# MTS: Improving Thoracic Surgery Based on Clustering and Classification Techniques

***Razieh Asgarnezhad[1], Karrar Ali Mohsin Alhameedawi[2]***

*[1]Department of Computer Engineering, Aghigh Institute of Higher Education Shahinshahr, Isfahan, Iran.*

*[2]Student, Department of Computer Engineering, Isfahan (Khorasgan) Branch, Islamic Azad University, Isfahan, Iran.*

*Corresponding Author: karrark658@gmail.com*

**Abstract: -** Thoracic and cardiothoracic surgery is one of the most dangerous diseases that may face a big problem. When performing such operations, cardiothoracic surgery is a medical field that specializes in surgery for diseases that affect the rib cage and heart. The major challenge is to address outliers and missing values problems for predicting thoracic disease and improving its performance. In this paper, we proposed an effective technical model to improve the performance of thoracic surgery. The current authors applied two experiments. The first experiment includes the application of the clustering task as one of the unsupervised machine learning methods. The supervised techniques were applied for classification tasks in the second experiment. Using these two experiments, we got good results, and it improved the performance of chest surgeries, where the highest accuracy and F1 reached 86.17% and 92.56%, respectively. Therefore, our model is considered an effective proposal and outperforms its peers.

**Key Words: —** ***Pre-processing, Thoracic surgery disease, Ensemble technique, Clustering, Classification.***

## I. INTRODUCTION

Thoracic surgery is one of the most dangerous diseases that affect the heart, the thoracic cage, and all age groups. Sometimes, it is more serious for the elderly as the biggest problem. In this disease, the missing values are considered a problem that should be treated and organized. Cardiac surgery includes the heart, blood vessels, thoracic surgery, and the lungs prompted many researchers to search for techniques and algorithms to improve this disease. The current authors proposed a model that predicts heart surgery disease and works to improve it. Where heart surgery is a part of thoracic surgery, it is the surgery performed by surgeons at a high level on the heart muscle, blood vessels, or the pericardial membrane close to the heart. Heart surgery is divided into two parts: open-heart surgery and closed heart surgery.

Manuscript revised May 28, 2022; accepted May 29, 2022. Date of publication May 30, 2022.

This paper available online at www.ijprse.com

ISSN (Online): 2582-7898; SJIF: 5.59

It is a dangerous area; the injured patient has to follow a focused diet and keep away from all the things that make him worse. For example, staying away from crises, shocks, fatigue, assumptions about jogging, and others that pose a threat to the patient's life.

In 2014, the authors contacted the patients who had performed thoracic surgery. After this operation, they interviewed via CPSP phone at their hospital where they examined them to confirm their conditions for Maguire with a group of the questionnaires were. Then, it was sent to the patients, who also developed a logistic analysis for them. To determine the SF equal to 36, a health and ID survey determined that the CPSP had not improved and found a CPSP of 32.5% (86/265 patients). It made many researchers, including us, search for excellent ways to make better predictions. After thoracic surgery, it was found that their bodily functions had significantly decreased, and their quality of life became worse. After thoracic surgery has decreased their physical function and quality of life worse [1]. In 2013, the authors proposed an effective model to select four relevant studies with patients matching the slope score of the six Tronic rules, including endpoints, number of patients surrounding thoracic surgery, and knowledge of individual complications after thoracic surgery. And how long does it take to recover from this

surgery? They suggested (VATS) which is a goal of a meta-analysis compared to the results. They obtained from video-assisted thoracic surgery (NSCLC), where they found that all VATS patients who died surrounding chest surgery were similar among those with general complications. Much less or multiples of much lower rates of air pollution [2].

Because of the seriousness and facing some problems of chest surgery disease, the current authors have been prompted to present an effective technical model that works to predict and improve the performance of chest surgery disease. We downloaded thoracic surgery data from the UCI educational site. But these values have problems containing stray values and missing values, as they were worked on by conducting two experiments in our work. In the first experiment, we applied unsupervised machine learning (ML) techniques for clustering techniques such as DBSCAN, k-means, k-means (fast), k-means (H2O), k-medoids, random clustering, support vector Clustering, and X-mean algorithms. We obtained good results and divided the data to the nearest neighbor, which represents the letter k. Cluster 0 where some values reached 54, 720, 42.764, 73.436, and 57.714 without pre-processing. Then, we carried out the preprocessing and got two tables, Cluster 0, Cluster 1, and Cluster 2, where these are the most important values 70.936, 57.714, and 56.838. In the second experiment, we obtained good values with an excellent prediction through other techniques, which are supervised techniques within the techniques of ML. We worked on the downloaded dataset by applying ensemble techniques such as bagging, stacking, voting, and booting with/without pre-processing. The dataset was processed on missing values to improve the performance of this disease. Without pre-processing, we obtained the following measurements with values of precision, recall, accuracy, and f1. The highest value reached 86.03, 100, 85.11, and 91.96%. These values are considered good in comparison to the previous works, and they proved to be superior to the previous works. The highest value reached 86.17, 100, 85.11, and 92.56% through boosting, voting, and bagging techniques. Our model has proven that it is superior to its peers in terms of evaluation metrics. With clustering and classification techniques, our work has proven to be good enough and has improved the performance and prediction of this disease.

In this paper, an effective technical model was proposed with unsupervised machine learning techniques, clustering techniques. In the first experiment, we applied DBSCAN, k-means, k-means(fast), k-means (H2O), k-medoids algorithm, random clustering, Support Vector Clustering, x-mean, and we got good results that predict well. In the second experiment, we applied supervised machine learning techniques, where we applied ensemble with bagging and stacking and voting and boosting with preprocessing and without preprocessing techniques.

This paper is organized as follows: We show a summary of the related works in Sect. 2. Then, the proposed method and experiments are presented in Sect. 3 and Sect. 4, respectively. Conclusion presented in Sect. 5.

## II. RELATED WORKS

Several works suggested to predict the in thoracic surgery years between 2015 to 2021. We summarized some of significant works herein.

In 2015, the authors introduced a method by reviewing the medical records of chest and heart surgery between (2008-2013) for clinical signs, personal imaging, surgical findings, and other pathological conditions. They collected results in 12 dogs who had VATS for automatic chest therapy. Their method is somewhat satisfactory. Thoracic surgery is one of the most dangerous diseases that a person may face. It prompted researchers and other scientists to research and find the best ways to predict chest surgeries, and among them, we presented in this paper a technical model that relies on automated learning for supervision and unsupervised ML [3]. Also, authors in 2015 presented a model to predict chest surgeries, where they used classification methods and classification algorithms to predict chest surgeries and improve their performance. Classification of thoracic surgery and this is what makes their work satisfactory, as the highest accuracy in their article reached 84% in comparison with our work, our work outperforms this work in terms of accuracy, results used, and methods for improving classification performance [4].

In 2016, the authors presented a model that predicts cardiothoracic surgery in which they used heart rust data obtained from 77 to develop an automated system. They developed the HCM ML model with 62 patients from HCM to develop an automated system. They applied three algorithms including random forests and artificial neural networks for excellent prediction and satisfactory results, and then they used the K-fold voting method. To make critical predictions for the improvement of heart disease and thoracic surgery requires the search for excellent ways to function perfectly [5]. In 2017, the authors proposed an effective model to improve thoracic surgery and lung cancer, as they worked on data containing 470 records and 17 features, where they extracted the most important results and the most important covariates that cause

chronic lung cancer using algorithms to detect knowledge, And the data such as the Naive Bayes algorithm, where they used prediction by applying the another algorithm, where they reached the development of the prediction of the risk of death after one year, and their method presented is good and satisfactory, as the authors also designed a calculator to determine and detect the risk of death after surgery one year based on a scorecard algorithm measurement [6].

In 2018, the authors proposed a method and model for forecasting in the thoracic surgery industry. They investigated the distribution of non-research, non-fundamentalist, and non-proprietary payments made to thoracic surgeons, as their goal was to explore the regularity of relationships. Finance among many industries and thoracic surgeons is at a high level. Their goal is to predict and improve the performance of thoracic surgery. They obtained annual statistical data from the paid data set for the payments program from 2014-2016. They were able to distinguish the distribution of annual payments with individual payments that exceed more than 1 0000 US Dollars. It confirms that their work is good, but our work has surpassed all the previous mentioned works here [7]. The authors in 2019 presented the continuation of technology due to the importance of thoracic surgery. The authors proposed a technical model whose goal is to conduct comprehensive trials of thoracic surgery and develop and improve its performance to be enhanced and provided by robots to work to achieve better results and improve classification performance, where they obtained a percentage of up to 44.7%. This study compared to our model and our proposal outperforms all previous works in terms of the results we have reached, and this makes us foresee excellently in the future [8].

In 2020, a model was suggested to improve the performance of thoracic surgery. The authors proposed a model that reviewed commonly used administrative databases and reviewed cancer registries in Health Services Research. Thoracic surgery and achieved satisfactory values due to the importance of thoracic surgeries, prompting many to conduct experiments and apply them to improve thoracic surgeries [9]. In 2021, the authors proposed an effective technical model given the danger of thoracic surgery and the difficulty of diagnosing this disease. They prompted their search for the best methods. They retrospectively analyzed diagnosed patients undergoing lobectomy and that lobectomy, segment, and pneumonectomy (January (January 2006 - December 2018). They used fifty characteristics before surgery to predict cardiopulmonary complications. Through their proposed model, they obtained the highest value, which reached 70%, and this confirms that our work has far outperformed this work, as we obtained an accuracy of 85.11%. It proves The success of our work and its superiority over its peers [10].

The current authors found that the highest result was reached by the rapid miner tool. The highest accuracy was 85.11% through the ensembles. In this paper, we conducted eight experiments with machine learning techniques unsupervised and supervised machine learning techniques with clustering and ensemble techniques. It has been shown that our work outperforms previous work, through the results obtained.

## III. THE PROPOSED MODEL

In this section, we introduced the proposed model (See Fig. 1).

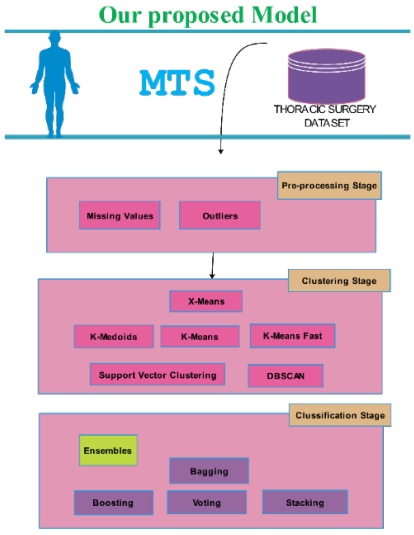


Fig.1. The proposed MTS model

We conducted two experiments on the data that we downloaded from the UCI website. This data needs to be processed. We processed it with the application of unsupervised ML techniques clustering with DBSCAN, k-means, k-means (fast), k- means (H2O), k-medoids algorithm, random clustering, Support Vector Clustering, x-mean with/without pre-processing. Then, we conducted a second experiment which is the application of supervised ML techniques and ensemble techniques with bagging, boosting, voting, and stacking with/without pre-processing, and our work turned out to be a good job, giving good results and better forecasting.

### *3.1 Pre-processing stage*

Here, we worked on the dataset from the site UCI, which is thoracic surgery data that contain missing values. We

performed the pre-treatment, where the first experiment was applied by pre-treatment with clustering techniques. We have applied the techniques of replacing the missing value with mean and detect outlier and obtained good results with the first experiment. Then, we have applied these techniques, but with ML techniques Supervised ensemble with replacing the missing value with mean and detect outlier. Pre-processing techniques are one of the most prominent techniques for treating outliers, excess values, and missing values [11-15]. It has made our work correctly and predicts excellently. Our proposal is good with these techniques and more predictive of this disease and works to improve it.

### *3.2 The clustering stage*

Here, unsupervised ML techniques were applied for processing and improving data performance. After addressing the problems in pre-processing, we applied the clustering such as BSCAN, k-means, k-means (fast), k-means (H2O), k-medoids algorithm, random clustering, Support Vector Clustering, x-mean. These aggregates are called a cluster, and they are divided according to the nearest neighbors, where the closest represents the letter k. Without pre-treatment, we got six tables, and each table represents a technique, and each table differs from the others. It confirms that our work gave distinctive and new results and predicts the disease of chest surgeries, and we will explain the techniques applied below in detail.

k-means: It is one of the most prominent techniques used by non-supervised ML techniques. Its work was divided as follows, where we considered that the k represents data aggregates. The data is divided and predicted according to the nearest neighbor and obtained two tables, a table with pre-processing, and contains good results and values, the most important of which are 54, 720, 42.764, 73.436, 57.714, and 62.739. We have also obtained another table without pre-treatment. We will mention some of them, 0.891, 0.941, 0.849, and 0.642. Thus, this process and algorithm is good for predicting better than others and predicting data improvement and development (See Fig. 2).

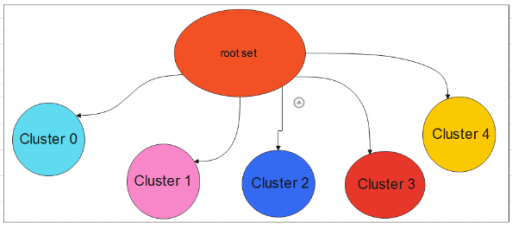


Fig.2. The used k-means model

Fast k-means: One of the algorithms and unsupervised ML techniques is considered one of the fastest algorithms in predicting the nearest neighbors by defining the data and dividing it by sum by giving a value for each group, for example, k = 5. It means we will have five groups: Cluster 4, Cluster 3, Cluster 2, Cluster 1, and Cluster 0. And these totals, each group contains more than 30 records, and each record contains values. For example, 54, 720, 42.764, 73.436, 57.714, and 62.739. This method is effective for evaluating and improving the performance of the disease through satisfactory results and the method of division and distribution that is applied to the nearest neighbors (See Fig. 3).

K-medoids algorithm: We searched for the best ways to improve classification performance and predict the best results. In this method, the data was divided into one-third of the totals, k =3. By dividing the data to the nearest third of the neighbors predicted well, where good and improved values were obtained in the tables below, and this indicates that our proposal is effective and works well, the process of dividing the data into totals in the figure (See Fig. 4).

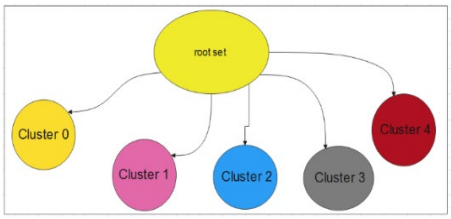


Fig.3. The used fast k-means model

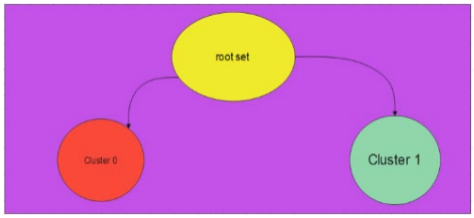


Fig.4. The used k-medoids model

DBSCAN: It is a type of clustering technique, and it is considered one of the unsupervised ML methods, and it is an improved method. After we downloaded the chest surgery data, we entered it into the Rapid Miner tool to work on it, where we applied the pre-processing methods to it. Then, the data were divided into five groups as shown in the below diagram, k = 5. We obtained values through this method, and this indicates that our work is well. This method, this method will predict and improve the performance of chest surgeries. (See Fig. 5).

Support vector clustering: Due to the spread of thoracic surgery, many researchers have been prompted to find the best results for forecasting. We have worked on unsupervised ML techniques in this article. This algorithm applied and gave a value for k = 12. It indicates that our work with this algorithm is a good job, and it was worked on and applied after the pre-treatment and with the pre-treatment, and we obtained satisfactory results that predict the diseases of chest surgeries. It processes the data we have downloaded and improves classification performance (See Fig. 6).

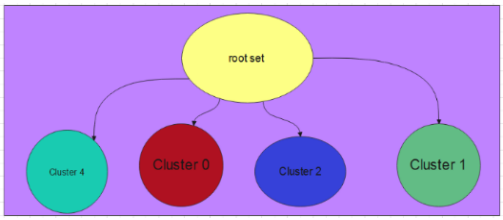


Fig.5. The DBSCAN model

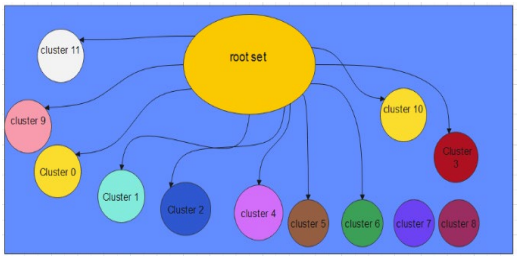


Fig.6. The used support vector clustering model

Random clustering: This method is considered good and gave satisfactory results as it is a type of clustering within the unsupervised ML methods. Our work was done with the Rapid Miner tool, and this method was applied to the data of chest surgeries to improve their level and predict the disease of chest surgeries due to the lack of excellent methods to reduce the risk. This disease is advised that the patient should be linked to a good diet and to stay away from exhausting things such as shocks and crises, k = 3. It means that the prediction of the nearest neighbors was divided into three groups, and this proves that the work of this algorithm is good and predicts well and gives satisfactory results, and outperforms its counterparts (See Fig. 7).

k-means (H2O): Here, the data was divided into three groups, each group consisting of a group of elements. It divided through the nearest neighbors, where the groups were divided as follows, k=2. We have two groups to predict the best results, as the results and this method are considered good and predict better. Using pre-processing on these data with this algorithm worked better, and this method gave satisfactory results. It confirms the success of the proposed model and confirms its superiority over its peers (See Fig. 8).

X-mean: Another method is one of the methods of unsupervised ML works and predicts well, the x-mean. We have proposed a technical model and applied high-accuracy techniques with k = 3. This data is divided into one-third of the totals and according to the nearest neighbors. This method is considered good to predict excellently and prove our work is good enough in the performance of classification (See Fig. 9).

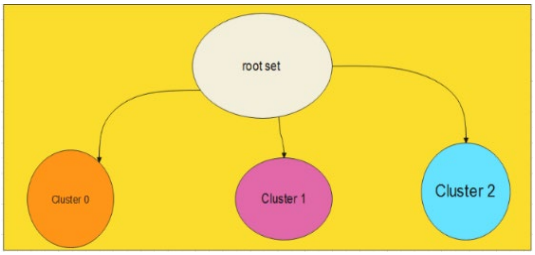


Fig.7. The used random clustering model

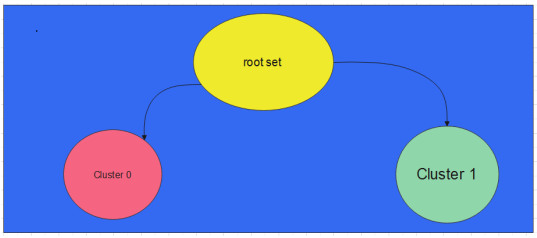


Fig.8. The used k-means (H2O) model

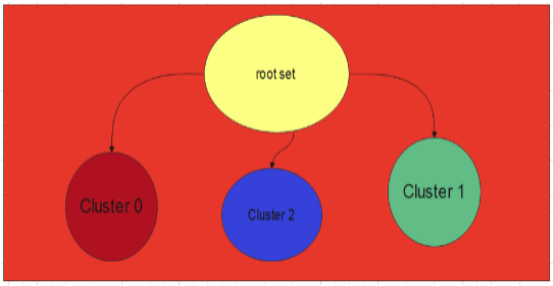


Fig.9. The used x-means model

### *3.3 The Classification Stage*

The ensemble method is considered one of the best methods used, as it always gives good results and predicts excellently [16-20]. It is considered one of the supervised machine learning methods. In this article, data for chest surgeries has been downloaded, but it needs treatment. We have applied two

experiments to it. The first experience is unsupervised machine learning techniques. With the supervision of clustering, good results were obtained, then in this section, ensemble techniques were applied with/without pre-treatment. well and improve the performance of classification and the performance of chest surgery disease (See Fig. 10).

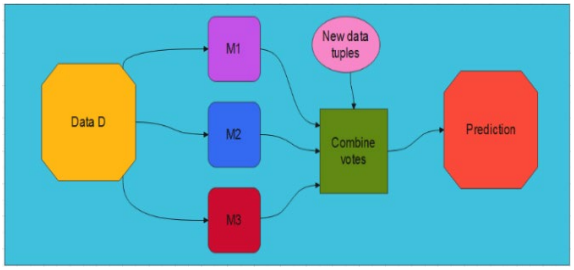


Fig.10. The used ensemble method generates a set of classification models, M1, M2, and M3 to given a new data tuple and combine the votes to return a class predication

Bagging: The bagging algorithm was applied with/without pre-processing. In the paper, the obtained precision, recall, accuracy, and F1 with bagging algorithm were 86.17, 100, 86.17, and 92.57%. These values are good enough and will improve the performance of the product and predict the disease of chest surgery.

Boosting: This method is considered one of the supervised ML methods as a type of clustering. It was applied to this data in the Rapid Miner tool, with/without pre-treatment, where its work is divided according to determining values. For example, in the education and supervision department and the training department, it has been implemented and we have obtained satisfactory results. The obtained precision, recall, accuracy, and F1 were 86.17, 100, 86.17, and 92.57%. These results are considered good, and this method has proven that it is on the right path and gave good values that predict excellently and improve chest surgeries.

Stacking: The stacking method is an aggregate method of ML subject to supervision, as it was applied in the Rapid Miner tool with/without pre-processing. The work is divided into the records education department and the training and data receiving department, where algorithms such as method tree and random forest. It is considered good because it gave good results. The obtained precision, recall, accuracy, and F1 with the stacking algorithm equal 86.17, 100, 86.17, and 92.57%. It indicates that this method is good and will make an excellent prediction with these results.

Voting: The last method is one of the methods of supervised ML, where this method is good for better prediction. It was worked with these data, and the techniques of ML supervised voting was applied, where its work is divided into two parts, a test section and a learning section, which were applied with/without pre-treatment, where we obtained two tables. The precision, recall, accuracy, and F1 were 86.17, 100, 86.17, and 92.57%. These values are good and will predict excellently, compared to previous work, that our work is a good job and outperforms its peers.

## IV. Results and Discussion

### *4.1 Data collection*

In this section, the thoracic surgery dataset was selected from among several datasets that we downloaded from the UCI website contains 470 instances and 17 attributes. It contains stray and missing values. We have worked on it as we applied pre-processing methods to replace the missing value with the mean and detect outliers. It determines the features of thoracic surgery data. We will develop and improve its performance in this paper and excellently predict this data. We will apply two experiments, the first with unsupervised machine learning methods, and the second with teaching methods. The supervised ensemble algorithm, where our proposal has been shown to work well and predict the best results.

### *4.2 Evaluation metrics*

To evaluation, accuracy, precision, recall, and F1 measures were applied. These measures were defined in Table 1.

Table.1. Evaluation Metrics

| Parameter | Equation |
|---|---|
| Accuracy | $(TP + TN) / (P + N)$ |
| Precision (P) | $(TP) / (TP + FP)$ |
| Recall (R) | $TP / P$ |
| F1 | $\frac{2 \times P \times R}{P+R}$ |

### *4.3 Experiments through clustering stage*

Experiment I: In the first experiment, we checked the data we downloaded from the UCI websites. The k-means algorithm was applied. This method is considered one of the unsupervised ML methods, and it is a good way to predict the disease of chest

surgeries. Its results have proven a good work and will improve classification performance.

We applied the clustering algorithms to obtain good results through k-means without pre-processing algorithms. We divided the data into five groups to predict the nearest neighbors. We got the best results, where these values in Table 2 represent the values of five groups of the data that we downloaded. The value of one of the features reached 0.922, 0.823, 0.738, and 1 0.8260. These results were obtained through the first experiment with unsupervised methods of teaching k-means, and this method showed well results in the first experiment. It confirms that we have done a good job in predicting thoracic surgery.

Experiment II: In the second experiment, we verified the implementation of clustering algorithms with k-means fast, with pre-processing with Rapid Miner tool. Table 3 shows us the results obtained results from applying the k-means fast algorithm without pre-processing. The data were divided into five groups, and each group represents a cluster, where we gave a value of k equal to five values. It means that we have five groups and these values in the table prove that our work is well and is distinguished from the rest of our study in terms of the accuracy of the results we obtained, and thus our proposal will predict well and improve the performance of these data.

Experiment III: In the third experiment, we verified the implementation of the k-medoids algorithm, where it was applied to the data that we downloaded, and we got a table containing two groups of elements predicted by us through the nearest neighbors.

According to the Table 4, we have applied unsupervised ML techniques to predict the best results and improve the performance of chest surgeries. Chest surgeries are an incurable disease because the chest area is close to the heart and rib cage. In the table, we applied the k-medoids algorithm. The data were divided into two groups, Cluster 0 and Cluster 1. The results showed that the application of clustering algorithms, especially this algorithm, gives good results that improve classification performance and predict the best results for improving chest surgeries.

Experiment IV: In the fourth experiment, we verified the implementation of the k-means algorithm with pre-processing and applied the techniques of replacing the missing value with mean and detect outliers. Through these steps, the missing values and stray values were addressed and predicted better.

According to the Table 5, we pre-processed the data using replace the missing value with the mean and detect outliers. With the application of the unsupervised ML algorithm with k-means, the data were divided into five groups to predict the best results. Through the closest neighbors, we reported good results. The performance of the data improved. The table showed the values of five groups. This method with pre-treatment proved a good work in predicting the best results and improving the performance of chest surgeries.

Experiment V: In the fifth experiment, we applied the k-medoids algorithm and verified the use of pre-processing techniques. With replace the missing value with mean and detect outliers, we have obtained good values that predict excellently.

According to the Table 6, it revealed that the results in the table are satisfactory. We used the k-medoids algorithm, which is one of the unsupervised ML methods, given the importance of this disease and the importance of this data. Using replacing the missing value with mean and detect outlier, Where the outliers and outlier's values were treated. In this way, the data were divided into two groups concerning the nearest neighbors, and the prediction is excellent. We obtained satisfactory values and predicted excellently.

Experiment VI: In the sixth experiment, we verified the implementation of the k-means fast unsupervised ML algorithm with pre-processing, as the Table 7 shows the most important values that we obtained.

The table 7 shows us the most important values that were obtained. We worked on the data that we downloaded, especially in chest surgeries. The x-means applied with pre-processing techniques to process this data, where the data in this technique was divided into one-third of the totals, each group contains values related to the data through the nearest neighbors.

Table.2. Centroid table for cluster model without pre-processing with k-means

| Attribute | Cluster 0 | Cluster 1 | Cluster 2 | Cluster 3 | Cluster 4 |
|---|---|---|---|---|---|
| AGE | 62.739 | 57.714 | 73.436 | 42.764 | 54.720 |
| PRE10 = F | 0.282 | 0.5714 | 0.238 | 0.352 | 0.395 |
| PRE10 = T | 0.717 | 0.428 | 0.761 | 0.647 | 0.604 |
| PRE11 = F | 0.8260 | 1 | 0.738 | 0.823 | 0.922 |

| | | | | | |
|---|---|---|---|---|---|
| PRE11 = T | 0.173 | 0 | 0.261 | 0.176 | 0.077 |
| PRE14 = OC11 | 0.369 | 0.285 | 0.357 | 0.588 | 0.387 |
| PRE14 = OC12 | 0.565 | 0.642 | 0.571 | 0.294 | 0.519 |
| PRE14 = OC13 | 0.038 | 0.071 | 0.031 | 0 | 0.054 |
| PRE14 = OC14 | 0.027 | 0 | 0.0396 | 0.117 | 0.038 |
| PRE17 = F | 0.913 | 0.9285 | 0.904 | 1 | 0.953 |
| PRE17 = T | 0.086 | 0.071 | 0.095 | 0 | 0.046 |
| PRE19 = F | 0.9943 | 1 | 1 | 1 | 0.992 |
| PRE19 = T | 0.005 | 0 | 0 | 0 | 0.007 |
| PRE25 = F | 0.972 | 1 | 0.984 | 1 | 0.992 |
| PRE30=F | 0.147 | 0.357 | 0.183 | 0.353 | 0.178 |
| PRE30=T | 0.853 | 0.643 | 0.817 | 0.647 | 0.822 |
| PRE32=F | 0.989 | 1 | 1 | 1 | 1 |
| PRE32=T | 0.011 | 0 | 0 | 0 | 0 |
| PRE4 | 3.234 | 3.158 | 2.980 | 4.040 | 3.556 |
| PRE5 | 2.507 | 71 | 2.232 | 3.290 | 2.750 |
| PRE6 = PRZ0 | 0.266 | 0.642 | 0.206 | 0.176 | 0.333 |
| PRE6 = PRZ1 | 0.695 | 0.357 | 0.650 | 0.823 | 0.651 |
| PRE6 = PRZ2 | 0.038 | 0 | 0.142 | 0 | 0.015 |
| PRE7 = F | 0.934 | 0.714 | 0.928 | 1 | 0.953 |
| PRE7 = F | 0.065 | 0.285 | 0.071 | 0 | 0.046 |
| PRE8 = F | 0.842 | 0.642 | 0.849 | 0.941 | 0.891 |
| PRE8 = F | 0.157 | 0.357 | 0.1507 | 0.058 | 0.108 |
| PRE9 = F | 0.956 | 0.571 | 0.944 | 0.941 | 0.930 |
| PRE9 = T | 0.043 | 0.428 | 0.055 | 0.058 | 0.069 |

Table.3. Centroid table for cluster model without pre-processing with k-means (fast)

| Attribute | Cluster 0 | Cluster 1 | Cluster 2 | Cluster 3 | Cluster 4 |
|---|---|---|---|---|---|
| PRE6 = PRZ1 | 0.698 | 0.375 | 0.250 | 0.659 | 0.500 |
| PRE6 = PRZ0 | 0.195 | 0.625 | 0.750 | 0.314 | 0.500 |
| PRE6 = PRZ2 | 0.106 | 0 | 0 | 0.026 | 0 |
| PRE7 = F | 0.9206 | 0.625 | 1 | 0.955 | 0.500 |
| PRE7 = T | 0.079 | 0.375 | 0 | 0.044 | 0.500 |
| PRE8 = F | 0.835 | 0.375 | 1 | 0.880 | 1 |
| PRE8 = T | 0.164 | 0.625 | 0.0 | 0.119 | 0.0 |
| PRE9 = F | 0.947 | 0.500 | 0.750 | 0.943 | 0.500 |
| PRE9 = T | 0.052 | 0.500 | 0.250 | 0.056 | 0.500 |
| PRE10 = T | 0.767 | 0.500 | 0.250 | 0.644 | 0.500 |
| PRE10 = F | 0.232 | 0.500 | 0.750 | 0.355 | 0.500 |
| PRE11 = T | 0.248 | 0.0 | 0.0 | 0.116 | 0.0 |
| PRE11 = F | 0.751 | 1.0 | 1.0 | 0.883 | 1.0 |
| PRE14 = OC14 | 0.031 | 0.0 | 0.0 | 0.041 | 0.0 |

| | | | | | |
|---|---|---|---|---|---|
| PRE14 = OC12 | 0.560 | 0.750 | 0.500 | 0.531 | 0.500 |
| PRE14 = OC11 | 0.375 | 0.125 | 0.500 | 0.382 | 0.500 |
| PRE14 = OC13 | 0.0317 | 0.125 | 0.0 | 0.044 | 0.0 |
| PRE17 = F | 0.904 | 1.0 | 0.750 | 0.940 | 1.0 |
| PRE17 = T | 0.095 | 0.0 | 0.25 | 0.059 | 0.0 |
| PRE19 = F | 1.0 | 1.0 | 1.0 | 0.992 | 1.0 |
| PRE19 = T | 0.0 | 0.0 | 0.0 | 0.992 | 1.0 |
| PRE25 = F | 0.978 | 1.0 | 1.0 | 0.007 | 1.0 |
| PRE25 = T | 0.0211 | 0.0 | 0.0 | 0.014 | 1.0 |
| PRE30 = T | 0.841 | 0.5 | 1.0 | 0.816 | 0.5 |
| PRE30 = F | 0.1582 | 0.5 | 0.0 | 0.183 | 0.5 |
| PRE32 = F | 1.0 | 1.0 | 1.0 | 0.992 | 1.0 |
| PRE32 = T | 0.0 | 0.0 | 0.0 | 0.007 | 0.0 |
| PRE4 | 3.044 | 2.92 | 3.027 | 3.4559 | 4.375 |
| AGE | 70.936 | 63.375 | 54.75 | 56.838 | 41.0 |

Table.4. Centroid table for cluster model without pre-processing with k-medoids algorithm

| Attribute | Cluster 0 | Cluster 1 |
|---|---|---|
| PRE6 = PRZ1 | 1.0 | 0.0 |
| PRE6 = PRZ0 | 1.0 | 0.0 |
| PRE6 = PRZ2 | 0.0 | 0.0 |
| PRE7 = F | 0.0 | 1.0 |
| PRE7 = T | 0.0 | 0.0 |
| PRE8 = F | 1.0 | 1.0 |
| PRE8 = T | 0.0 | 0.0 |
| PRE9 = F | 1.0 | 1.0 |
| PRE9 = T | 0.0 | 0.0 |
| PRE10 = T | 0.0 | 1.0 |
| PRE10 = F | 1.0 | 0.0 |
| PRE11 = T | 0.0 | 1.0 |
| PRE11 = F | 1.0 | 0.0 |
| PRE14 = OC14 | 0.0 | 0.0 |
| PRE14 = OC12 | 1.0 | 1.0 |
| PRE14 = OC11 | 0.0 | 0.0 |
| PRE14 = OC13 | 0.0 | 0.0 |
| PRE17 = F | 1.0 | 1.0 |
| PRE17 = T | 0.0 | 0.0 |
| PRE19 = F | 1.0 | 1.0 |

| | | |
|---|---|---|
| PRE19 = T | 0.0 | 0.0 |
| PRE25 = F | 1.0 | 1.0 |
| PRE25 = T | 0.0 | 0.0 |
| PRE30 = T | 1.0 | 1.0 |
| PRE30 = F | 0.0 | 0.0 |
| PRE32 = F | 1.0 | 1.0 |
| PRE32 = T | 0.0 | 0.0 |
| PRE4 | 4.72 | 1.96 |
| PRE5 | 3.56 | 1.68 |
| AGE | 51.0 | 79.0 |

Table.5. Centroid table for cluster model with pre-processing with k-meeans

| Attribute | Cluster 0 | Cluster 1 | Cluster 2 | Cluster 3 | Cluster 4 |
|---|---|---|---|---|---|
| AGE | 62.739 | 57.714 | 73.436 | 42.764 | 54.720 |
| PRE10 = F | 0.282 | 0.5714 | 0.238 | 0.352 | 0.395 |
| PRE10 = T | 0.717 | 0.428 | 0.761 | 0.647 | 0.604 |
| PRE11 = F | 0.8260 | 1 | 0.738 | 0.823 | 0.922 |
| PRE11 = T | 0.173 | 0 | 0.261 | 0.176 | 0.077 |
| PRE14 = OC11 | 0.369 | 0.285 | 0.357 | 0.588 | 0.387 |
| PRE14 = OC12 | 0.565 | 0.642 | 0.571 | 0.294 | 0.519 |
| PRE14 = OC13 | 0.038 | 0.071 | 0.031 | 0 | 0.054 |
| PRE14 = OC14 | 0.027 | 0 | 0.0396 | 0.117 | 0.038 |
| PRE17 = F | 0.913 | 0.9285 | 0.904 | 1 | 0.953 |
| PRE17 = T | 0.086 | 0.07 | 0.095 | 0 | 0.046 |
| PRE19 = F | 0.9943 | 1 | 1 | 1 | 0.992 |
| PRE19 = T | 0.005 | 0 | 0 | 0 | 0.007 |
| PRE25 = F | 0.972 | 1 | 0.984 | 1 | 0.992 |
| PRE30=F | 0.147 | 0.357 | 0.183 | 0.353 | 0.178 |
| PRE30=T | 0.853 | 0.643 | 0.817 | 0.647 | 0.822 |
| PRE32=F | 0.989 | 1 | 1 | 1 | 1 |
| PRE32=T | 0.011 | 0 | 0 | 0 | 0 |
| PRE4 | 3.234 | 3.158 | 2.980 | 4.040 | 3.556 |
| PRE5 | 2.507 | 71 | 2.232 | 3.290 | 2.750 |
| PRE6 = PRZ0 | 0.266 | 0.642 | 0.206 | 0.176 | 0.333 |

| | | | | | |
|---|---|---|---|---|---|
| PRE6 = PRZ1 | 0.695 | 0.357 | 0.650 | 0.823 | 0.651 |
| PRE6 = PRZ2 | 0.038 | 0 | 0.142 | 0 | 0.015 |
| PRE7 = F | 0.934 | 0.714 | 0.928 | 1 | 0.953 |
| PRE7 = T | 0.065 | 0.285 | 0.071 | 0 | 0.046 |
| PRE8 = F | 0.842 | 0.642 | 0.849 | 0.941 | 0.891 |
| PRE8 = T | 0.157 | 0.357 | 0.1507 | 0.058 | 0.108 |
| PRE9 = F | 0.956 | 0.571 | 0.944 | 0.941 | 0.930 |
| PRE9 = T | 0.043 | 0.428 | 0.055 | 0.058 | 0.069 |

Table.6. Centroid table for cluster model with pre-processing with k-medoids algorithm

| Attribute | Cluster 0 | Cluster 1 |
|---|---|---|
| PRE6 = PRZ1 | 1.0 | 0.0 |
| PRE6 = PRZ0 | 1.0 | 0.0 |
| PRE6 = PRZ2 | 0.0 | 0.0 |
| PRE7 = F | 0.0 | 1.0 |
| PRE7 = T | 0.0 | 0.0 |
| PRE8 = F | 1.0 | 1.0 |
| PRE8 = T | 0.0 | 0.0 |
| PRE9 = F | 1.0 | 1.0 |
| PRE9 = T | 0.0 | 0.0 |
| PRE10 = T | 0.0 | 1.0 |
| PRE10 = F | 1.0 | 0.0 |
| PRE11 = T | 0.0 | 1.0 |
| PRE11 = F | 1.0 | 0.0 |
| PRE14 = OC14 | 0.0 | 0.0 |
| PRE14 = OC12 | 1.0 | 1.0 |
| PRE14 = OC11 | 0.0 | 0.0 |
| PRE14 = OC13 | 0.0 | 0.0 |
| PRE17 = F | 1.0 | 1.0 |
| PRE17 = T | 0.0 | 0.0 |
| PRE19 = F | 1.0 | 1.0 |
| PRE19 = T | 0.0 | 0.0 |
| PRE25 = F | 1.0 | 1.0 |
| PRE25 = T | 0.0 | 0.0 |
| PRE30 = T | 1.0 | 1.0 |

| | | |
|---|---|---|
| PRE30 = F | 0.0 | 0.0 |
| PRE32 = F | 1.0 | 1.0 |
| PRE32 = T | 0.0 | 0.0 |
| PRE4 | 4.72 | 1.96 |
| PRE5 | 3.56 | 1.68 |
| AGE | 51.0 | 79.0 |

Table.7. Centroid table for cluster model with pre-processing with x-means

| Attribute | Cluster 0 | Cluster 1 | Cluster 2 |
|---|---|---|---|
| AGE | 70.93650793650794 | 57.714285714285715 | 56.838951310861425 |
| PRE4 | 3.0444444444444447 | 3.158571428571429 | 3.4559925093632966 |
| PRE5 | 2.3016402116402124 | 70.99999999999999 | 2.6901872659176034 |
| PRE19 = F | 1.0 | 1.0 | 0.9925093632958801 |
| PRE32 = F | 1.0 | 1.0 | 0.9925093632958801 |
| PRE25 = F | 0.9788359788359788 | 1.0 | 0.9850187265917603 |
| PRE9 = F | 0.9470899470899471 | 0.5714285714285714 | 0.9438202247191011 |
| PRE7 = F | 0.9206349206349206 | 0.7142857142857143 | 0.9550561797752809 |
| PRE17 = F | 0.9047619047619048 | 0.9285714285714286 | 0.9400749063670412 |
| PRE30 = T | 0.8412698412698413 | 0.6428571428571429 | 0.8164794007490637 |
| PRE8 = F | 0.8359788359788359 | 0.6428571428571429 | 0.8801498127340824 |
| PRE10 = T | 0.7671957671957672 | 0.42857142857142855 | 0.6441947565543071 |
| PRE11 = F | 0.7513227513227513 | 1.0 | 0.8838951310861424 |
| PRE6 = PRZ1 | 0.6984126984126984 | 0.35714285714285715 | 0.6591760299625468 |
| PRE14 = OC12 | 0.5608465608465608 | 0.6428571428571429 | 0.5318352059925093 |
| PRE14 = OC11 | 0.37566137566137564 | 0.2857142857142857 | 0.38202247191011235 |
| PRE11 = T | 0.24867724867724866 | 0.0 | 0.11610486891385768 |
| PRE10 = F | 0.2328042328042328 | 0.5714285714285714 | 0.35580524344569286 |
| PRE6 = PRZ0 | 0.19576719576719576 | 0.6428571428571429 | 0.3146067415730337 |
| PRE8 = T | 0.164021164021164 | 0.35714285714285715 | 0.1198501872659176 |
| PRE30 = F | 0.15873015873015872 | 0.0 | 0.18352059925093633 |
| PRE6 = PRZ2 | 0.10582010582010581 | 0.07142857142857142 | 0.026217228464419477 |
| PRE17 = T | 0.09523809523809523 | 0.07142857142857142 | 0.0599250936329588 |
| PRE7 = T | 0.07936507936507936 | 0.2857142857142857 | 0.0449438202247191 |
| PRE9 = T | 0.05291005291005291 | 0.42857142857142855 | 0.056179775280898875 |
| PRE14 = OC14 | 0.031746031746031744 | 0.0 | 0.04119850187265917 |

| | | | |
|---|---|---|---|
| PRE14 = OC13 | 0.031746031746031744 | 0.07142857142857142 | 0.0449438202247191 |
| PRE25 = T | 0.021164021164021163 | 0.0 | 0.0149812734082397 |
| PRE19 = T | 0.0 | 0.0 | 0.00749063670411985 |
| PRE32 = T | 0.0 | 0.0 | 0.00749063670411985 |

### *4.4 Experiments through classification stage*

Experiment I: In this experiment, we worked with supervised ML algorithms, unlike the first experiments with unsupervised ML techniques, where we investigated this experiment with the implementation of algorithms without pre-processing.

Table.8. The obtained results through ensembles without pre-processing with Dt and RF with Rapid Miner

| | Precision | Recall | Accuracy | F1 |
|---|---|---|---|---|
| Bagging | 85.11 | 100 | 85.11 | 91.96 |
| Boosting | 85.11 | 100 | 85.11 | 91.96 |
| Voting | 85.11 | 100 | 85.11 | 91.96 |
| Stacking | 86.03 | 97.50 | 84.40 | 91.40 |

We have worked on all the data that was related to thoracic surgery. Table 8 shows that the results were obtained with the application of supervised ML algorithms. The ensemble algorithms were applied without pre-processing. It showed our results predict well. The obtained precision, recall, accuracy, and F1 were 86.03, 100, 85.11, and 91.96%. These values are considered values for one-third of the algorithms, and they are the highest as shown in the table, where the highest accuracy in one-third of the algorithms is bagging, boosting, and voting, the highest accuracy in these algorithms, and this paper reached 86.03%, Given the difficulty of this data and this disease, this result is considered good and will be an excellent predictor.

Experiment II: According to this experiment, we verified the application of pre-treatment to treat missing values by applying pre-treatment techniques replacing the missing value with mean, and detect outliers with an ensemble algorithm, where we obtained good values as shown in the Table 9.

According to the table, we have reached good values from the data was collected from the UCI website, and this data is good, but it needs processing. Here, we applied the ensemble algorithms with pre-treatment, where we obtained good values that will predict the disease of chest surgeries, as we proved that our proposal is good and outperforms.

It ranked its peers where the values of precision, recall, accuracy, and F1 were 86.17, 100, 86.17, 92.57%. These values are considered the highest values obtained in this article, as shown in the Table 9.

Table.9. The obtained results through ensembles with pre-processing and DT and RF with Rapid Miner

| | Precision | Recall | Accuracy | F1 |
|---|---|---|---|---|
| Bagging | 86.17 | 100 | 86.17 | 92.57 |
| Boosting | 86.17 | 100 | 86.17 | 92.57 |
| Voting | 86.17 | 100 | 86.17 | 92.57 |
| Stacking | 86.17 | 100 | 86.17 | 92.57 |

In this article, we used Rapid Miner to do the processing and algorithmic work on the data we downloaded from the UCI Thoracic Surgical Registry website. Our work in this article is divided as follows. We conducted two experiments. The first experiment used clustering methods with DBSCAN, k-means, k-means (fast), k-means (H2O), k-medoids algorithm, random clustering, Support Vector Clustering, mean Where pre-treatment methods have also been applied. These techniques were applied with/without pre-treatment. The first experiment is we applied the k-means technique without pre-treatment, where the table appears in the first experiment, where we divided the data into five groups, where k = 5. Through the nearest neighbor was predicted, where it contains many values in the first table, including 54.720, 42.764, 73.436, 57.714, 62.73. The second experiment was conducted without pre-treatment by applying the unsupervised k-means (fast) machine learning algorithm, where the data was divided into five groups. The value of one of the attributes in the table with five totals is 0.500, 0.659, 0.250, 0.375, and 0.698. It indicates that the second experiment is well and predicts excellently. We did a third experiment without pre-processing with the k-medoids algorithm, where the results showed through the table. The data were divided into two groups, where k = 2, where the nearest neighbors were predicted, and these values are indicative of the

success of the methods used. A third experiment with pre-processing with k-means was used, where the data were divided into five groups, and each group contains a group of data, where one of the values reached in the fourth table 54, 720, 42.764, 73.436, 57.714, 62.739. This method is considered well for predicting chest surgeries, and we conducted the fifth experiment with unsupervised ML methods. It was used with pre-processing with the k-medoids algorithm, where the data were divided into two groups to predict better doubts through the nearest neighbors. This technique proved good through the fifth table, and it was found that it is good and predicts the best results. The values are in the sixth table, and this shows that our work is good and does an excellent job and predicts the best ways as our proposal has proven that it does a good job and will improve the performance of the classification. In the seventh experiment, we used supervised ML techniques, working on the downloaded data for thoracic surgery. Using the Rapid Miner tool, the ensemble techniques were applied without pre-treatment, where we used techniques without pre-treatment, wherein the seventh table the values we obtained through the application of bagging, stacking, voting, and boosting techniques, where the values of precision, recall, and accuracy and F1 without pre-processing with a DT and RF. The highest criteria in the Table 8 reached 86.03, 100, 85.11, 91.96%. These values are good and predict our proposal has proven that it is good and outperforms its peers. In the eighth experiment, we used supervised ML techniques, the ensemble with bagging, stacking, voting, and boosting with pre-processing techniques. The Table 9 showed values of precision, recall, accuracy, and F1 were 86.17, 100, 86.17, and 92.57%. It proves that our work and our proposal is a good proposal and predicts and improves thoracic surgery. The comparison among the obtained results through ensembles with pre-processing and other works.

Figures 10-27 illustrate the ROC curves that were obtained from the classifications. These figures are considered to show that our work is good and that our proposal is good where tables, values , and figures confirm that our work is superior to its counterparts. Figures 11-17 are considered drawings without pre-processing with clustering tasks. Figures 18-23 represent the ROC for the best of the current authors with pre-processing with the Rapid Miner tool and clustering. Figures 24-28 represent the ROC for the best of the current authors without/with pre-processing with the Rapid Miner tool with the ensemble. It indicates and proves that our work will predict and improve the performance of chest surgery disease and improve its performance.

According to the Table 9, we made a comparison among our work and the previous works, as the accuracy of our work amounted to 86.17%, which outperformed the previous works as shown in this Table 10. It proves that our work is successful work and our proposal has been excellently predicted and is considered an improvement in the performance of chest surgery and predicts it through the results obtained.

According to the Table 10, we made a comparison between our work and the previous works added in related works. It found that the proposed model for the current financiers is superior to its counterparts in terms of results and performance, where the highest accuracy is the accuracy that we obtained, reaching 86.17%, and this confirms our superiority over the previous works through the accuracy we obtained.

Table.10. A comparison among the obtained results through ensembles with pre-processing and other works

| | Precision | Recall | Accuracy | F1 |
|---|---|---|---|---|
| Koklu et al. | _ | _ | 84 | _ |
| Salati et al. | _ | _ | 70 | _ |
| Siu et al. | _ | _ | 44.7 | _ |
| **Our** | **86.17** | **100** | **86.17** | **92.56** |

Figures 11-17 show the ROC for the best of the current authors without pre-processing and figures 18-23 show the ROC for the best of the current authors with pre-processing with Rapid Miner tool with clustering. Figures 24-28 show the ROC for the best of the current authors without/with processing with Rapid Miner tool with ensemble.

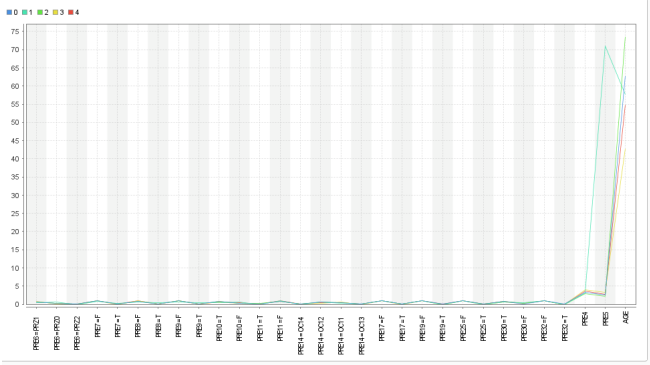

Fig.11. The ROC for cluster model with k-means without pre-processing

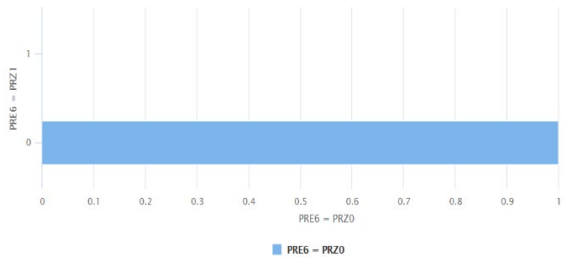


Fig.12. The ROC for Bar (Horizontal) with k-means

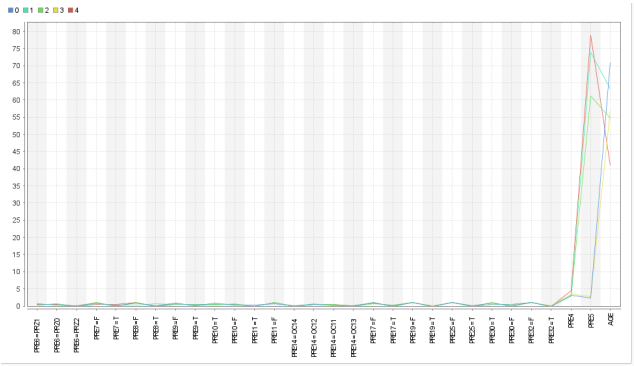

Fig.13. The ROC for cluster model with k-means fast

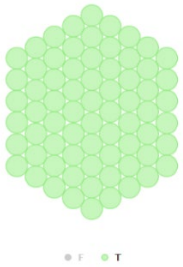

Fig.14. The ROC for packed bubble with DBCCAN

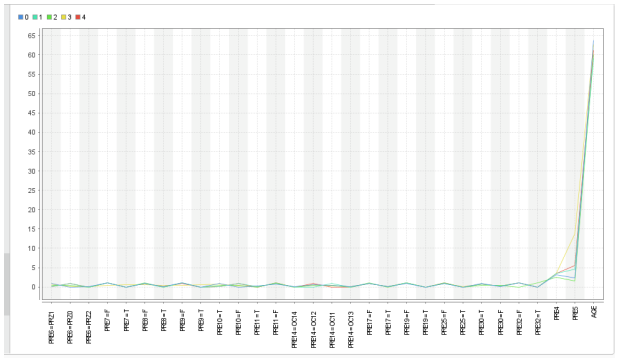

Fig.15. The ROC for cluster model with k-means (H2o)

Fig.16. The ROC for pyramid with random clustering

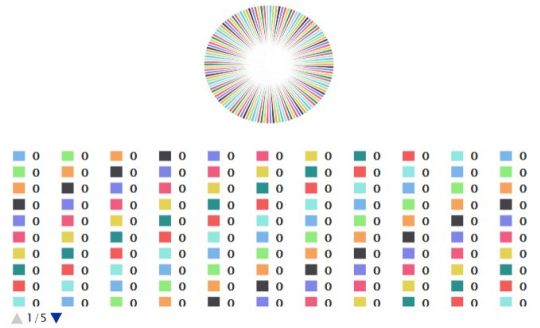

Fig.17. The ROC for pie/donut with DBSCAN

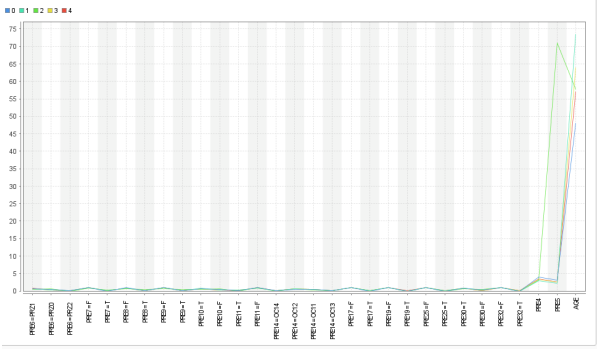

Fig.18. The ROC for cluster model with k-means

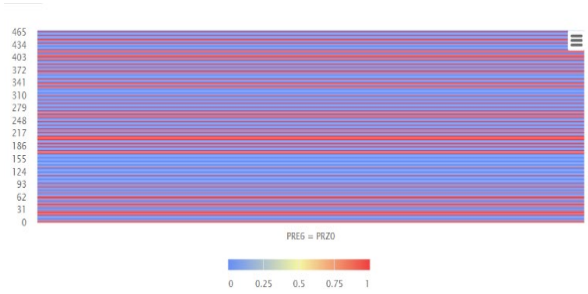


Fig.19. The ROC for with Heat map with k-means

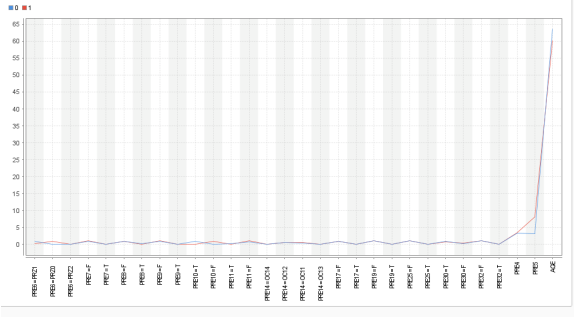

Fig.20. The ROC for cluster model with k-means (H2O)

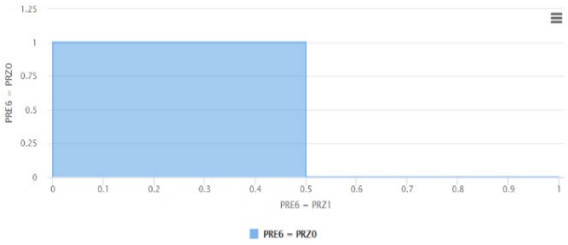


Fig.21. The step area with k-medoids algorithm

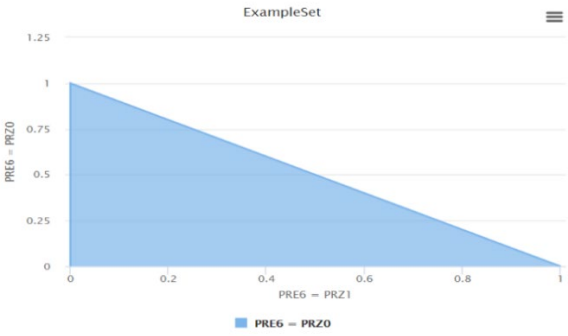


Fig.22. The ROC for area with random clustering

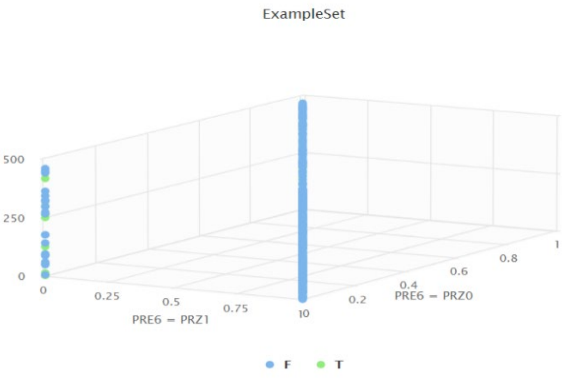


Fig.23. The ROC for with scatter3D with Support Vector Clustering

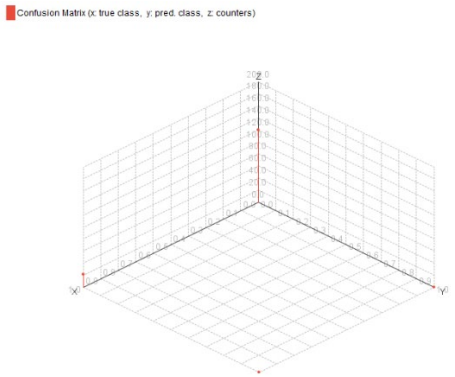


Fig.24. The ROC for with confusion with bagging

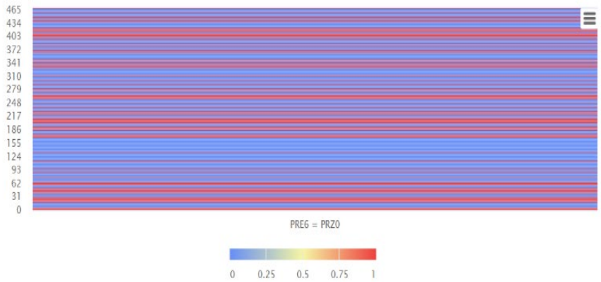


Fig.25. The ROC for with Heat map with Boosting

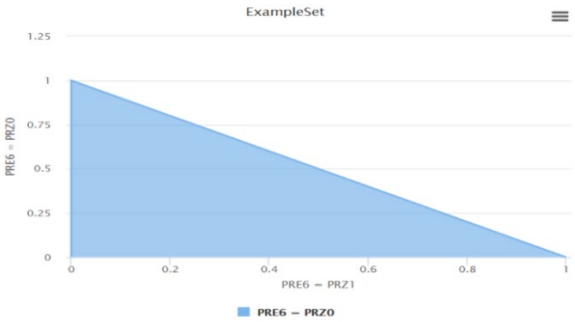


Fig.26. The ROC for with Area with stacking

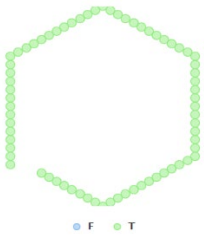


Fig.27. The ROC for Becket Bubble

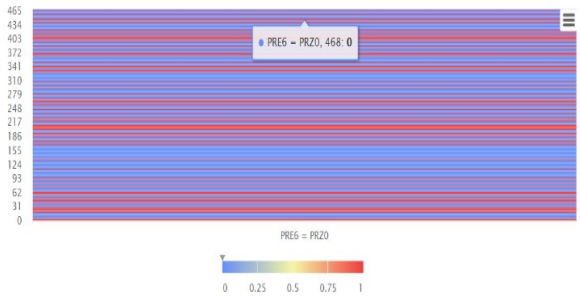


Fig.28. The ROC for with Heat map with voting

## V. CONCLUSION

Thoracic surgery is one of the most dangerous diseases that affect all age groups. It has the lost values and the missing values that could be addressed to ensure an excellent prediction. For example, the heart is part of the chest surgeries as a sensitive place. Outliers and missing values are two important problems that affect the performance of models. In this study, we proposed models for the diagnosis of these diseases through two experiments. The first experiment was used unsupervised ML techniques through clustering algorithms. This experiment was applied to pre-process with DBSCAN, k-means, k-means(fast), k-means (H2O), k-medoids algorithm, random clustering, Support Vector Clustering, x-mean. The obtained results proved that the proposed model is well and predicts excellently. The obtained values with k-means were 0.069, 0.058, 0.055, 0.428, and 0.043. Through these results, our proposed work has proven that it is superior to its peers. As this data is considered well-known data, and we worked on developing it and improving its performance. That is why we applied these techniques with pre-treatment and without pre-treatment. Experiments from number one to number 6 were obtained through good experiments and gave values and yields excellently. It confirms that our work is good and improves the performance of the disease. In the second experiment, we applied ensemble with/without pre-processing techniques. Our work was divided as follows in the seventh table. We applied these algorithms to work on them with the Rapid Miner tool. The work of these techniques was divided into the Education and Training Department to predict excellently. The highest accuracy, recall, precision, and F1 without pre-processing were 86.03, 100, 85.11, 91.96%. These values are considered high and will predict and improve the performance of thoracic surgery. The highest accuracy, recall, precision, and F1 with ensembles were 86.17, 100, 86.17, and 92.57%. The model presented to the current authors has proven to be superior to its peers. In future work, we will apply a heuristic model to improve data performance and predict excellently.

## REFERENCES

[1]. Peng Z, Li H, Zhang C, Qian X, Feng Z, and Zhu S, (2014). "A retrospective study of chronic post-surgical pain following thoracic surgery: prevalence, risk factors, incidence of neuropathic component, and impact on quality of life," PLoS One, vol. 9, p. e90014.

[2]. Cao C, Manganas C, Ang SC, Peeceeyen S, and Yan TD, (2013). "Video-assisted thoracic surgery versus open thoracotomy for non-small cell lung cancer: a meta-analysis of propensity score-matched patients," Interactive cardiovascular and thoracic surgery, vol. 16, pp. 244-249.

[3]. Case JB, Mayhew PD, and Singh A, (2015). "Evaluation of video-assisted thoracic surgery for treatment of spontaneous pneumothorax and pulmonary bullae in dogs," Veterinary Surgery, vol. 44, pp. 31-38.

[4]. Koklu M, Kahramanli H, and Allahverdi N, (2015). "Applications of rule-based classification techniques for thoracic surgery," in Management, Knowledge and Learning-Joint International Conference 2015-Technology, Innovation and Industrial Management TIIM, pp. 1991-1998.

[5]. Narula S, Shameer K, Salem Omar AM, Dudley JT, and Sengupta PP, (2016). "Machine-learning algorithms to automate morphological and functional assessments in 2D echocardiography," Journal of the American College of Cardiology, vol. 68, pp. 2287-2295.

[6]. Hachesu PR, Moftian N, Dehghani M, and Soltani TS, (2017). "Analyzing a lung cancer patient dataset with the focus on predicting survival rate one year after thoracic surgery," Asian Pacific journal of cancer prevention: APJCP, vol. 18, pp. 15-31.

[7]. Na X, Guo H, Zhang Y, Shen L, Wu S, and Li J, (2018). "Mining open payments data: analysis of industry payments to thoracic surgeons from 2014-2016," Journal of medical Internet research, vol. 20, pp. e11655.

[8]. Siu ICH, Li Z, and Ng CS, (2019). "Latest technology in minimally invasive thoracic surgery," Annals of translational medicine, vol. 7.

[9]. Groth SS, Habermann EB, and Massarweh NN, (2020). "United States administrative databases and cancer registries for thoracic surgery health services research," The Annals of thoracic surgery, vol. 109, pp. 636-644.

[10]. Salati M, Migliorelli L, Moccia S, Andolfi M, Roncon A, Guiducci GM, et al., (2012). "A Machine Learning Approach for Postoperative Outcome Prediction: Surgical Data Science Application in a Thoracic Surgery Setting," World Journal of Surgery, vol. 45, pp. 1585-1594.

[11]. Asgarnezhad, R, Monadjemi SA, and Aghaei MS, (2021). "A new hierarchy framework for feature engineering through multi-objective evolutionary algorithm in text classification," Concurrency and Computation: Practice and Experience.
[12]. Asgarnezhad, R and Monadjemi SA, (2021). "Persian sentiment analysis: feature engineering, datasets, and challenges," Journal of applied intelligent systems & information sciences, vol. 2, no. 2, pp. 1-21.
[13]. Asgarnezhad, R and Monadjemi SA, (2021). "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.
[14]. Asgarnezhad, R, Monadjemi SA, and Soltanaghaei M, (2021). "An application of MOGW optimization for feature selection in text classification," The Journal of Supercomputing, vol. 77, no. 6, pp. 5806-5839.
[15]. Asgarnezhad, R and Ali Mohsin Alhameedawi K, (2021). "MVO-Autism: An Effective Pre-treatment with High Performance for Improving Diagnosis of Autism Mellitus," Journal of Electrical and Computer Engineering Innovations, vol. 10, no. 1, pp. 209-220.
[16]. Asgarnezhad, R and Nematbakhsh N, (2015). "A reliable and energy efficient routing algorithm in WSN using learning automata," Journal of Theoretical & Applied Information Technology, vol. 82, no. 3, pp. 401-411.
[17]. Asgarnezhad R and Ali Mohsin Alhameedawi K, (2021). "MVO-Autism: An Effective Pre-treatment with High Performance for Improving Diagnosis of Autism Mellitus," Journal of Electrical and Computer Engineering Innovations (JECEI), 2021.
[18]. Asgarnezhad R and Ali Mohsin Alhameedawi K, (2022). "PEML-E: EEG eye state classification using ensembles and machine learning methods," Journal of Advances in Computer Engineering and Technology, vol. 7, no. 2, pp. 147-156.
[19]. Asgarnezhad R and Ali Mohsin Alhameedawi K, (2022). "Improving of Diabetes diagnosis using ensembles and machine learning methods," Majlesi Journal of Telecommunication Devices, vol. 11, no. 1, pp. 33-41.
[20]. Asgarnezhad R and Monadjemi SA, (2021). "NSE: An effective model for investigating the role of pre-processing using ensembles in sentiment classification," Journal of Advances in Computer Research, vol. 12, no.3, pp. 27-41.