Journal of Theoretical and Applied Information Technology

15th September 2023. Vol.101. No 17

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

A PRACTICAL MODEL FOR USING CLASSIFICATION ALGORITHMS TO ENHANCE CARDIAC DIAGNOSIS

LIQAA ALWAN ADHOOB[1], RAZIEH ASGARNEZHAD[2, *]

[1]Department of Computer Engineering, Isfahan (Khorasgan) Branch, Islamic Azad University, Isfahan, Iran

[2]Department of Computer Engineering, Aghigh Institute of Higher Education Shahinshahr, 8314678755, Isfahan, Iran

E-mail: [1]leqaaali9@gmail.com, [2]r.asgarnezhad@aghigh.ac.ir

ABSTRACT

According to historical records, one of the most prevalent ailments is heart disease. All age groups are affected by this illness, including teenagers, adults, and the elderly. Since there is never a powerful and effective treatment that may significantly lessen the severity of this condition and there is always a failure in clinical cardiac situations, it is thought to be incurable. Individuals are at grave risk from heart disease since it has lately emerged as a severe condition that threatens people. All age groups, from young to the elderly, are often affected. The preparation of data and finding a fix for record failures are the major challenges in this area. Clinical heart failure data, where a successful, high-performance strategy was suggested to treat loss and ameliorate heart disease. The authors of this paper used the method for cleaning data and used the base classification techniques. For cleaning data, the worthwhile preprocessing step is employed to recognize the missing values and outliers in conjunction with KNN. Two experiments were handled to replace these challenges with RF, DT, and KNN. We achieved success using classification algorithms to greatly forecast and enhance the performance of heart disease. Through the outcomes of this study, this model showed a clear advantage over its competitors. The highest results obtained were 98% for all evaluation metrics. It means that we have a 3% improvement. It demonstrates that this model will make accurate predictions and enhance the performance of the data we have focused on.

Keywords: *Data Mining, Pre-processing, Heart Disease, classification, Machine Learning Techniques*

1. INTRODUCTION

According to historical records, one of the most prevalent ailments as heart disease. All age groups are affected by this illness, including teenagers, adults, and the elderly. Since there is never a powerful and effective treatment that may significantly lessen the severity of this condition and there is always a failure in clinical cardiac situations, it is thought to be incurable. This study encouraged several researchers to look for strategies to counteract and make up for the shortcomings that mostly happened. By offering methods and algorithms that work to lower the risk of this disease, anticipate it, and improve its performance, the suggested model enhances the performance of this disease. Avoid smoking, exercise, and regulate weight since excess weight is dangerous and harmful to the patient. You should also avoid crises, nervous circumstances, and stress because they weaken the heart and cause the pulse to stop rapidly. The aortic valve, mitral valve, tricuspid valve, and mitral valve are the three valves that open and shut to guide blood flow from the heart. When one of these valves fails, it may do so to a number of different causes that result in narrowing, leakage, or a failure to close normally. It is one of the issues the condition has. Additionally, because it is a pulsing component of the human body, this condition is hazardous when ignored.

This study's motivation is the importance of heart disease and the related available data on all existing conditions. Heart disease affects other diseases and the whole body system. In that case, this subject is the most important challenge in the health care system.

In this study, the hypothesis that can define the proposed method is the utility of the preprocessing step to improve the evaluation metrics. On the other hand, the usage of cleaned data can improve these metrics significantly.

However, our work is superior to theirs in terms of the methods used and the results obtained, where our work showed that its results are superior

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

to those of its peers. In 2015, the authors proposed a method and comprehensively reviewed the data on fatal and non-fatal rheumatic heart diseases for the period between 1990 and 2015. They achieved good results. According to the authors' investigation, there were 319,400 fatalities (95% confidence interval: 297,300 to 337,300). Their good work is confirmed by one death from rheumatic heart disease, but our work is superior to it [1]. Musicians presented their systematic review and meta-analysis of longitudinal observational data in 2016 using a technological approach. This disease is extremely hazardous because of its severity and the lack of a treatment that may prevent or significantly reduce the frequency of infections. The purpose of this study was to develop a method for investigating loneliness and social isolation by conducting a systematic review and meta-analysis. They demonstrated the quality of their job, which has been projected to enhance this disease's functionality. In terms of outcomes and approach, our study performs better than this work [2]. We were successful in treating the clinical failure of cardiac disease and enhancing its function as well [3].

Risk factor for heart disease: Evidence points to a number of risk factors for heart disease: age, sex, high blood pressure, blood lipids, diabetes, tobacco smoking, processed meat consumption, excessive alcohol consumption, sugar, family genetics, obesity, lack of physical activity, psychosocial factors and air pollution. While the contribution of each risk factor is different between different communities or ethnic groups, the overall contribution of these risk factors in epidemiological studies is significant. Some of these risk factors, such as age, gender, or family history, are immutable. But many of these important cardiovascular risk factors can be modified by lifestyle changes, social changes, drug treatment, and prevention of high blood pressure, lipids, and diabetes.

Hear disease diagnosis methods: Most of the time, the treating doctor does a lot of investigations to get enough information to diagnose your disease and make the right decision about the appropriate treatment. The set of tests that your doctor requests for you may not be necessary in another person with the same condition [21]. The results of a survey may provide information that reveals the need for more surveys to complete the information. You may also need to repeat the same test several times to determine exactly how your heart is responding to different medications, surgical procedures, or other treatments. Medical history is also a determining factor in requesting appropriate examinations [22]. There are several methods to diagnose heart disease, which we will discuss below.

We provide a technical model to enhance heart disease, lessen its severity, and considerably forecast it due to the illness's prevalence. We suggest a model with two experiments in it. The impact of classification approaches on the preprocessing step is examined in the first experiment. These methods were used in the second experiment without any preprocessing. We tackle the challenges of missing values and outliers during the preprocessing stage. These methods are applied in this study to get successful outcomes.

The technical model presented in this article uses KNN, NN, NB, SVM, RF, and DT with or without preprocessing and is based on classification algorithms. When these methods were used with the Rapid Miner program, positive outcomes were attained. We have demonstrated that our work surpasses earlier work in terms of outcomes and better prediction in reducing heart disease by analyzing the data that we got from the UCI website. Our model is accurate and makes solid predictions about how this disease will progress.

This paper is organized as follows: Sect. 2 shows a summary of the related works followed by the proposed method is presented in Sect. 3 and evaluated by the experiment explained in Sect. 4. Finally, the paper presents a conclusion in Sect. 5.

2. RELATED WORK

Several works suggested predicting the autism diagnosis in the years between 2015 and 2021. We summarized some of the significant works herein.

The authors wrote about a random experiment in 2015. Through a randomized trial approach, the two musicians developed a method to enhance the performance of cardiac disease and better anticipate it. They included qualified individuals between the ages of 18 and 80 with concrete proof that they have ischemic heart disease. They worked on these records to enhance their performance, and their experiences in the present article support the success and accuracy of our work [4]. The authors were driven to concentrate on heart data and offer their approach in 2016 as a result of the illness's spread and the dearth of alternatives to treat heart disease and lessen this damage. It was suggested that they clinically combine the outcomes into temporal

research to gather information. They applied this approach to clinical research to better understand the cardiac disease and how it functions and got the following outcomes: They established that in the 1960s and 1950s, an industry-sponsored research program helped make their models and exploration successful by enhancing model performance [5].

A novel model for presenting the specific features with the use of machine learning (ML) techniques was developed by the authors in 2019. With the use of ML algorithms like the random forest, they were able to greatly enhance and forecast cardiac disease (RF) Given that they were able to achieve an accuracy of 88.7%, their work must have gone well [6]. In 2017, the authors analyzed the correlations between consuming 10 particular nutrients and death from natural causes. Their trials showed that their model was acceptable and may enhance classification performance, therefore they offered a model to lower the risk of heart disease data and the associated ambiguity. It is a significant issue because this dataset doesn't have any treatment. As a result, we proposed strategies in our article to significantly enhance the performance of this illness, and it is anticipated that this would have the opposite effect of the prior studies they presented, but not to the same extent as our work [6].

The authors provided a technological methodology to project childhood obesity in 2020. The measures that dealt with teenage weight were the most crucial. For Americans who are 35 years of age or older, they offered a policy model of coronary artery disease that is a computer and natural simulation. Through these models, they were able to make better predictions. Their model produced good results, but the need to find alternative treatments for heart disease made it necessary to find the best ways to treat the condition. This prompted us to present a technical model in our paper that produced results that outperformed both our earlier work and this work, where we attest that our work excelled in treating heart disease [7]. The authors detailed the risks of visceral heart in middle-aged and elderly patients in 2021 and then assessed the long-term outcomes using a technological model. The work of these writers is regarded as excellent since they accurately anticipated heart disease, yet it is challenging to discover a cure for such a devastating illness. Research and effort are needed to develop a technological approach that produces good results in order to increase its performance, and this is what we have done in this study. Compared to earlier research, ours has produced good findings and significantly improved heart disease prediction [8].

The current writers did a lot of connected works study. The model using clustering techniques that used SVM, RF, DT, NB, NN, and KNN with or without preprocessing to replace the missing value with the mean to detect outliers was shown to be more accurate. Through the use of methodologies, we were able to predict heart disease with high accuracy by utilizing missing values and significant prediction, which produced excellent results when the data were divided into five groups. It demonstrates and validates that our work performs better than any earlier work. The model we described is outstanding, highly predictive, and provides superior outcomes for treating cardiac disease, as has been demonstrated [9].

Authors in [10] a new classification for cardiovascular diseases using deep learning to detect signs of heart diseases. The used method including VGG16, ResNet, MobileNet, Inception V3, and Xception. The highest results were 100% for F1 and 100% for accuracy.

A literature reviews based machine learning on heart disease diagnosis was proposed in [11]. 49 references considered in the literature including: heart disease type, algorithms, applications, and solutions.

A New Framework for Diagnosis of Cardiovascular proposed in [12]. The authors investigated various heartbeat acoustic. Long short-term memory (LSTM)-convolutional neural network (CNN), recurrent neural network (RNN), LSTM, Bi-LSTM, CNN, K-means, and support vector machine (SVM) as methods were used. The highest accuracy was 93.38%. Authors in [13], tried to give a perspective of the epidemiology and highlight challenges and proposed an algorithm for the classification of cardiac organ. Authors in [14] produced a system for patients undergoing coronary CT angiography (CCTA). They wanted to enhance the initial reporting system through considering new technical developments. Authors in [15] investigated the effect of the classification and regression tree (CART) modeling on heart disease. The research was included 263 IHD and 181 DCM patients. They employed a CART algorithm to create classification models based on HRV features. The highest accuracy was 73.3%. Authors in [16] showed a review on artificial intelligence techniques for diagnosis of cardiovascular diseases.

The authors investigated the diagnosis of myocarditis in Cardiac Magnetic Images through CycleGAN and Deep PreTrained Models [17]. They applied artificial intelligence (AI) techniques

for diagnosis of myocarditis. Also, they offered a new method through deep learning (DL) models. The highest accuracy was 99.33%.

Authors in [18] done a research on the diagnosis of cardiac abnormalities based on heart sound signal. In this article, they suggested a new feature extraction method based on fuzzy matching. Convolutional features of test signal and Gaussian wavelets were applied. They handle a tradeoff between matching degree and matching energy features. To evaluate performance of the research, machine learning algorithms such as K-nearest neighbor, support vector machine, random forest and multilayer perceptron were used. They reveal that the best classification accuracy belonged to random forest with 96.5%. Accuracy, sensitivity and specificity obtained 99.0%, 99.4% and 99.7%, respectively.

The authors in [19] proposed a framework for diagnosis of the Type 2 diabetes mellitus (T2DM) special cardiovascular event. Authors in [20] investigated the effect of genetic algorithm for diagnosis Cardiac. They obtained the greatest results for diagnostic. They improved the performance of classification up to 5%. The authors showed their framework for diagnosis heart disease was effective.

To sum up, we can review some related works in this scope and compare them. The model expressed in [6] used ML techniques in 2019. They achieved an accuracy of 88.7% and their work was well done. In 2021, the long-term outcomes using a technological model showed an accuracy of 94.3%. This work was excellent since they accurately anticipated heart disease [8]. In [23], the authors used ML techniques and achieved an accuracy of 95% and this work was well. Compared to earlier research, ours has produced good findings and significantly improved heart disease prediction. Our highest accuracy obtained 98% that it shows a 3% improvement approximately.

3. THE PROPOSED MODEL

We have carried out two studies to investigate the influence of categorization and ML techniques both before and after preprocessing stages. When preparing KNN, we used a method to address the outlier problem. The mechanism that replaces missing values with the mean was also used to correct missing values. SVM, RF, DT, NN, NB, and KNN are some of the ML approaches employed. The assortment is also employed. It's crucial to give a quick rundown of the tactics employed. The phases of our model are described in this section. We processed the data we obtained from the UCI website. Information about instances of diabetes. In this study, we divided the participants into the first track and combined the RF, DT, and KNN classifiers utilizing SVM, RF, DT, NN, NB, and KNN classification techniques using the Rapid Miner tool without any preprocessing. Then, utilizing ensemble techniques in conjunction with preprocessing, the highest values possible were obtained, surpassing the alternatives. Figure 1 shows the steps of the proposed model.

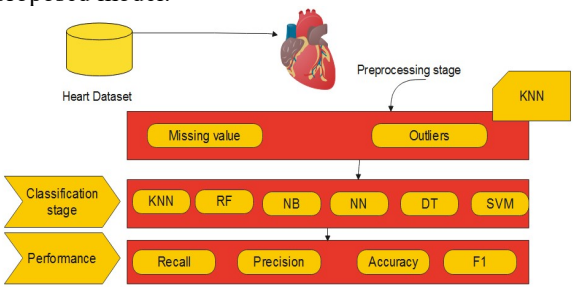


Figure 1: The proposed Model

2.1 Data Collection

Since this data has erroneous and missing values, we have obtained the clinical heart failure data from the UCI website. We have used preprocessing approaches to identify outliers and substitute missing information in order to enhance this disease's performance, make predictions that will be greatly better, and make the process of resolving the problem of failure in instances easier. Clinical algorithms were used first, followed by clustering algorithms including KNN, NB, NN, NB, RF, and SVM, which produced outstanding results for predicting chronic heart disease and enhancing its functionality. It demonstrates how superior our work is to everything done before. Because they provide efficient and positive results, these strategies are regarded as the finest among all other techniques. Since heart disease is seen as an incurable condition.

2.2 Pre-processing Stage for Missing Values

Since clinical heart failure records sometimes have stray and missing information, we retrieved data for those records here. We pre-treated them by employing methods like substituting the missing value with the mean and identifying outliers using the Rapid Miner tool. We were able to substitute the missing numbers, which led to fantastic outcomes. With the addition of additional values, our work's effectiveness was demonstrated, just as it had been with the pre-treatment; it improved the performance of the data and produced fresh, better than expected findings. This demonstrates how our work will benefit and help patients with heart failure.

2.3 Classification Stage through Classification and ML Methods

The classification process makes use of SVM, RF, NN, NB, KNN, and DT. These techniques are currently being utilized to improve classification performance and generate output with good features and high accuracy. The categorization stage is when the preprocessing stage is combined to create effective results. In this investigation, the greatest accuracy was 98%, which is thought to be the best accuracy. It has a higher value than its predecessors and takes into consideration high accuracy.

Support Vector Machines (SVM): is a supervised machine learning technique used in both regression and classification. Although we also refer to them as regression issues, categorization is where they fit in best. Finding a hyperplane in an N-dimensional space that clearly classifies the data points is the goal of the SVM method. The number of features determines the hyperplane's size. The hyperplane is essentially a line if there are just two input characteristics. The hyperplane turns into a 2-D plane when the number of input characteristics is raised to three. Imagining something with more than three characteristics gets challenging.

Decision Tree Algorithm (DT): The supervised learning algorithm family includes the decision tree algorithm. The decision tree technique, in contrast to other supervised learning methods, is capable of handling both classification and regression issues. By learning straightforward decision rules derived from previous data, a decision tree is used to generate a training model that may be used to predict the class or value of the target variable (training data). In decision trees, we begin at the tree's base when anticipating a record's class label. We contrast the root attribute's values with that of the attribute on the record. We follow the branch that corresponds to that value and goes on to the next node based on the comparison.

Random Forest Algorithm (RF): Popular machine learning algorithm Random Forest is a part of the supervised learning methodology. It may be applied to classification and regression issues in machine learning. It is predicated on the idea of ensemble learning, which is the procedure of merging many classifiers to address a complicated issue and enhance the model's performance. Random Forest, as the name implies, is a classifier that uses a number of decision trees on different subsets of the provided dataset and averages them to increase the dataset's prediction accuracy. Instead of relying just on one decision tree, the random forest considers each tree's forecast and decides based on the predictions that received the most votes.

Neural networks (NN): Artificial neurons, which are a set of interconnected units or nodes that loosely resemble the neurons in a biological brain, are the foundation of an ANN. Like the synapses in a human brain, each link has the ability to send a signal to neighboring neurons. An artificial neuron may signal neurons that are linked to it after receiving, processing, and acting on inputs. The output of each neuron is calculated by some non-linear function of the sum of its inputs, and the "signal" at a connection is a real number. Edges refer to the connections. The weight of neurons and edges often changes as learning progresses. The weight alters a connection's signal intensity by increasing or decreasing it. A neuron's threshold may be set that a signal is only transmitted if the aggregate signal crosses that threshold.

Naive Bayes (NB): Naive a probabilistic classifier called Bayes returns the likelihood that a test point belongs to a class rather than the test point's label. It's one of the most fundamental Bayesian network models, however by using kernel density estimates, it could be more accurate. Unlike many other ML algorithms, which often do both classification and regression tasks, this approach is exclusively relevant to classification problems. Naive The assumptions made by the Bayes algorithm are regarded as naïve since they are almost impossible to verify with actual data. To compute the sum of the individual probabilities, conditional probability is used. This indicates that the method takes into account the existence or absence of a particular class characteristic that is unrelated to the presence of other features.

K-Nearest Neighbor(KNN): When there is little or no prior knowledge about the distribution of the data, K-nearest-neighbor (KNN) classification should be one of the first options considered. It is one of the most fundamental and basic classification techniques. When accurate parametric estimates of probability densities are unknown or challenging to calculate, the discriminant analysis must be used. This is when the K-nearest-neighbor classification was created. Fix and Hodges presented the k-nearest neighbor rule, a non-parametric technique for pattern categorization, in a 1951 unpublished study from the US Air Force School of Aviation Medicine.

4. EXPERIMIENTS AND RESULTS

4.1 Experiment I

We go through our third example of employing highly precise algorithms in this

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

procedure. It typically yields favorable results when predicting the best outcomes, but does not show favorable results when put to use. In order to evaluate the downloaded data in connection to diabetes, we used the Rapid Miner application together with SVM, RF, DT, NN, NB, and KNN. Table 1 displays the recall, precision, accuracy, and f-measure values for categorization, RF, and DT. The results of our work were excellent. The accuracy was good, reaching 98%. These processes are anticipated and are among the most popular ways to improve categorization performance.

Table 1: The obtained results through classification with pre-processing with RF and DT with Rapid Miner tool

Classifier	Precision	Recall	Accuracy	F1
SVM	98 %	98 %	98 %	98 %
RF	97%	97%	97%	97%
DT	95%	95%	95 %	95 %
NN	95%	95%	95%	95%
NB	96%	96%	96%	96%
KNN	97%	97%	97%	97 %

In order to offer high values for accuracy, recall, precision, and F1, we used the techniques listed in Table 1. These results showed high and distinct results. The highest accuracy rate using DT was 98%. These are considered high values that came from our study and were applied in this study, where high-precision changes were performed to improve and foresee the troublesome and widespread diabetes at younger ages. Additionally, it demonstrates that our work is superior to others and will have favorable effects. Additionally, it will perform better than competitors and yield reliable results.

4.2 Experiment II

The outcomes of our experiment using preprocessing to improve and forecast chronic diabetes are shown in Table 2. We performed the collection procedures using a partial data set for diabetes provided by the UCI, which contained stray and missing values. The SVM, RF, DT, KNN, NN, and NB algorithms were used. The maximum degree of accuracy with DT was 98%, which is thought to be the optimum value for treating diabetes and forecasting the best outcomes. Table 2 displays high-accuracy values for accuracy and F1, as well as accuracy and each recall and precision.

Table 2: The obtained results through classification with pre-processing with DT and KNN with Rapid Miner Tool

Classifier	Precision	Recall	Accuracy	F1
SVM	98 %	98%	98%	98%
RF	98%	98%	98%	98%
DT	98%	98%	98%	98%
NN	95%	95%	95%	95%
NB	96%	96%	96%	96%
KNN	95%	95%	95%	95%

4.3 Evaluation Metrics and Discussion

Accuracy, precision, recall, and F1 were used to evaluate. Table 3 provides a definition of these metrics [24].

Table 3: Parameters definitions

Metrics	Evaluation
Accuracy	$(TP+TN)/(P+N)$
Precision	$(TP)/(TP+FP)$
Recall	TP/P
F1	$\frac{2\times precision\times recall}{precision+recall}$

Using the Rapid Miner tool, we used ensemble approaches in this case, including SVM, RF, DT, NN, NB, and KNN, both with and without preprocessing. These techniques show outstanding expertise and overwhelming effectiveness in delivering results [25]. Our assignment was further divided into two parts for each of two parts. Without any preprocessing, we combined DT and RF with categorization in the first segment. Recall, precision, accuracy, and F1 scores were all equal to 98%, 98%, 98%, and 98%, as shown in Table 1.

This demonstrates the dependability, predictive accuracy, and value of our study. As shown, we used the second branch of the first section of SVM, RF, DT, NN, NB, and KNN to get good results in Table 2 without any aggravating treatment. This is a great number that raises both the performance and prediction accuracy of the classifier. With a DT accuracy of 98%, Table 2's categorization technique produced the greatest results for us. However, as shown in Table 1, we now applied the preprocessing, DT, and RF classification techniques. The findings in Table 3 were fairly good; the greatest accuracy was 98% for SVM, RF, NN, KNN, and NB. Our rating system's performance will be significantly forecasted and enhanced, and it exhibits the best value in our sector. In Table 1, we utilized the methods for classification with DT, KNN, and preprocessing.

Table 4 compares our labor to that of other workers. Our efforts were equally precise and more effective than those of others (98%). Our study has demonstrated that applying such algorithms will result in the right course of action and favorable outcomes, thus it is highly desired for improved prediction. Using such algorithms also shows that you are doing a fantastic job since your findings are correct and your forecasts are more accurate. It is well known that our study significantly improved how the results were displayed.

Table 4: A comparison among the obtained results through classification with pre-processing and other works

	Precision	Recall	Accuracy	F1
[6]	88.7%	88.7%	88.7%	88.7%
[23]	95%	95%	95%	95%
[8]	94.3%	94.3%	94.3%	94.3%
Our work	98 %	98 %	98 %	98 %

A comparison of our work with past attempts at this issue is shown in Table 4. As our paper had a maximum accuracy of 98%, performance has to be improved. In comparison to previous models, ours improves classification performance while still requiring more development and forecast labor. Figure 2 compares our results to their counterparts. Figure 3 compares the outcomes of our work using classification, pre-processing, DT, and RF. Figure 4 compares results among our work through classification with pre-processing in conjunction with DT and KNN.

Compared to mentioned research, ours has produced good findings and significantly improved heart disease prediction. Our highest accuracy obtained 98%. It means that we have a 3% improvement rather than the proposed models in [23], a 10% improvement rather than [6], and a 4% improvement rather than [8] approximately. The model expressed in [6] used ML techniques in 2019. They achieved an accuracy of 88.7%. In 2021, the long-term outcomes using a technological model showed an accuracy of 94.3%. [8]. In [23], the authors used ML techniques and achieved an accuracy of 95%. It revealed that a proper preprocessing step can offer a good enough improvement in the classification performance.

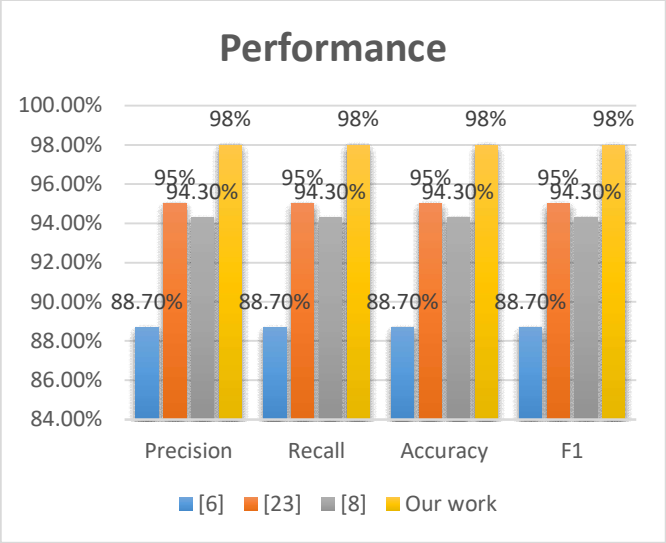


Figure 2: The Comparison results among our work and others

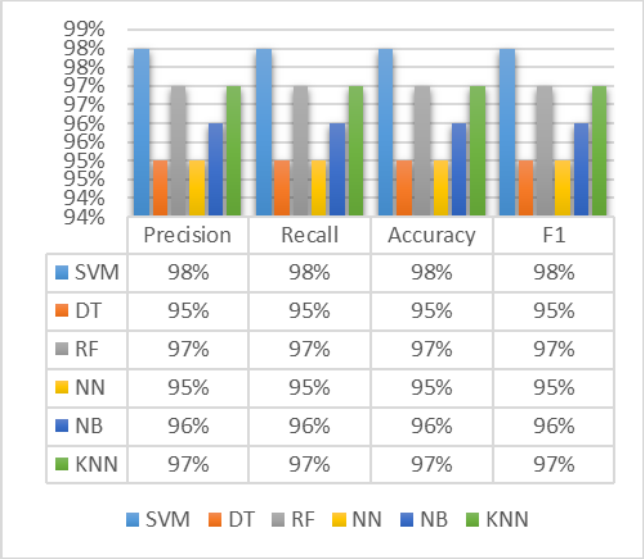
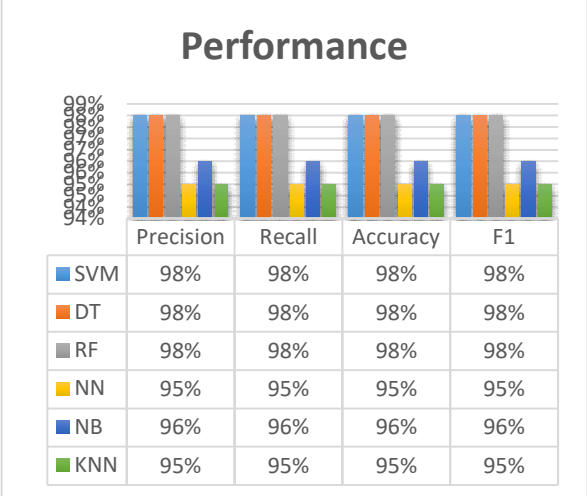


Figure 3: The comparison of the outcomes of our work using classification, pre-processing, DT, and RF



	Precision	Recall	Accuracy	F1
SVM	98%	98%	98%	98%
DT	98%	98%	98%	98%
RF	98%	98%	98%	98%
NN	95%	95%	95%	95%
NB	96%	96%	96%	96%
KNN	95%	95%	95%	95%

Figure 4: The Comparison results among our work through classification with pre-processing in conjunction with DT and KNN

5. CONCLUSION

One of the ailments that have evolved over the years is heart disease. We provide a high-resolution model that makes use of categorization methods. Numerous research provides reliable methods to diagnose this condition, but they do not investigate how it operates. Our study is broken down as follows: Using the quick miner tool, we did two tests on clinical heart failure data with missing values and outliers. We conducted two trials for this.

In this study, we proposed a model that improves the Heart's performance. This model makes use of the Rapid Miner and the classification algorithms SVM, RF, DT, NN, NB, and KNN. We used these tactics with the subsequent algorithms to provide better outcomes and more precise predictions: These methods combine the usage of KNN, DT, and RF. Each component of our work is divided into two branches, each of which contains further information on the topic. In the first portion, DT, RF, and Rapid Miner are combined with the categorization without preprocessing. These techniques were used, and the results showed that germination in terms of memory, precision, accuracy, and F1 had improved. 98%, 98%, 98%, and 98%, respectively, were the measurements for this segment. Without any preprocessing, we use classification with Rapid Miner, DT, and KNN in the second portion. The results showed that 98% was the best value for recall, precision, accuracy, and F1. The best and most accurate figure for predicting diabetes performance is this one. Next, we use ensemble together with DT, RF, and preprocessing using Rapid Miner in the third segment. Our values were really high, which is good for a more precise prognosis. The maximum accuracy was attained by SVM, RF, DT, NN, NB, and KNN, at 99%. The three most crucial figures in our essay are those mentioned above, and by using them, we may substantially forecast and enhance the performance of the category. We were able to arrive at a solution that delivers forecasts rapidly and accurately because of the high values that were also acquired. The preceding attempts were outperformed by these fantastic outcomes. The accuracy for SVM, RF, DT, NN, NB, and KNN in the fourth table was 98%. The stacking value turns out to be the best in this situation, and while the others are all excellent, it generates outstanding forecasts and will result in excellent results. The techniques we employed in this study resulted in excellent results and a superb model, which quickly improved the precision of diabetes classification and prediction. This model proved to be more effective than alternative strategies and prior initiatives.

To get the finest outcomes, we will continue to create and use new methods in our future work. We will present works in which we strive for the best outcomes, such as our work, using methods like associated, summing, and recommender systems, which will enable us to provide a lot of research in the future and enhance the functionality of some diseases and some crucial data that must be worked on and developed to enhance their functionality.

REFERENCES:

[1] P. W. Serruys, B. Chevalier, D. Dudek, A. Cequier, D. Carrié, A. Iniguez, et al., "A bioresorbable everolimus-eluting scaffold versus a metallic everolimus-eluting stent for ischaemic heart disease caused by de-novo native coronary artery lesions (ABSORB II): an interim 1-year analysis of clinical and procedural secondary outcomes from a randomised controlled trial," The Lancet, vol. 385, 2015, pp. 43-54.

[2] R. Asgarnezhad and M. Karim Qaseem, "Toward a high-accuracy hybrid system for cardiac patient data analysis using C-Means fuzzy clustering in neural network structure," Majlesi Journal of Telecommunication Devices, 2023.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

[3] N. K. Valtorta, M. Kanaan, S. Gilbody, S. Ronzi, and B. Hanratty, "Loneliness and social isolation as risk factors for coronary heart disease and stroke: systematic review and meta-analysis of longitudinal observational studies," Heart, vol. 102, 2016, pp. 1009-1016.

[4] M. M. Nentwich and M. W. Ulbig, "Diabetic retinopathy-ocular complications of diabetes mellitus," World journal of diabetes, vol. 6, 2015, p. 489.

[5] C. E. Kearns, L. A. Schmidt, and S. A. Glantz, "Sugar industry and coronary heart disease research: a historical analysis of internal industry documents," JAMA internal medicine, vol. 176, 2016, pp. 1680-1685.

[6] S. Mohan, C. Thirumalai, and G. Srivastava, "Effective heart disease prediction using hybrid machine learning techniques," IEEE access, vol. 7, 2019, pp. 81542-81554.

[7] K. Bibbins-Domingo, P. Coxson, M. J. Pletcher, J. Lightwood, and L. Goldman, "Adolescent overweight and future adult coronary heart disease," New England Journal of Medicine, vol. 357, 2007, pp. 2371-2379.

[8] M. Fedchenko, Z. Mandalenakis, K. W. Giang, A. Rosengren, P. Eriksson, and M. Dellborg, "Long-term outcomes after myocardial infarction in middle-aged and older patients with congenital heart disease—a nationwide study," European heart journal, vol. 42, 2021, pp. 2577-2586.

[9] R. Asgarnezhad, "SRV: A Striking Model based on Meta-Classifier for Improving Diagnosis Type 2 Diabetes," Journal of Advances in Computer Research, vol. 13, 2022, pp. 1-13.

[10] M. K. MAlnajjar and S. S. Abu-Naser, "Heart Sounds Analysis and Classification for Cardiovascular Diseases Diagnosis using Deep Learning," 2022.

[11] M. M. Ahsan and Z. Siddique, "Machine learning-based heart disease diagnosis: A systematic literature review," Artificial Intelligence in Medicine, 2022, p. 102289.

[12] S. Pandya, T. R. Gadekallu, P. K. Reddy, W. Wang, and M. Alazab, "InfusedHeart: A novel knowledge-infused learning framework for diagnosis of cardiovascular events," IEEE Transactions on Computational Social Systems, 2022.

[13] M. A. Talle, E. Ngarande, A. F. Doubell, and P. G. Herbst, "Cardiac Complications of Hypertensive Emergency: Classification, Diagnosis and Management Challenges," Journal of Cardiovascular Development and Disease, vol. 9, 2022, p. 276.

[14] R. C. Cury, J. Leipsic, S. Abbara, S. Achenbach, D. Berman, M. Bittencourt, et al., "CAD-RADS™ 2.0-2022 Coronary Artery Disease-Reporting and Data System: An Expert Consensus Document of the Society of Cardiovascular Computed Tomography (SCCT), the American College of Cardiology (ACC), the American College of Radiology (ACR), and the North America Society of Cardiovascular Imaging (NASCI)," Journal of cardiovascular computed tomography, 2022.

[15] A. Accardo, L. Restivo, M. Ajčević, A. Miladinović, K. Iscra, G. Silveri, et al., "Toward a diagnostic CART model for Ischemic heart disease and idiopathic dilated cardiomyopathy based on heart rate total variability," Medical & Biological Engineering & Computing, vol. 60, 2022, pp. 2655-2663.

[16] S. Singhal and M. Kumar, "A Systematic Review on Artificial Intelligence-Based Techniques for Diagnosis of Cardiovascular Arrhythmia Diseases: Challenges and Opportunities," Archives of Computational Methods in Engineering, 2022, pp. 1-24.

[17] A. Shoeibi, N. Ghassemi, J. Heras, M. Rezaei, and J. M. Gorriz, "Automatic Diagnosis of Myocarditis in Cardiac Magnetic Images Using CycleGAN and Deep PreTrained Models," in International Work-Conference on the Interplay Between Natural and Artificial Computation, 2022, pp. 145-155.

[18] W. Yang, J. Xu, J. Xiang, Z. Yan, H. Zhou, B. Wen, et al., "Diagnosis of cardiac abnormalities based on phonocardiogram using a novel fuzzy matching feature extraction method," BMC medical informatics and decision making, vol. 22, 2022, pp. 1-13.

[19] A. Unnikrishnan, R. Sahay, U. Phadke, S. Sharma, P. Shah, R. Shukla, et al., "Cardiovascular risk in newly diagnosed type 2 diabetes patients in India," PloS one, vol. 17, 2022, p. e0263619.

[20] F. Stafford, N. Krishnan, E. Richardson, A. Butters, C. Burns, B. Gray, et al., "The role of genetic testing in diagnosis and care of inherited cardiac conditions in a specialised multidisciplinary clinic," medRxiv, 2022.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

[21] D. Roa, J. Bautista, N. Rodríguez, M. D. P. Villamil, A. Jiménez, and O. Bernal, "Data mining: A new opportunity to support the solution of public health issues in Colombia," in 2011 6th Colombian Computing Congress (CCC), 2011, pp. 1-6.

[22] S. Palaniappan and R. Awang, "Intelligent heart disease prediction system using data mining techniques," in 2008 IEEE/ACS international conference on computer systems and applications, 2008, pp. 108-115.

[23] H. F. Ahmad, H. Mukhtar, H. Alaqail, M. Seliaman, and A. Alhumam, "Investigating Health-Related Features and Their Impact on the Prediction of Diabetes Using Machine Learning," Applied Sciences, vol. 11, 2021, p. 1173.