HFC: Towards an Effective Model for the Improvement of heart Diagnosis with Clustering Techniques

Razieh Asgarnezhad*
Department of Computer Engineering
Aghigh Institute of Higher Education Shahinshahr,
8314678755, Isfahan, Iran
r.asgarnezhad@aghigh.ac.ir

Karrar Ali Mohsin Alhameedawi
Department of Computer Engineering
Al-Rafidain University of Baghdad
Baghdad, Iraq
karraralimohsin125@gmail.com

***Abstract*—Heart disease pretends great danger to people, as heart disease has recently become a dangerous disease that acts as a threat to humans. It usually affects all groups from young to old. The biggest challenge in this paper is data pre-processing and discovering a solution to the failure of records Clinical heart, where an effective high-performance model is proposed to enhance heart disease and treat failure in the clinical heart failure records. The current authors applied the techniques of clustering with k-means, expectation-maximization clustering, DBSCAN, support vector clustering, and random clustering herein. Using cluster techniques, we gained good enough results for significantly predicting and improving the performance of heart disease. The goal of the model is a suggestion of a reduction method to find features of heart disease by applying several techniques. Our most important results are to predict faster and better. It indicates that the proposed model is excellent and gives excellent results. This model demonstrated a great superiority over its counterparts through the results obtained in this research. We obtained some values of 130, 980, 183, 125.133, 133, 203, and 125.800. It confirms that this model will predict significantly and improve the performance of the data that we have worked on this.**

Keywords—Data Mining; Pre-processing; Heart Disease; Clustering; Machine Learning Techniques.

1. Introduction

Heart disease reflected as one of the most common diseases in previous centuries. It is a disease that affects all groups of people, adolescents, adults, and aging. This disease considered an incurable disease, as no good treatment reduces the severity of the disease, and there is always a failure in clinical heart cases. This study aided to prompt many researchers to find ways to block and compensate for the failures occurred mainly. The proposed model improves the performance of this disease by providing techniques and algorithms that work to reduce the risk of this disease and predict this disease and work to improve its performance. [1]. Avoid smoking, exercising, and controlling weight, as excess weight is considered a great danger and causes negative harm to the patient, as well as following a diet, and staying away from crises, nervous conditions, and stress, as this weakens the heart and makes the pulse fail quickly. The aortic valve, mitral valve, mitral valve, and tricuspid valve, where these valves open and close to direct blood flow from the heart, and this failure may cause damage to these valves for various reasons different, which leads to narrowing, leaking, or not closing naturally. It is one of the problems facing the disease. Also, this disease is dangerous when we ignore it, as it is the pulsating part of the human body [2].

In 2015 the authors suggested a method and systematically reviewed the data on fatal and non-fatal rheumatic heart diseases for the period between 1990 to 2015, where they obtained good results. Still, our work is superior to their work in terms of the methods used and the results obtained, where our work showed that the results are superior to its peers. As the authors explored, there were 319,400 deaths (95% uncertainty period, 297,300 to 337,300). Death due to rheumatic heart disease in one year confirms that their work is excellent, but our work is superior [3].

In 2016, musicians provided a technical model to present their systematic review and meta-analysis of longitudinal observational studies. The researchers studied this disease to provide a technique where they did a systematic review and meta-analysis to investigate loneliness and social isolation. They proved that their work is good, and it has been predicted to improve the performance of this disease. Our work outperforms this work in terms of results and methods used. We were able to improve the performance of heart disease, treat the clinical failure of heart disease and improve its performance [4].

Due to the widespread of this disease, we present a technical model to improve heart disease, reduce its severity, and predict it significantly. Here, the current authors proposed a model consisting of two experiments. The first one investigates the effect of clustering techniques in the pre-processing stage. The second one applied these techniques without the pre-processing. In the pre-processing step, the missing values and outlier problems governed. These techniques are used in this paper to give good results. During the application of these techniques, this data divided into five clusters, where each cluster includes a set of data, where the results were displayed, cluster 0 for 123 items, cluster 1 for 92 items, cluster 2 for 42 items, cluster 3 for 39 items, and cluster 4 for 3 items. It proves that our work is to take the path of righteousness to give good results with these divisions where all clusters. It includes many values, and the highest values considered to be higher than 83%. It confirms that our work is good enough and has given good results that remarkably

10.22133/ijwr.2022.304795.1105 20.1001.1.26454335.2021.4.2.3.1

predict this disease and give good results to fill the empty values in this article.

Here, a technical model based on clustering techniques proposed in which k-means, expectation-maximization clustering, DBSCAN, support vector clustering, and random clustering with/without pre-processing are used. These techniques applied with the Mail program, and good results obtained. We have proven that our work outperforms the previous work in terms of results and better prediction in improving heart disease by processing the data downloaded from the UCI website. Our model is well, and it predicts well the improvement of this disease.

This paper is organized as follows: Sect. 2, shows a summary of the related works. 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. RELATER WORK

Several works suggested predicting the heart Disease diagnosis in the years between 2015 to 2022. We summarized some of the significant conducts herein.

In 2015, the authors wrote about a randomized experiment. The two musicians introduced a technique to improve the performance of heart disease and better predict it through a randomized trial process. They enrolled eligible patients between the ages of 18-80 years with hard evidence, as these suffer from the ischemic disease in the tuning heart. They worked on these records to improve their performance, and his experience in the current paper confirms that our work is successful with good prediction [3].

Due to the spread of this disease and the lack of finding alternative ways to improve heart disease and limit these injuries, the authors 2016 motivated to work on heart data and present their method. It proposed to aggregate the results clinically into a temporal study to collect data. In the case of clinical research to improve heart disease and its performance, they presented this method and obtained the following results. They proved that the industry-sponsored a research program in the sixties and fifties, where their model and exploration became good because it improved the model performance [5]. In 2017, the authors estimated the associations of eating ten specific factors of nutrients with due mortality. They presented a model to reduce the risk of heart disease data and the corresponding uncertainty, as their experiments proved that their model is acceptable and may improve classification performance. The lack of treatment in this dataset is a big problem. It was why our study presented ways to improve the performance of this disease dramatically [6].

In 2019, the authors proposed a model to present a new model for finding the specific features through the application of machine learning (ML) techniques. They used this technology to improve heart disease and predict it significantly, where they used ML algorithms such as random forest (RF). It indicates that their work did well, with an accuracy of 88.7% [7]. The authors presented a technical model to estimate obesity in adolescent children in 2020. The most important step was related to adolescent weight [8].

In 2021, the authors presented a technical model to describe the risks of visceral heart in middle-aged and elderly patients and then evaluated the long-term results. To enhance the performance of this work, research and hard work are required to provide a technical method that gives good results, and this is what we have done in this research. Compared to previous work, our work has given excellent results and greatly predicts heart disease [9].

In 2020, Authors in [10] proposed a model to predict the performance of Heart Disease Diagnosis by six classifiers. The highest accuracy was 99.4% for Cleveland-Hungarian (CH) datasets. The experimental results proved that the combination of chi-square with PCA obtains greater performance in most classifiers. The utility of PCA directly from the raw data computed lower results and would require greater dimensionality to improve the results.

The authors proposed a novel framework based on IoT for heart disease diagnosis in 2020. The proposed framework is validated by examples of real case studies. The experimental results indicate that the proposed framework provides a viable solution that can work at a wide range, a new platform for millions of people getting benefits over the decreasing mortality and cost of clinical treatment related to heart failure.

In [11], a smart healthcare monitoring system for heart disease prediction is based on ensemble deep learning. In this system, the feature fusion method combines the extracted features from sensor data and electronic medical records to generate valuable healthcare data. Second, the information gain technique eliminates irrelevant and redundant features and selects the important ones, which decreases the computational burden and enhances the system performance. At last, the ensemble deep learning model was trained for heart disease prediction [12]. The proposed system is evaluated with heart disease data and compared with traditional classifiers. The proposed system obtains an accuracy of 98.5%, which is higher than existing systems

The current authors studied several works in related works. It found that the model presented a better predictive model where clustering algorithms include both k-means, expectation-maximization clustering, DBSCAN, support vector clustering, and random clustering with/without pre-processing with replacing the missing value with mean to detect outlier. Through the application of these techniques, we have reached well results for predicting heart disease using treating missing values and prediction significantly, which gave good results by dividing the data into five groups and gave good results and better prediction. It confirms and proves that our work outperforms all previous work. It has proven that the model we presented is an excellent model, highly predictive, and gives better results for improving heart disease.

3. THE HFC MODEL

Here, we conducted two experiments. The first experiment includes applications of pre-processing techniques by replacing the missing value with the mean to detect with outlier with cluster techniques including k-means, k-means fast, expectation-maximization clustering, DBSCAN, support vector clustering, and random clustering. In the pre-processing process, we addressed missing values and stray values. We applied clustering techniques to improve and diagnose chronic heart disease, reduce heart failure, improve its performance, and predict well. The current authors applied a hierarchy of based clustering methods to select the best clusters for the used dataset (See Fig. 1).

3-1. The first stage

Here, we downloaded data for clinical heart failure records, as these data contain missing values and stray values. We applied pre-treatment to them using techniques, including replacing the missing value with the mean and detecting outliers with the Rapid Miner tool. We obtained excellent results as we were able to replace the missing values. The success of our work was confirmed with new values.

3-2. The second stage

It is one of the higher education techniques. It used to collect and work on data where we can use the clustering algorithm to classify data points to specific aggregates. As these data are similar in characteristics and tasks, the clustering algorithm is considered one of the techniques used to collect data and improve its performance. In this paper, we downloaded data from the UCI website to improve its performance, expectation-maximization clustering, DBSCAN, support vector clustering, and random clustering with pre-processing. The cluster is one of the unsupervised ML methods. It is a method for processing common data and the task in need of processing, wherein our work good results obtained that outperformed the previous work. The achieved results showed this, cluster 0 for 123 items, cluster 1 for 92 items, cluster 2 for 42 items, cluster 3 for 39 items, and cluster 4 for 3 items. It confirms our work is a successful business, and we obtained cluster values higher than 83% for better forecasting (See Fig. 2).

k-means: It is one of the methods of unsupervised ML, where this technique divides the data into *k*. Let $N = \{x_1, \dots, x_n\}$ be the set of n objects to be clustered by a similarity criterion, where $x_i \in R$ for $i = 1, \dots, n$ and $d \geq 1$ is the number of dimensions. Additionally, let $k \geq 2$ be an integer and $K = \{1, \dots, k\}$. For a k-partition, $P = \{G(1), \dots, G(k)\}$ of N, let μ_j denote the centroid of cluster G(j), for j ∈ K, and let $M = \{\mu_1, \dots, \mu_n\}$ and $W = \{w_{11}, \dots, w_{ij}\}$ .Therefore, the clustering problem can be formulated as an optimization problem, which is described by (1):

$$P: \min imize\ z(W, M) = \sum_{i=1}^{n} \sum_{j=1}^{k} w_{ij} d(x_i, \mu_j)$$

$$subject\ to \sum_{j=1}^{k} w_{ij} = 1, for\ i = 1, \dots, n, \quad (1)$$

$$w_{ij} = 0\ or\ 1, for\ i = 1, \dots, n, and\ j = 1, \dots, k,$$

where $w_{ij} = 1$ implies object x_i belongs to cluster G(j) and $d(x_i, \mu_j)$ denotes the Euclidean distance between x_i and μ_j for i = 1,...., n and j = 1,...., k [13]. For example, giving each cluster equal 5, wherewith this division our dynasty will become five clusters, and the data is distributed into five groups for excellent prediction. The k algorithm starts with the k, and we deal with it as if it were an intermediate nodal wherein this work. We worked on data that contained missing values, and through this algorithm, a table of results and values was obtained and divided into five groups. It confirms that the work is very predictable with this algorithm for all the examples repeated several times.

k-means fast: An unsupervised ML, where is more than the aggregation method designed. It predicts the best neighbors by specifying the k sums, which predicts better and predicts well, where this algorithm is much faster than other algorithms. It proved its effectiveness through the results obtained, where we applied this algorithm with/without pre-processing to get the best results and predict faster, as the data divided into five groups. Cluster 0 equals 123 items, Cluster 1 equals 3 items, Cluster 2 equals 39 items, Cluster 3 equals 42 items, and Cluster 4 equals 92 items. The total number of items is equal to 299. It proves that the work of this algorithm is well, and it has outperformed its peers and previous work, as it will better predict and improve heart failure to improve its performance. It outperformed the previous work where the results showed this, cluster 0 for 123 items, cluster 1 for 92 items, cluster 2 for 42 items, cluster 3 for 39 items, and cluster 4 for 3 items. It confirms although our work is a successful business, we obtained cluster values higher than 83% for better forecasting (See Fig. 3).

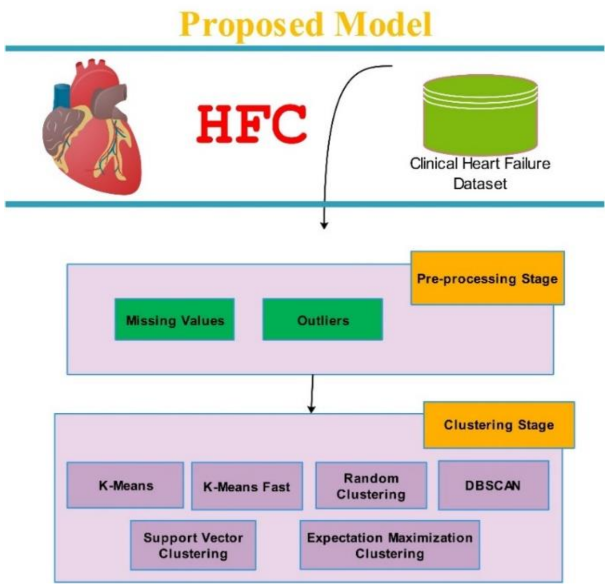

Fig. 1. The proposed model

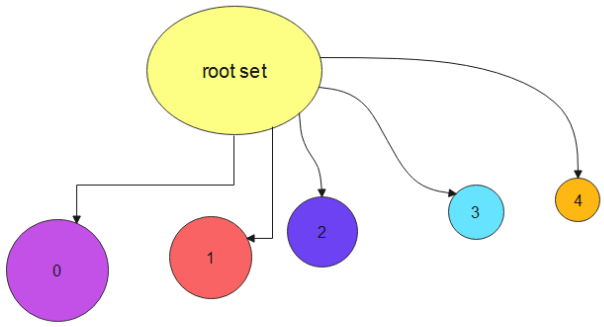

Fig. 2. The K-means

Expectation-maximization clustering: Due to the complexity of the data that we downloaded and worked on, we want to reach the highest values to improve classification performance. The expectation-maximization clustering algorithm was applied herein, which is one of the unsupervised algorithms. It is one of the methods of machine learning. It is a viable way to verify the values of these data where this data divided into groups, and each group represents the k. It is merit for iterative computation to manage the maximum probability, to implement the estimation within the full data framework. There is a need to process the data by applying pre-processing methods to improve these data and obtain better predictive values. This data is better divided into Cluster 0: 282 items, and Cluster 1:17 items. The total number of items equal to 299, and cluster probabilities: Cluster 0: 0.9434205719274595, Cluster 1: 0.0565794280725406, where this method has proven to be successful, germinates well, and outperforms the previous works (See Fig. 4).

DBSCAN: It is a type of unsupervised ML. It is considered one of the aggregation algorithms and applied in this study. It considered one of the most important algorithms for collecting data from different sources, processing it, and improving its performance due to the problems that exist in this data that we have worked on where we applied this algorithm with/without pre-processing, where this data is divided into groups. Each group represents the letter k, which is the prediction of the nearest neighbors. The data in this algorithm divided as follows: Cluster 0 equals 299 items, and the total number of items equals 29. The number of clusters is 1.000. Where this unsupervised algorithm is considered good through its results, it is highly predictable and excellent, as it has proven that its work is superior to the previous work in terms of accuracy and extracted results (See Fig. 5).

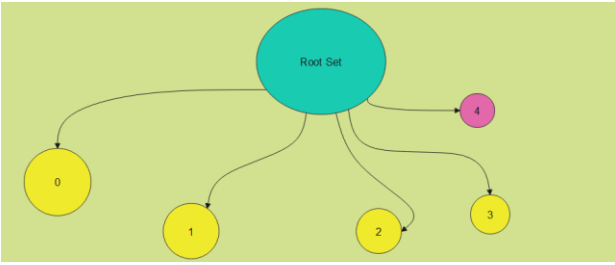

Fig. 3. The K-means fast

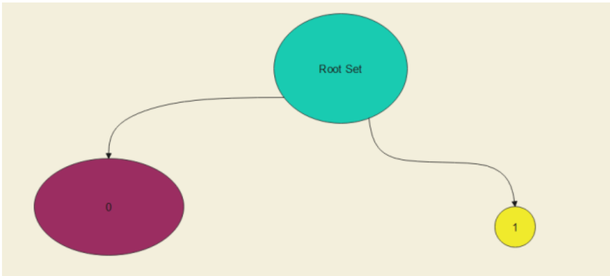

Fig. 4. The expectation-maximization clustering

Support vector clustering: Due to the seriousness and spread of heart disease and the occurrence of heart failure in many cases, we started this paper to use more than one algorithm. The goal of this technique is to divide the data into groups. According to the evaluation metrics, each group represented by the letter k, where this algorithm applied with/without pre-processing. It is the best way to improve this data and predict better to reach satisfactory results as it depends on the nearest neighbor. To divide this data, it divided as follows: Cluster 0 equals 299 items, the total number of items is 299, as this technology gave excellent results and better germination, and this indicates that our work outperforms all previous work, and this proves the success of our work (See Fig. 6).

Random clustering: This algorithm is considered one of the types of ML. It is the collection technique to give good results with the pre-processing of this data. This algorithm is concerned with aggregation together to apply better and give the best results. This technique implements the most prominent flat clusters where the data divided into the nearest neighbors were giving the value of the collection is with the letter k where the input port expected on Example Set. It is the output of the recovery factor in the attached example process. The output of other operators can also be used as input. The totals divided as follows: Cluster 0 equals 99 items, Cluster 1 equals 101 items, Cluster 2 equals 99 items, and the total number of items is 299. This technology, combined with the Rapid Miner tool and pre-processing, gave good results that outperform its peers, and this proved that our work with cluster algorithms, good algorithms that better predict the improvement and prediction of heart disease (See Fig. 7).

4. SIMULATION AND RESULTS

4-1. Data collection

Web resources, datasets, and web research used to explore data. In this context, we have downloaded the clinical heart

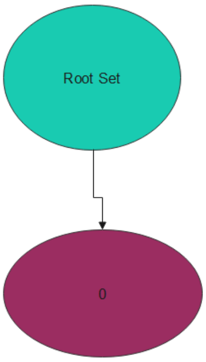

Fig. 5. The DBSCAN

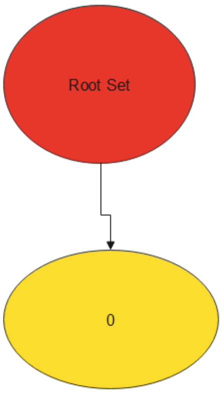

Fig. 6. The Support vector clustering

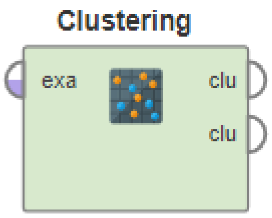

Fig. 7. The random clustering

failure data from the UCI website, as these data contain missing values and stray values [14] [15]. The current authors applied techniques to replace missing values, and detect outliers with pre-processing to improve the performance of this disease, predict significantly to improve, facilitate steps to solve the problem of failure in cases Clinical. Then clustering algorithms were applied, including k-means, expectation-maximization clustering, DBSCAN, support vector clustering, and random clustering, which gave excellent results for predicting chronic heart disease and improving its performance. It proves that our work has outperformed all previous work. Give high-accuracy results, and predict the fastest time, and these techniques considered the best among the rest of the techniques because they give effective and worthwhile results, whereas heart disease considered an incurable disease, which prompted researchers and scientists to search for the best ways to improve and predict this disease, which gave good results.

4-2. Experiment I

In the first experiment, we verified the implementation of clustering algorithms with k-means with pre-processing with the Rapid Miner tool.

Table I shows the process of clustering with preprocessing with k-means. Pre-processing applied with the aggregation technique, ML k-means, where the results showed effectiveness through Table I, which includes five groups, including Cluster 0, Cluster 1, Cluster 2, Cluster 3, and Cluster 4. The values in the Table I are the values that were obtained by dividing the data by the nearest neighbor, where one of the above values amounted to Death-event, 0.285, 0.293, 0.357, 0.462, and 0.333. The obtained results from the Table I proved successful. It proved to be good work and well predictive, and the performance of chronic heart disease improved.

4-3. Experiment II

In the second experiment, we verified the implementation of clustering algorithms with k-means fast, with pre-processing with the Rapid Miner tool.

In the second experiment, we pre-processed operations within the Rapid Miner tool with the k-means fast by replacing the missing value and detecting the outlier clustering algorithm. The data divided into five groups, each group represented by a cluster, where these totals are as follows Cluster 0, Cluster 1, Cluster 2, Cluster 3, and Cluster 4. Where these totals represent the process of dividing the data. and many values were obtained, including 0.285, 0.333, 0.462, 0.357, and 0.293. As these results are considered good and outperformed the previous works, it confirmed that this work outperformed all previous works, the prediction greatly improved, and the performance of heart disease improved. Its peers in terms of accuracy, methods used, and the results obtained indicate that this work has been achieved significantly to obtain excellent results due to the seriousness of this disease, which prompted us to search for alternative methods that work on finding a solution to the failure in some clinical cases.

4-4. Experiment III

In the third experiment, we will verify the application of the pre-processing with the expectation-maximization clustering algorithm, wherewith the Rapid Miner tool. We obtained a results table consisting of two groups of cluster 0 and cluster 1 (Table II), where we obtained highly predictive results that gave excellent results for improvement in heart disease chronic.

We obtained good values through Table III. We divided the data into two groups, each group containing certain values and predicting better than others. By these results, we proved that our study is a good enough outperforms its counterparts.

4-5. Experiment IV

In the fourth experiment, we verified the implementation of clustering algorithms with k-means fast, without pre-processing with the Rapid Miner tool.

According to Table IV, we implemented an application within the Rapid Miner tool, where we applied the k-means algorithm without pre-processing. It gave effective results that outperformed its peers. In the Table IV, the data divided into five groups as follow: Cluster 0, Cluster 1, Cluster 2, Cluster 3, and Cluster 4, where many values were obtained, including the times, were equal to 130.980, 183, 125.133, 133,203, and 125.800. The results of this table are considered excellent results that outperformed their peers and gave excellent results that work to improve the performance of this disease and gave good results for predicting chronic heart disease and treating clinical failure.

4-6. Discussions

Due to the importance of heart disease and its frightening spread in recent times, and the lack of good results that limit the severity of this disease, many researchers have prompted us to research and investigate to get good predictions and improve the data on the clinical failure of heart disease. We worked on a clinical heart dataset using two experiments. In the first experiment, we used the Rapid Miner tool with the pre-processing using cluster techniques, ML k-means, k-means fast, expectation-maximization Clustering, DBSCAN, support vector clustering, and random clustering with pre-processing. We obtained data division and obtained the best results, where among them we obtained 0.285, 0.293, 0.357, 0.462, and 0.333. The Table II with the k-means fast algorithm, where some values within this table amounted to 0.285, 0.333, 0.462, 0.357,0.293. We have also obtained Table III with pre-treatment. These values are among the most important values obtained, as we proved the superiority of our work over the previous studies. In the second experiment, the clustering algorithms applied with k-means and k-means fast, expectation-maximization clustering, DBSCAN, support vector clustering, and random clustering without pre-processing through the Rapid Miner tool, where we obtained good values. Dividing the data into five groups, each group containing certain values, where we obtained excellent values that sprout excellently. It considered the success of our work, as we obtained some values in the Table IV, 130, 980, 183, 125.133, 133, 203, and 125.800. It confirms to us the success of our work in terms of accuracy. In the previous work, we have predicted expertly. It confirms that we will improve heart disease and improve its performance. The results of our work were represented with several drawings, starting from Figs 7-19, where these illustrations are an indication that our work is good, as it exhibits that our work is superior to all previous works, and this makes our work an elaborate work and of high value, as our work has proven that our results are well and will be highly predictive and will address missing data and clinical failure in heart disease and this is a good business compared to its counterparts.

Figs 8-16 show the ROC for the best of the current clustering with pre-processing through the Rapid Miner tool. Figs 17-20 show the ROC for the best of the current clustering without pre-processing through the Rapid Miner tool.

5. Conclusion

Heart disease is one of the diseases that spread in the previous centuries. We propose a high-resolution model using clustering techniques. Many studies create the appropriate ways to test this disease. But they do not examine this disease and its action. In this paper, our work divided as follows, where we conducted two experiments with the rapid miner program, where we downloaded data on clinical heart failure, and these data contain missing values and outliers. For this, we used two experiments. The first experiment applied clustering techniques that include both k-means and k-means, expectation-maximization clustering, DBSCAN, support vector clustering, and random clustering with pre-processing with the Rapid Miner tool. Tables I to IV obtained, these table contains all totals, each group represents the nearest sample as it contains all values, and the data divided as follows: Cluster 0, Cluster 1, Cluster 2, Cluster 3, and Cluster 4, where this division is good to give better results and predict significantly, as we obtained some values in the first table, 0.285, 0.293, 0.357, 0.462, and 0.333. And then, we obtained in the second table with pre-processing as well. Good values, as we will mention some values from the second table: 0.285, 0.333, 0.462, 0.357, and 0.293. Our work and the most important results to predict faster and better. It indicates that the proposed model is excellent and gives excellent results. In the second experiment, we used clustering techniques that include k-means, k-means fast,

TABLE I. Centroid table for cluster model with pre-processing through k-means

Attribute	*Cluster 0*	*Cluster 1*	*Cluster 2*	*Cluster 3*	*Cluster 4*
Anemia	0.382	0.478	0.454	0.462	0.333
Creatinine-phospokin	641553	499.163	541.024	608.538	893.333
Diabetes	0.447	0.391	0.429	0.359	0.667
Ejection-fraction	38.146	38.370	39.524	35.385	41.667
High-blood-pressure	0.366	0.359	0.381	0.256	0.333
Platelets	282593.096	214423.913	407333.333	126592.308	737666.667
Serum-creatinine	1.306	1.357	1.482	1.674	1.267
Serum-sodium	136.179	136.739	137.405	136.718	139.333
Sex	0.626	0.674	0.595	0.744	0.333
Smoking	0.350	0.261	0.310	0.359	0.667
Time	125.789	133.043	134.929	131.154	151.333
Death-event	0.285	0.293	0.357	0.462	0.333

TABLE II. Centroid table for cluster model with pre-processing through k-means fast

Attribute	*Cluster 0*	*Cluster 1*	*Cluster 2*	*Cluster 3*	*Cluster 4*
Anemia	0.382	0.333	0.462	0.452	0.478
Creatinine-phospokin	641.553	893.163	608.538	541.024	499.163
Diabetes	0.447	0.667	0.359	0.429	0.391
Ejection-fraction	38.146	41.667	35.385	39.524	38.370
High-blood-pressure	0.366	0.333	0.256	0.381	0.359
Platelets	282593.096	737666.667	126592.308	407333.333	214423.913
Serum-creatinine	1.306	1.267	1.674	1.482	1.357
Serum-sodium	136.179	139.333	136.718	137.405	136.739
Sex	0.626	0.333	0.744	0.595	0.674
Smoking	0.350	0.667	0.359	0.310	0.261
Time	125.789	151.333	131.154	134.929	133.043
Death-event	0.285	0.333	0.462	0.357	0.293

TABLE III. Centroid table for cluster model with pre-processing through expectation maximization clustering

Cluster No.	*Centroid*
Cluster 0	0.43613249064927917 591.0554870400457 0.40077376881549653 37.78069798216367 0.3616935542300597 251239.96613222425 1.3908638905213304 136.76862581931874 0.6630385351194862 0.3227040795097773 129.07902494828505 0.3226077422664821
Cluster 1	0.3531630489476152 428.1690374759626 0.7062988057818785 43.134482357320955 0.17570748072830433 465417.8669136395 1.4441643257433032 134.23753344181281 0.41190302234722304 0.2938270573884112 149.96726396365793 0.29543341044399474

TABLE IV. Centroid table for cluster model without pre-processing through k-means fast

Attribute	*Cluster 0*	*Cluster1*	*Cluster2*	*Cluster 3*	*Cluster 4*
Anemia	0.405	0	0.467	0.475	0.457
Creatinine-phospokin	618.915	1176	313.267	469.136	636.371
Diabetes	0.412	1	0.467	0.356	0.457
Ejection-fraction	37.908	47.500	38.267	36.966	39.100
High-blood-pressure	0.359	0	0.400	0.305	0.371
Platelets	2046006.214	796000	481200	145866.102	33814.286
Serum-creatinine	1.378	1.050	1.369	1.479	1.373
Serum-sodium	136.118	140	136.400	137.136	137.257
Sex	0.647	0.500	0.467	0.780	0.586
Smoking	0.294	0.500	0.333	0.373	0.329
Time	130.980	183	125.133	133.203	125.800
Death-event	0.301	0	0.400	0.373	0.314

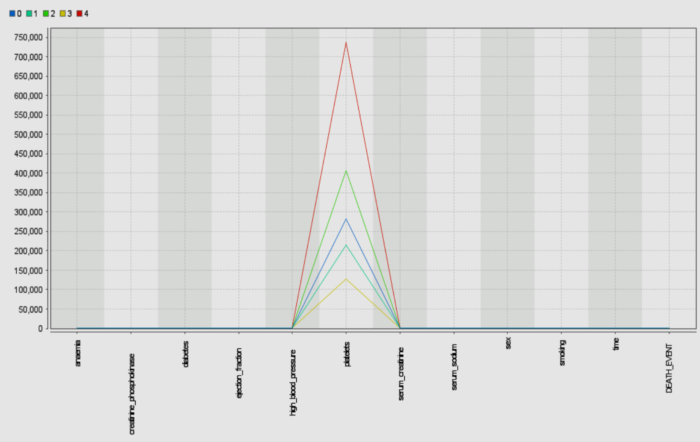

Fig. 8. The ROC for cluster model (clustering) with k-means

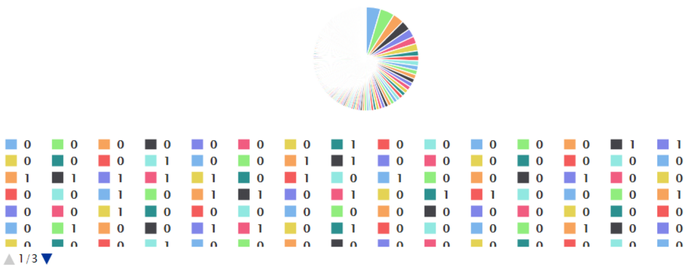

Fig. 9. The ROC for ple/dount with k-means

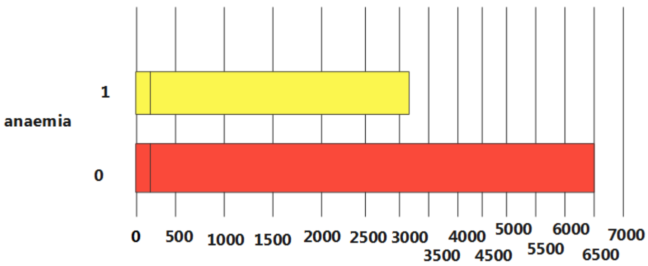

Fig. 10. The ROC for Bar(horizontal)with k-means

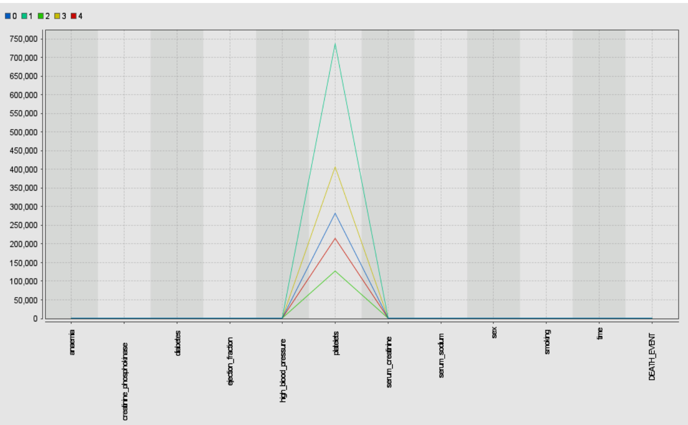

Fig. 11. The ROC for cluster model (clustering) with k-means fast

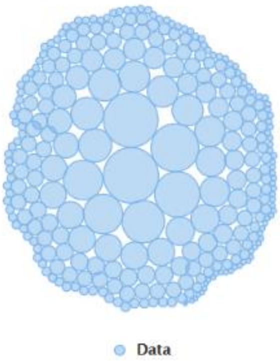

Fig. 12. The ROC for packed bubble with k-means fast

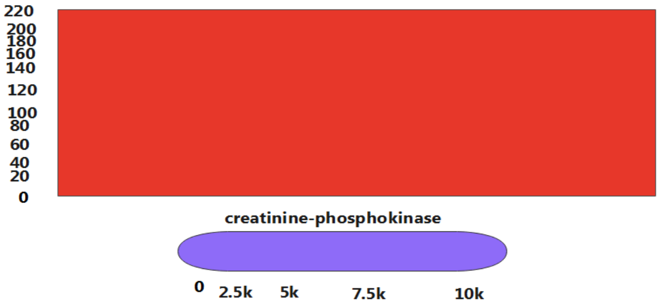

Fig. 13. The ROC for heatmap with DBSCAN

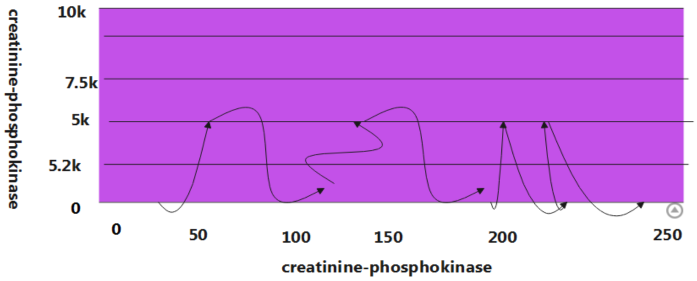

Fig. 14. The ROC for with step line with expectation maximization clustering

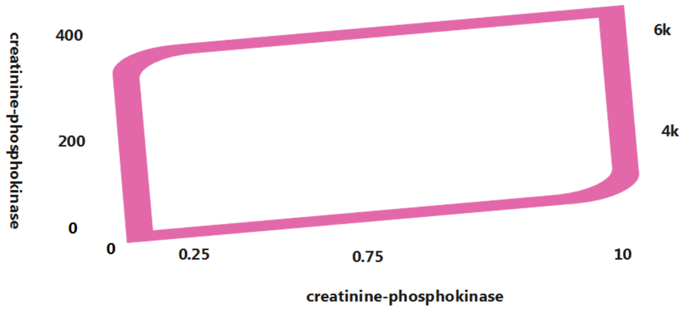

Fig. 15. The ROC for scatter 3d with support vectore clustering

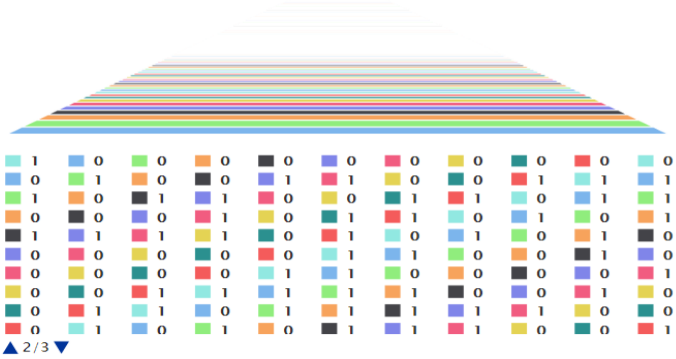

Fig. 16. The ROC for pyramid with random clustering

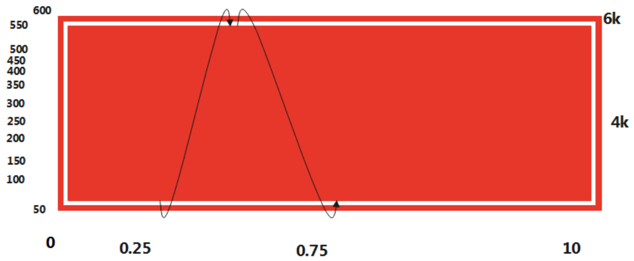

Fig. 17. The ROC for cluster model (clustering) with k-means fast

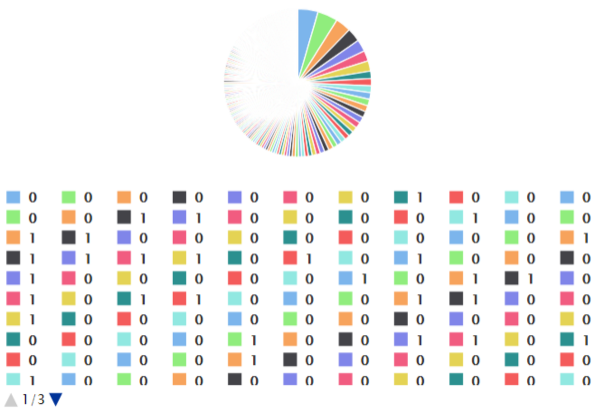

Fig. 18. The ROC for pie/dount with k-means

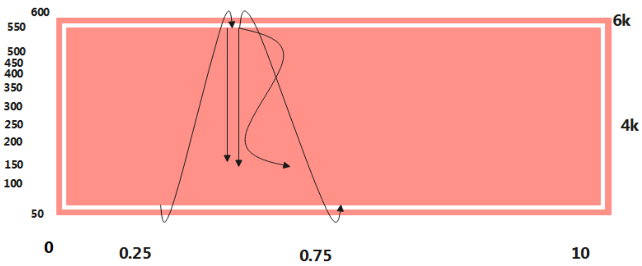

Fig. 19. The ROC for cluster model (clustering) with k-means fast

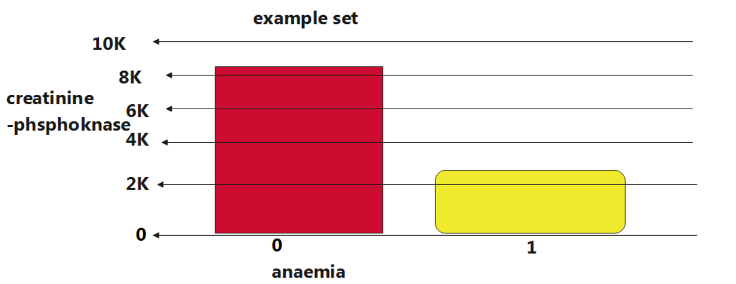

Fig. 20. The ROC for Bar colum with k-means

expectation-maximization clustering, DBSCAN, support vector clustering, and random clustering without pre-processing with the Rapid Miner tool. We obtained new values and high values that are highly predictive and worked on them. Improving heart disease, as the fourth table has proven excellent results that our study is superior to the previous works. In the fourth table, the data we downloaded and worked on were divided into five groups as follows: Cluster 0, Cluster 1, Cluster 2, Cluster 3, and Cluster 4, where these data revealed Good results and high improvement with these techniques and without pre-treatment, as we obtained good and high values, which are values that outperform their peers, as they contain some values. For example, times were 130.980, 183 125.133, 133.203, and 125.800. We proved in the second experiment that our work is also good and outperformed the previous works, and this indicates that the technical model that we proposed in this paper worked well and predicted excellently for improving heart disease, treating clinical failure, and improving the performance of heart disease, and our work has proven its superiority over its counterpart.

In our future works, we will develop more and apply other techniques to get the best results. We will use techniques such as associated, summing, and recommender system, and many of the works that we think to present works where we seek to reach the best results, such as our study. This makes us provide much research in the future and improve the performance of some diseases and some important data that must be worked on and developed and improve their performance.

ACKNOWLEDGMENT

The authors thank the editor and all anonymous reviewers.

REFERENCES

[1] R. Asgarnezhad and S.A. Monadjemi, "Persian sentiment analysis: feature engineering, datasets, and challenges," *Journal of applied intelligent systems & information sciences*, Vol. 2, No. 2, pp. 1-21, 2021.

[2] R. Asgarnezhad and S.A. Monadjemi, "NB VS. SVM: A contrastive study for sentiment classification on two text domains," *Journal of applied intelligent systems & information sciences*, Vol. 2, No. 1, pp. 1-12, 2021.

[3] 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, pp. 43-54, 2015.

[4] 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, pp. 1009-1016, 2016.

[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, pp. 1680-1685, 2016.

[6] R. Micha, J. L. Peñalvo, F. Cudhea, F. Imamura, C. D. Rehm, and D. Mozaffarian, "Association between dietary factors and mortality from heart disease, stroke, and type 2 diabetes in the United States," *Jama*, vol. 317, pp. 912-924, 2017.

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

[8] 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, pp. 2371-2379, 2007.

[9] 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, pp. 2577-2586, 2021.

[10] K. Gárate-Escamila, A. H. El Hassani, and E. Andrès, "Classification models for heart disease prediction using feature selection and PCA," I*nformatics in Medicine Unlocked*, vol. 19, p. 100330, 2020.

[11] M. Abdel-Basset, A. Gamal, G. Manogaran, L. H. Son, and H. V. Long, "A novel group decision making model based on neutrosophic sets for heart disease diagnosis," *Multimedia Tools and Applications*, vol. 79, pp. 9977-10002, 2020.

[12] F. Ali, S. El-Sappagh, S. R. Islam, D. Kwak, A. Ali, M. Imran, and K. S. Kwak, "A smart healthcare monitoring system for heart disease prediction based on ensemble deep learning and feature fusion," *Information Fusion*, vol. 63, pp. 208-222, 2020.

[13] J. Pérez-Ortega, N. N. Almanza-Ortega, A. Vega-Villalobos, R. Pazos-Rangel, C. Zavala-Díaz, and A. Martínez-Rebollar, "The k-means algorithm evolution," in *Introduction to Data Science and Machine Learning*, ed: IntechOpen, 2019.

[14] A. Asuncion and D. Newman, "UCI machine learning repository," ed: Irvine, CA, USA, 2007.

S. D. Bay, D. Kibler, M. J. Pazzani, and P. Smyth, "The UCI KDD archive of large data sets for data mining research and experimentation," *ACM SIGKDD explorations newsletter*, vol. 2, pp. 81-85, 2000.

Razieh Asgarnezhad received her B.Sc. and MSc. degrees in Computer Engineering from Kashan Azad University in 2009 and Arak Azad University in 2012, respectively. She received Ph.D. degree from Isfahan Azad University in 2020. Her current researches include Data Mining, Text Mining, Learning Automata, Recommendation System, Sentiment Analysis, and Wireless Sensor Network.

Karrar Ali Mohsin Alhameedawi received his B.Sc. degrees in Computer Engineering from the Al-Rafidain University of Baghdad in 2018. He received a certified international development trainer specialized in human development from Germany in 2019. After one year, he received the title of the best humanitarian personality in Iraq in 2020. He has experience as a website developer, programmer, the highest educational social networking site in Iraq, and an electronic contest program. He is a member of seven international institutions and has received 200 certificates and a letter of thanks from European and local countries. He is currently an MSc. student at Azad Isfahan University. His current researches include Data Mining and Big Data.