Garcia, “Using KEEL software as a educational tool: A case of study teaching data mining,” in *2011 7th International Conference on Next Generation Web Services Practices*, IEEE,

2011, pp. 464–469.
[4] S. B. Patil and Y. S. Kumaraswamy, “Extraction of significant patterns from heart disease warehouses for heart attack prediction,” *IJCSNS*, vol. 9, no. 2, pp. 228–235, 2009.
[5] M. Ilayaraja and T. Meyyappan, “Efficient data mining method to predict the risk of heart diseases through frequent itemsets,” *Procedia Comput Sci*, vol. 70, pp. 586–592, 2015.
[6] S. Mallik, A. Mukhopadhyay, and U. Maulik, “RANWAR: rank-based weighted association rule mining from gene expression and methylation data,” *IEEE Trans Nanobioscience*, vol. 14, no. 1, pp. 59–66, 2014.
[7] C. Yadav, S. Wang, and M. Kumar, “An approach to improve apriori algorithm based on association rule mining,” in *2013 Fourth International Conference on Computing, Communications and Networking Technologies (ICCCNT)*, IEEE, 2013, pp. 1–9.
[8] M. E. Brickner, L. D. Hillis, and R. A. Lange, “Congenital heart disease in adults,” *New England Journal of Medicine*, vol. 342, no. 5, pp. 334–342, 2000.
[9] R. El-Bialy, M. A. Salamay, O. H. Karam, and M. E. Khalifa, “Feature analysis of coronary artery heart disease data sets,” *Procedia Comput Sci*, vol. 65, pp. 459–468, 2015.
[10] J. Su and H. Zhang, “A fast decision tree learning algorithm,” in *Aaai*, 2006, pp. 500–505.
[11] S. Mardikyan, İ. Aksoy, and B. Badur, “Finding hidden patterns of hospital infections on newborn: A data mining approach,” *İstanbul Üniversitesi İşletme Fakültesi Dergisi*, vol. 39, no. 2, pp. 210–226, 2010.
[12] V. Špečkauskienė and A. Lukoševičius, “Methodology of Adaptation of Data Mining Methods for Medical,” *Elektronika ir Elektrotechnika*, vol. 90, no. 2, pp. 25–28, 2009.
[13] E. Liljegren, “Usability in a medical technology context assessment of methods for usability evaluation of medical equipment,” *Int J Ind Ergon*, vol. 36, no. 4, pp. 345–352, 2006.
[14] I. Ufumaka, “Comparative analysis of machine learning algorithms for heart disease prediction,” *Int. J. Sci. Res*, vol. 11, pp. 339–346, 2021.
[15] M. Cueto, “The World Health Organization,” in *Global Health Essentials*, Springer, 2023, pp. 421–424.
[16] S. B. Patel, P. K. Yadav, and D. P. Shukla, “Predict the diagnosis of heart disease patients using classification mining techniques,” *IOSR Journal of Agriculture and Veterinary Science (IOSR-JAVS)*, vol. 4, no. 2, pp. 61–64, 2013.
[17] O. Maimon and L. Rokach, *Decomposition methodology for knowledge discovery and data mining*. Springer, 2005.
[18] D. Napoleon and S. Pavalakodi, “A new method for dimensionality reduction using k-means clustering algorithm for high dimensional data set,” *Int J Comput Appl*, vol. 13, no. 7, pp. 41–46, 2011.
[19] W. Zhu, N. Zeng, and N. Wang, “Sensitivity, specificity, accuracy, associated confidence interval and ROC analysis with practical SAS implementations,” *NESUG proceedings: health care and life sciences, Baltimore, Maryland*, vol. 19, p. 67, 2010.
[20] N. Singh, A. G. Mohapatra, and G. Kanungo, “Breast cancer mass detection in mammograms using K-means and fuzzy C-means clustering,” *Int J Comput Appl*, vol. 22, no. 2, pp. 15–21, 2011.
[21] R. Chitra and V. Seenivasagam, “Heart attack prediction system using fuzzy C means classifier,” *IOSR J Comput Eng*, vol. 14, no. 2, pp. 23–31, 2013.
[22] H. Kahramanli and N. Allahverdi, “Design of a hybrid system for the diabetes and heart diseases,” *Expert Syst Appl*, vol. 35, no. 1–2, pp. 82–89, 2008.
[23] S. J. Khiabani, A. Batani, and E. Khanmohammadi, “A hybrid decision support system for heart failure diagnosis using neural networks and statistical process control,” *Healthcare Analytics*, vol. 2, p. 100110, 2022.
[24] D. Hassan, H. I. Hussein, and M. M. Hassan, “heart disease prediction based on pre-trained deep neural networks combined with principal component analysis,” *Biomed Signal Process Control*, vol. 79, p. 104019, 2023.