Computer & Robotics

CKD-PML: Toward an effective model for Improving diagnosis of Chronic Kidney Disease

Razieh Asgarnezhad [a,*], Karrar Ali Mohsin Alhameedawi [a, b]

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

[b] *Department of Computer Engineering, Al-Rafidain University of Baghdad, Baghdad, Iraq*

Deceived 04 December 2021 , Accepted 05 February 2022

Abstract

Chronic Kidney Disease is one of the most common metabolic diseases. The challenge in this area is a pre-processing problem. Artificial Intelligence techniques have been implemented over medical disease diagnoses successfully. Classification systems aim clinicians to predict the risk factors that cause Chronic Kidney Disease. To address this challenge, we introduce an effective model to investigate the role of pre-processing and machine learning techniques for classification problems in the diagnosis of Chronic Kidney Disease. The model has four stages including, Pre-processing, Feature Selection, Classification, and Performance. Missing values and outliers are two problems that are addressed in the pre-processing stage. Many classifiers are used for classification. Two tools are conducted to reveal model performance for the diagnosis of Chronic Kidney Disease. The results confirmed the superiority of the proposed model over its counterparts.

Keywords: Pre-processing, Chronic Kidney Disease, Classification, Machine Learning Techniques

1.Introduction

Chronic kidney disease (CKD) is sometimes a dangerous disease for humans. Neglecting early detection of this disease leads to negative results and may lead to acute kidney failure. In this condition, the kidneys need to wash several times. It is a disease that affects children, adults, and the elderly. This disease is considered a gradual loss of the functions of chronic kidney disease. The symptoms of nephritis are not specific and may include interruption of human appetite in terms of eating. This disease is diagnosed by examining adults with age who suffer from diabetes and chronic pressure. These are the most exposed to this disease, or perhaps close relatives who contract this disease may be transmitted through genetics. This disease is considered dangerous in previous years. For example, for people who carry complications of anemia, blood and heart disease is one of the most necessary problems that this disease faces. There is no specific treatment to limit the aggravation of this disease [1]. This disease is treated by regular chronic kidney dialysis, and this is not enough to prompt us. We work on the application and analysis of some algorithms to improve the adoption of this disease and results [2] [3].

In 2016, the authors were presented a technical model using the Naïve Bayes (NB) Classifier to reduce the rules for the steps of chronic nephritis. The treatment is organized according to working times and the lack of conflict in times. They also decreased the number of features to 80%. Their research was good, but our research is better and significantly superior to their work. Our work reached the highest accuracy that can be 100% through the NB algorithm with proposed pre-processing. Our results indicate that our work has excellently predicted this disease and outperformed other works previously [4]. In 2018, authors have presented a technical system by proposing the automatic classification of the diagnosing algorithms to predict and diagnose chronic kidney disease. They received good enough values, and the highest value

* Corresponding Author Email: razyehan@gmail.com

was 94.602% through support vector machine (SVM). In comparison to our work, we obtained the highest accuracy of up to 100% through NB. Our results confirm that our research gave excellent results, and we predicted and improved satisfactorily. Chronic kidney disease is unparalleled, as we reached impressive results in this paper [5].

In this article, a technical model was presented to predict the best results in chronic nephropathy and improve its performance. We downloaded new data from the UCI website, where these data contained 400 records and 25 traits. But the data contain anomalous values and missing values. Here, we employed two experiments. The first experiment was with the Weka tool, where we applied high-accuracy techniques and gave good results. Many classifiers such as decision tree (DT), K-nearest neighbor (KNN), random forest (RF), neural network (NN), SVM, NB, with/without pre-processing applied for the classification stage. We got good results through the Weka program, which reached the highest measurement with the RF algorithm. The values of precision, recall, accuracy, f1 with/without pre-processing achieved 97.09, 99.02, 99, 99.02%, respectively. These values are revealed that our work obtained very high results than our counterparts.

In the second experiment, we applied the following algorithms with the program Rapid Miner tool DT, RF, NN, KNN, NB, SVM with pre-processing. We applied techniques to address the missing value problem. We used to replace the missing value with the mean and detect outliers. The obtained results proved that our pre-processing is well, and the highest accuracy values, precision, recall, and f1 were 99.02, 98.33, 98.75, and 98.67% with the RF algorithm without pre-processing, respectively. On the other hand, the pre-processing gave results of the highest values in this paper. Precision, recall, accuracy, and f1 were the same and equal to 100% with the NB algorithm. These values are considered the best and highest in his paper, as well as this percentage. It outperformed all previous works in predicting CKD. It also proved that this paper is considered successful research because it outperformed the rest of the works.

We have designed an effective technical model to distinct from its peers by giving high values. We used pre-processing methods to process the data that we downloaded from the UCI website and applied good techniques, including DT, KNN, RF, NN, NB, SVM. We reached high-accuracy results and succeeded in predicting. The highest accuracy through the NB algorithm was 100%, and this indicates that our innovation has proven its quality and proven that it works well for predicting CKD and improving classification performance.

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.Related work

Text should be produced within the dimensions shown on these pages; each column 8.47 cm wide with 0.85 cm middle margin, total width of 17.78 cm and a maximum length of 21cm on the first page and 23.5cm

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

In 2015, the authors proposed a model for the CRISP Investigators. They present a technical model to analyse and compare simultaneous ultrasound and magnetic resonance imaging to predict the best results and improve the performance of chronic nephritis. They obtained good values, but our work far exceeded this work, as our work reached an accuracy of 100%, and this value is considered very high compared to the previous work [6].

In 2016, the authors presented a technical model to predict chronic nephropathy. Then, they got good values up to the precision of 91, recall of 95, the accuracy of 90, and f1 of 92% through KNN. These values were considered the highest values in their work through six techniques, and it was considered a good value. But our work far exceeded this work, as we predicted excellently with an accuracy that reached 100% [7].

In 2017, the authors presented a dietary prediction for patients with the CKD by considering blood potassium level, and Machine Learning Algorithms. They were presented with a technical model to work on determining diet plans by predicting the potassium area. Good results were obtained [8].

In 2018, the authors proposed a technology model to predict chronic kidney disease. In their work, classification of CKD was applied using Logistic Regression (LR), Feedforward NN, and Wide & Deep Learning. They used modern computer-assisted methods for CKD, random forest techniques used. They applied three machine learning (ML) techniques to predict chronic kidney disease, logistic regression, nutrient neural networks, and wide

networks. The highest accuracy, precision, recall, and f1 were 97, 99, 99, and 99%, respectively [9].

In 2019, authors examined changes in N-acetyltransferase 8 in kidney tubular cells: injury, recovery, and mesenchymal stromal cell-based therapy. They proposed a model to improve kidney disease and prediction of this disease significantly. The presented model contributes to the protective enzyme NAT8. They performed operations inside the laboratory and showed satisfactory results in their work, but our present work, where the current authors suggested a technical model that gave excellent results and predicted kidney disease significantly and improved the performance of this classification.

In 2020, authors presented a NN-based Model for Predicting CKD. They proposed a high model for predicting chronic kidney disease. They applied the artificial NN to analyse 14 features and linked them to CKD. They reached satisfactory results [10].

In 2021, the authors proposed a technical model whereby it develops and improves the detection of CKD. They applied techniques and algorithms to predict better results, as both reached accuracy of 96.76, the precision of 92, recall of 94.74, and f1 of 97.30% by SVM. These values were well for predicting chronic nephritis. Washing the kidneys, which prompted many researchers to write and search for other techniques that improve the performance of classification, but our work has proven a great success because of its superiority over all previous work [11].

In this section, we studied many works and results. In this article, we applied two experiments. An experiment runs with the WEKA tool with/without a pre-processing stage. We also conducted the second experiment with the Rapid Miner tool with/without pre-processing stage through DT, RF, NN, SVM, NB, KNN classifiers. The data we worked on was processed, and we reached the highest value. All evaluation metrics were 100%, and these values are considered the highest values obtained in this article with the NB algorithm. It is proven that our technical proposal in this paper is a well and successful proposal that will improve the performance of CKD and predict better. Our results show how our work was distinguished in everything and will improve the classification performance and work on its improvement, and prediction faster in terms of accuracy.

In 2020, Khan et al., investigated a study to classify chronic kidney patient records and proposed a technical model that includes the application of seven techniques of supervised ML methods such as NB-Tree, J48, SVM, LR, Multilayer Perceptron, NB, and Composite Hypercube on Iterated Random Projection (CHIRP). They obtained the highest value in their paper. The following measurements reached recall of 99.98, precision of 99.98, accuracy of 98, and f-measure of 99.97%. These values are considered good and through them, the authors were able to quickly predict kidney disease. This disease is difficult to diagnose and there are always impurities in the diagnosis and classification, and thus this value is considered good compared to the previous work [12].

Alaiad et al. build a system that predicts the diagnosis of CKD and its main problems. They used supervised ML methods that included both DTs, SVM, KNN, and NB. The authors downloaded from the educational website for CKD, used mining tools such as Weka and SPSS to analyse data, train, develop, extract and evaluate results. The precision of 98, MAE of 1.50, accuracy of 98.50, sensitivity of 99.59, and specificity of 96.75% were achieved. With the KNN algorithm which is considered the highest in their presented research, after applying the integrative approach for all between supervised ML and mining algorithms in the domains of correlation rules, they were able to predict this disease, improve its quality and improve diagnosis [13].

In 2021, Parthiban et al. propose a technical model for scaling up kidney disease performance. They introduce the advantage of selecting this FS based on the Hybrid Filter Wrapper Embedded (HFWE). The authors intend to identify a subset of features relative to the CKD records and data. This disease is complex as many authors have done several experiments to detect this disease and with this research of their own, they implemented this algorithm based on three important things or rather functions. The most important methods that the authors applied in their work included NB-HFWEFS, ANNHFWE-FS, SVM - HFWE-FS, NB, NN, SVM. They reached good results and improved the performance of CKD. The most important findings of the authors were: sensitivity of 95.45, specificity of 87.50, precision of 95.45, f-measure of 95.45, and accuracy of 93.33%. This is the ratio of the results that they have reached to its goodness and the improvement of the classification performance [14].

The authors submitted a technical proposal that includes the application of supervised ML

techniques to compare their work with previous work. The main goal in their paper was to find and predict CKD and improve its performance and classification performance. They downloaded a data set of chronic kidney patients from the ML repository at the University of California, Irvine. Through their work, the authors developed eight ML models that used Python language. To detect this disease, the most important findings of the slanderers through their experiments, where they used these criteria to evaluate the models used included all of the accuracy, accuracy, sensitivity, F1, and ROC-AUC. Among these parameters, the highest accuracy of their research was reached, reaching 99.75%. This result is good compared to previous work. They were able to predict this disease and improve its performance [15].

3. The proposed model

Here, we have done two experiments to investigate the impact of pre-processing and ML techniques on our model, CKD-PML. For pre-processing, we used some methods to solve missing value and outlier problems. The employed ML techniques are RF, SVM, KNN, DT, NB, and NN [16] [17]. Figure 1 shows a flowchart of the proposed model. The brief description of the applied methods is of concern (Figure 2).

3.1. Pre-processing Stage

At this stage, we downloaded data from the UCI website and worked on it. This data contains missing values, and we applied some techniques to process it with the Weka and Rapid Miner tools. We applied the mean to detect outliers, and with new values obtained good results. We exceeded our work over the previous works because the CKD-PML model exceeded the limit of imagination. The results were well so that this indicates our model predicts popularly with chronic nephritis. We improved the level of these data and worked to present significant research that predicts and improves classification.

3.2. Classification Stage

In this paper, we will describe the methods that we used. An algorithm was applied to give impressive results, and these techniques are DT, KNN, RF, NB, NN, SVM with/without pre-processing. Using two tools, where we obtained very high values, and the highest accuracy in our article reached 100% with NB. This value is excellent for predicting and improving chronic kidney disease. We proved that our work outperforms the previous work, as we will explain these algorithms in detail below.

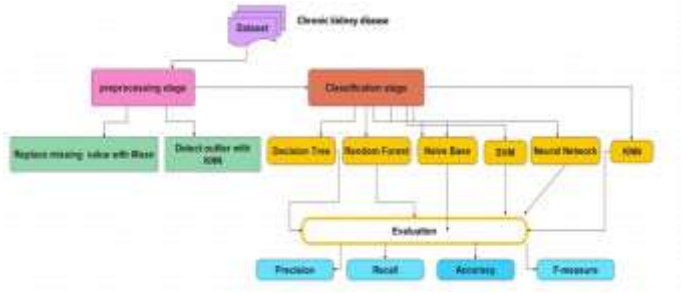


Fig .1. The diagram of the CKD-PML model

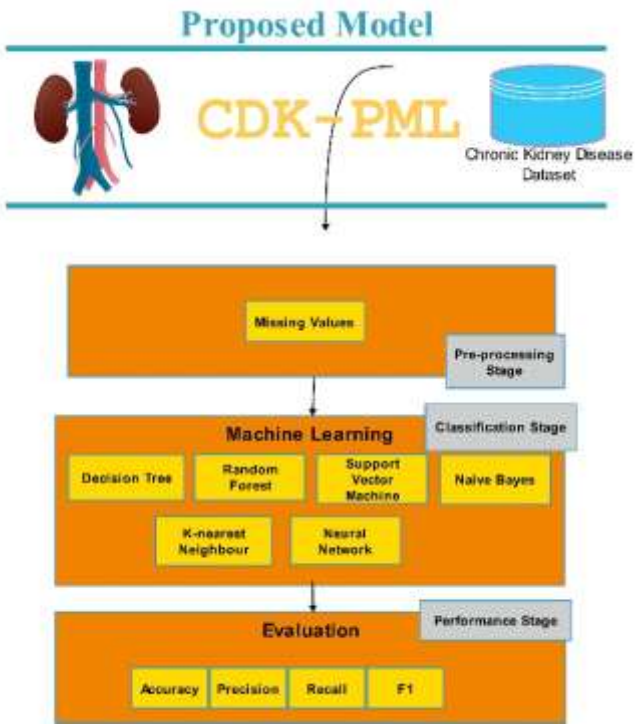


Fig .2. The phases of the CKD-PML model

DT: We applied the DT in our article with two experiences. The first was with the WEKA tool, with/without a pre-processing stage. This algorithm is a decision support tool used as an illustration showing the shape of a tree. This technique is one of the classification techniques that work to provide good values sometimes. In every work it is applied, the DT is used in researching the process in daily life. It is done with another experience with the Rapid Miner tool. It yielded good results as well, reaching 96.75% with/without a pre-processing

stage. It also reached accuracy with pre-processing with the Rapid Miner tool, where its accuracy was 97.50%. As shown in Table 3, these values are well for an excellent prediction, and the DT is good for the improvement of CKD.

RF: The RF tree is one of the algorithms that predict chronic kidney disease, as it consists of a set of individual decision trees that work as aggregates and always give good results. It is one of the classifiers that predict well. It is applied in this article with two experiments, the first experiment with the WEKA and Rapid Miner tool with/without a pre-processing stage. Good values were obtained, and it outperforms the previous ones, where it showed an accuracy of 99% with/without pre-processing in the WEKA tool. It is the second value after the NB value, and it is considered a high value that predicts excellent improve kidney disease. And then, we got a very high value with the rapid miner program, reaching 98.75, 99.17%. Using with/without a pre-processing, where the random forest is an algorithm that improves the performance of classification and prediction excellently and has outperformed its previous work and gave good results.

NN: The NN is a type of classification that we worked on in this article. In the first experiment, we obtained the results better than good, as this experiment reached 95, 95, 95, 95%, with/without a pre-processing in the WEKA tool. These values are considered high. The NN is a network of circles of neurons where the use of this network of predictive modeling provides good results. Then, this algorithm was applied with the Rapid Miner tool with/without a pre-processing to detect outliers. Results were obtained for precision, recall, accuracy, and f1 94.12, 96, 95, 95.05, respectively. These values are considered well and predicted skilfully, as they outperformed the previous ones in their work. It confirms the success of our work in predicting a large percentage of this disease and improving its performance.

NB: This technique was applied in our article, and the highest value was achieved. It is considered the best among the techniques that were used in our article. It is also considered one of the most famous methods of ML, data analysis, where this algorithm is characterized by fast processing and giving better results. It is also considered one of the classifications that are important because it sometimes gives impressive results. We applied this algorithm with/without pre-processing. The highest value was obtained in the article, reaching 100% with pre-processing, and this value is considered the best and outperforms the previous work. It will predict excellently and improve the classification performance, and it outperforms the previous work in terms of results.

KNN: It is considered one of the most important algorithms under the supervision of machines for ML. This algorithm works to determine the nearest neighbours. Then, it replaces the missing values with new values, where you choose the ten nearest neighbours and work to give a new section and replacing the stray values. We applied this algorithm with two tools with/without pre-processing. The high values were obtained with the first and second experiments 98.05% with the two tools. We have obtained this value, which is considered a good value compared to the previous works. We made an excellent prediction with this algorithm that works to improve the performance of chronic nephritis to improve the performance classification and work on giving high accuracy values and outperforming the previous works in its work.

SVM: This algorithm is considered to be another alternative for classification that predicts excellently and always gives good results. This algorithm is considered to be supervised learning and router support networks in ML. It is considered as data points closer to the super level and affects the direction of the site because kidney disease Chronic is considered a dangerous disease. This technique was applied with two experiments, WEKA and Rapid Miner. By replacing the missing value through pre-processing and without it, we obtained well results that give an excellent prediction. It proves the success of this technique, as all metrics were obtained equal to 95%. These are values that were obtained through this algorithm. It outperformed the previous work and predicted a high percentage. The results were handled with the Rapid Miner tool with/without pre-processing. We reached values with the following measurements with this technique. The highest accuracy, recall, precision, and F1 were 95.92, 97.33, 96.67, and 96.61%, respectively. These values are considered excellent for prediction, obtaining effective values, and improving chronic nephritis. The results proved that our model was successful, and it was able to improve the classification performance.

3.3. Evaluation Stage (Datasets and Metrics)

We collected the data for this research by searching for data related to the disease, CKD. It was found by searching the UCI website for this disease. This data contains 400 records and contains 25 traits. It contains missing values and stray values as well. We will work on this data and applied algorithms whose results are very well. We will apply classification algorithms, including KNN, RF, DT, NB, NN, SVM, and we worked on this data to improve the classification performance and predict CKD. This data was downloaded from the UCI website and worked on and obtained excellent results.

To evaluation, accuracy, precision, recall, and F1 measures were applied. These measures were defined in Table 1 [18] [19] [20].

Table 1
Parameters Definitions

Parameter	Equation
Accuracy	$(TP + TN) / (P + N)$
Specificity (TN Rate)	$(TN) / (TN + FP)$
Sensitivity (TP Rate)	$TP / (TP + FN)$
Precision	$(TP) / (TP + FP)$
Recall	TP / P
F-measure	$\frac{2\times precision\times recall}{precision+recall}$

4. Experiments

4.1. Experiment I

In Experiment I, we investigate the effect of ML techniques without/with a pre-processing stage on CKD through the WEKA tool.

According to Table 2, we have worked on the data that was downloaded from the UCI website. The data which we downloaded contains missing values and abnormal values. We obtained excellent values. The KNN, DT, RF, NN, SVM, NB were applied with/without pre-processing. We obtained very high values with the two experiments, which are the methods of applying to pre-process. The high values were reached through Table 2. The highest accuracy was reached in an experiment without pre-processing with the RF 99%. This value is considered to be a very high predictive value for kidney disease improvement. As for Table 2, we conducted another experiment with a pre-processing, where missing values with mean and detect outlier were applied. The highest accuracy reached with the RF algorithm is 99%. This is a very high value and outperforms the previous work and will predict better and gave excellent results CKD. Thus, we proved that our model in this research is a good enough model and improved the classification performance.

Table 2
The Obtained Results Through ML Techniques Without/With Pre-processing Stage in WEKA Tool (%)

ML Techniques		P	R	SE	SP	ACC	F1
Without pre-processing	NN	95	95	94.33	96.66	95	95
	NB	94	94	93.66	95.88	94	94
	KNN	96.08	98.08	97.66	98.77	98.05	98
	SVM	95	95	94.50	96.93	95	95
	DT	93.01	97.04	96.55	97.55	96.75	97.06
	RF	97.09	99.02	99.70	98.77	99	99.02
With pre-processing	NN	95	95	96.44	96.85	95	95
	NB	94	94	95.33	96.55	94	94
	KNN	96.08	98.08	97.64	98.04	98.05	98
	SVM	95	95	94.60	94.55	95	95
	DT	93.01	97.04	96.42	96.55	96.75	97.06
	RF	97.09	99.02	98.53	98.60	99	99.02

Note: P=Precision, R=Recall, SE=sensitivity, SP=specificity, ACC=Accuracy

4.2. Experiment II

In Experiment II, we investigate the effect of ML techniques without/with a pre-processing stage on CKD through the Rapid Miner tool.

According to Table 3, we have used the data that was collected from the UCI website, wherein this experiment and the results that are in Table 3. We worked with the Rapid Miner tool applied these algorithms, DT, KNN, RF, NN, SVM, NB. It was applied with/without pre-processing. The high values were obtained without pre-processing with the RF algorithm. This is considered a high value, but the highest in Table 3 that was reached with the pre-processing was obtained better and higher value in this research. The value of accuracy reached 100% with the NB algorithm. This value is well for predicting and considered one of the high values that outperformed all previous work. We have obtained a

value that will improve the performance of CKD herein. This is a fictional thing because we processed this data and applied algorithms in our model to give significant and high results for predicting and improving this disease.

The highest results are belonging to the NB classifiers. Hence, the computational complexity of model will be equal to this classifier. The current authors apply the pre-processing methods to can increase accuracy and other metrics in this context.

Table 3

The Obtained Results Through ML Techniques Without/With Pre-processing Stage in Rapid Miner Tool (%)

ML Techniques		P	R	SE	SP	ACC	F1
Without pre-processing	NN	94.12	96	95	96	95	95.05
	NB	98	99	98	99	98.75	98.49
	KNN	96.08	98.08	97	98	98.05	98
	SVM	95.92	97.33	96	96	96.67	96.61
	DT	98.07	96.66	95	96	97.50	97.35
	RF	99.02	98.33	97	98	98.75	98.67
With pre-processing	NN	94.12	96	95	98	95	95.05
	NB	100	100	3.4	1	100	100
	KNN	96.08	98.08	97	97	98.05	98
	SVM	95.92	97.33	96	98	96.67	96.61
	DT	95.75	95.33	96	97	95.83	96.02
	RF	99.34	98.89	97	98	99.17	99.67

Note: P=Precision, R=Recall, SE=sensitivity, SP=specificity, ACC=Accuracy

4.3. Discussions

Here, we will discuss our model in detail, and we will explain what are the methods used in this article. We have experimented using two tools. We have done the first experiment by working with the WEKA tool with/without pre-processing. The experiment with pre-processing used techniques to replace the missing value with the mean and detect outliers. The highest precision, recall, accuracy, and F1 with RF were 97.09, 99.02, 99, and 99.02%, respectively. These values are considered high in comparison to the previous works and confirm that our model is good enough and will predict and improve the classification performance in a large proportion since chronic nephritis is a disease that affects all groups. Its problem is that there is no treatment to reduce the aggravation of the number of injuries. This confirms that our first experiment with the WEKA tool is successful and improved the classification performance and germination significantly. These results are in Table 2, and these are the values that have become highly predictive values and improve the classification performance. In the second experiment, we worked on the data, but with the Rapid Miner tool, where we applied the following algorithms with/without pre-processing. These algorithms are DT, KNN, RF, NN, NB, SVM. In the first stage with pre-processing, techniques processed the data and replaced missing values. We applied these to replace the missing value with the mean and detect outliers. The results showed well with/without pre-processing. Without pre-processing, the measurements for precision, recall, accuracy, and F1 reached 99.02, 98.33, 98.75, 98.67%, respectively. Using the RF algorithm and with pre-processing in Table 3, we obtained high values, which are considered the best and highest values in the article. All metrics were equal to 100% through the NB algorithm, where these values are considered the highest and the best in this article, as they outperformed all previous works.

In Table 4, we compared our model with the previous works, where our work proved high results and outperformed the previous works in terms of accuracy. The accuracy of our work reached 100%, which is superior to all previous works, and this confirms that our work will predict excellently as chronic kidney disease in the years. The previous one was widely spread, and the lack of a treatment that limits the number of injuries. This is the reason that we prompted our model to search for alternative methods and techniques that predict better doubts and give better results. Many researchers worked and applied techniques, but our work proved to be the best and surpassed the previous works. Figures 3 to 7 show the comparison as well.

In Table 4, we made a comparison between our work and previous work, where the results showed that our work obtained the highest values for all evaluation metrics and was equal to 100% which indicates that the path of our paper is a good one and will greatly predict the improvement of data mining tasks. With this value, we will prove that corrected the path of this data and processed the data that we downloaded from the UCI website developed and we got significant results, which indicates our work outperformed all works and gave impressive results that improved the performance of CKD.

Table 4
A Comparison Among the Best Obtained Results and Other Existing Works (%)

Work	P	R	ACC	F1
[9]	99	99	99	97
[7]	91	95	90	92
[11]	92	94.74	96.67	97.30
CKD-PML (our)	100	100	100	100

Note: P=Precision, R=Recall, ACC=Accuracy

The ROC for the best classification of the current authors with pre-processing with the WEKA tool is shown in Figures 8 and 9. Figure 8: The ROC for the best classification of the current authors without pre-processing with the Rapid Miner tool showed in Figure 9: The ROC for the best classification of the current authors with pre-processing with the Rapid Miner tool showed in figures 10-15.

These figures showed the results of our work and the most important percentages that we obtained, as well as showing the way we worked, how we obtained well results, and how we will predict the improvement of chronic kidney disease, and this confirms the success of this paper in terms of the results we obtained.

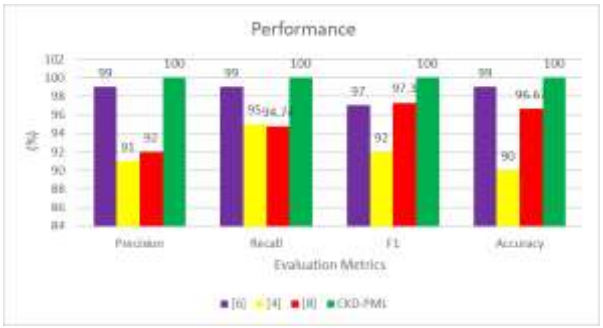


Fig .4. The comparison among our best results and its counterparts

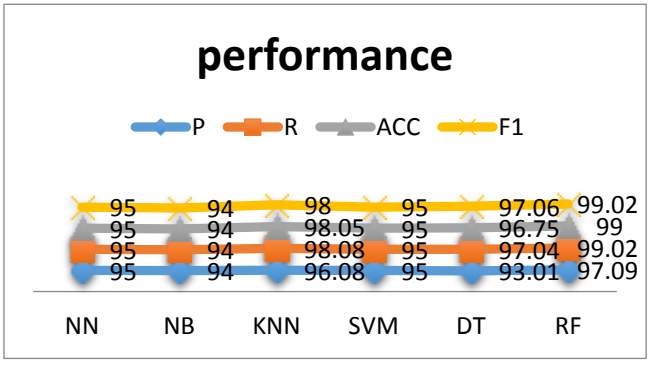


Fig .4. The obtained results through ML techniques with pre-processing stage in WEKA tool (%)

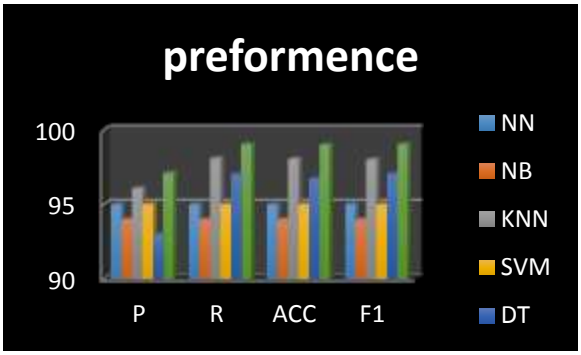


Fig .5. The obtained results through ML techniques without pre-processing stage in WEKA tool (%)

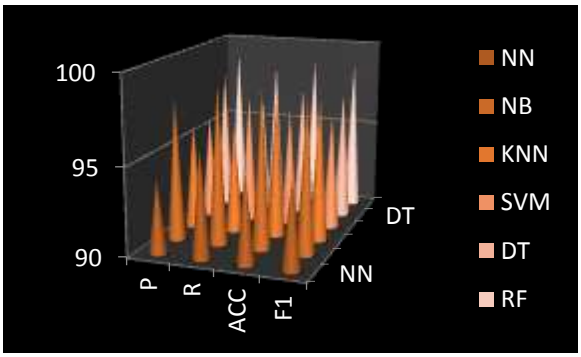


Fig .6. The obtained results through ML techniques with pre-processing stage in Rapid Miner tool (%)

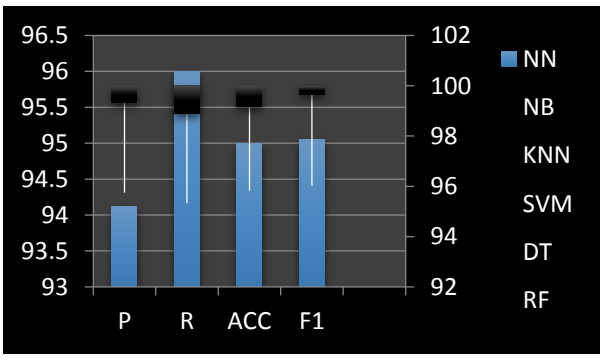


Fig .7. The obtained results through ML techniques without pre-processing stage in Rapid Miner tool (%)

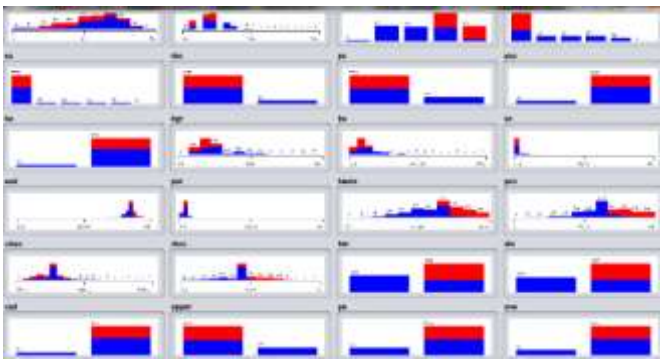


Fig .8. The ROC for the NB through WEKA tool

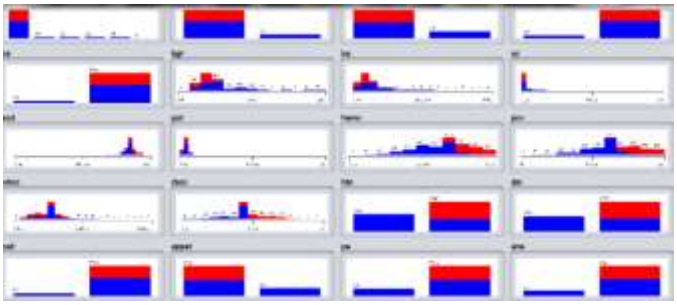
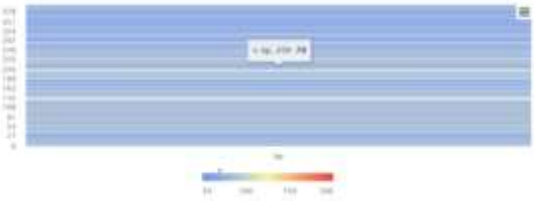
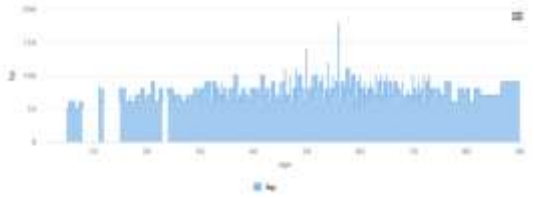


Fig .9. The ROC for the RF through WEKA tool

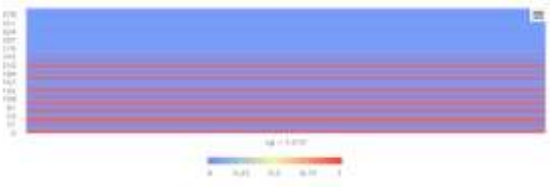


Heat map

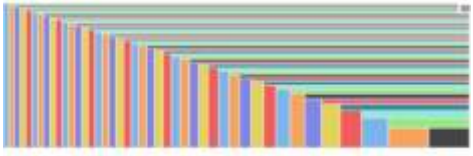


Step Area

Fig .10. The ROC for DT with Heat map and step Area through Rapid Miner tool



Heat map

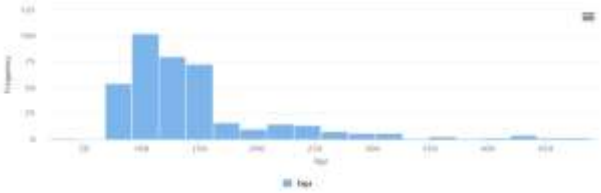


Tree map

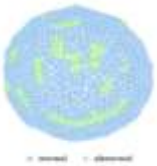
Packed Bubble

pyramid

Fig .11. The ROC for the DT with Heat map, tree map, packed Bubble, and pyramid through Rapid Miner tool



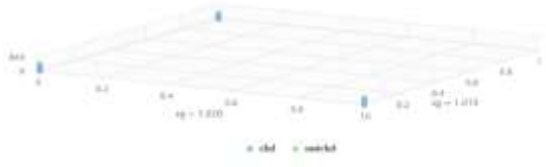
Histogram



packed Bubble

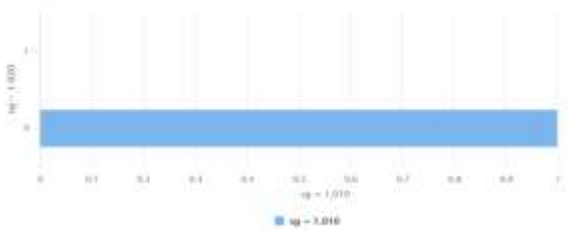
Ple/donut

Fig .12. The ROC for the DT with Histogram, Packed Bubble, and Ple/donut through Rapid Miner tool



Scatter 3d

Funnel packet Bubble



Bar (Horizontal)

Fig .13. The ROC for the DT with Scatter 3d, Funnel Packet Dabble, and Bar (Horizontal) through Rapid Miner tool

ple/donut

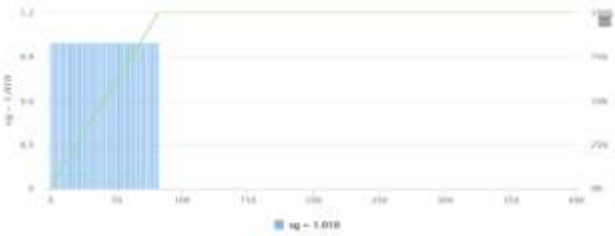
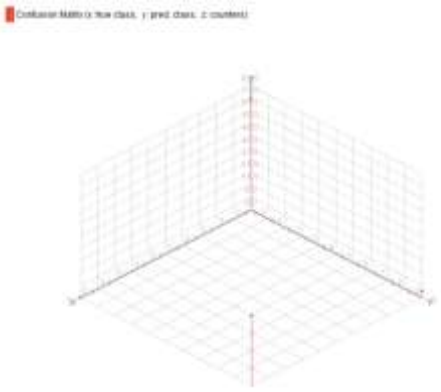
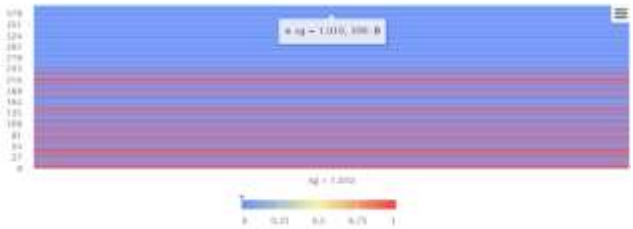


Fig .14. The ROC for the DT with Ple/donut and Pareto through Rapid Miner tool

Ple/donut



Confusion Matrix



Heat map

Fig .15. The ROC for the DT with Ple/donut, confusion Matrix, and Heat map through Rapid Miner tool

5. Conclusion

The used data were downloaded from the UCI website, where these data contain missing and stray values. In this research, we applied two experiments, where the first experiment we applied algorithms, namely DT, RF, NN, KNN, SVM, NB with/without pre-processing in The WEKA tool, which applied with the pre-processing techniques such as replace the missing value with mean and detect outliers. We obtained high values in the first experiment. Precision and recall and accuracy and f1 reached high values, and the highest value reached in Table 2, as shown in his paper. These reach 97.09, 99.02, 99, 99.02 with the RF algorithm, which is considered the best in giving high values. With these results, we have significantly improved the performance of classification and prediction, and this confirms that our model is superior to the previous work, as our results showed a remarkable development in the classification performance. Chronic nephrotic is considered one of the most widespread diseases in recent times, as it is also considered a very dangerous disease for people with chronic diseases and the elderly. In the second experiment, we worked with the Rapid Miner tool, where we applied the following algorithms DT, RF, KNN, NN, NB,

SVM with/without pre-processing. We used pre-processing to replace the missing value with the mean and detect outlier, where we obtained high values. In Table 3, we obtained values without pre-processing with the following measurements: precision, recall, accuracy, F1. The highest value without pre-processing reached 99.02, 98.33, 98.75, 98.67%, where these values are considered very high. But it is not the highest thing in this article with the RF algorithm. The value we obtained in this article is considered the highest, and the highest is with the pre-processing. The results were obtained in Table 3 with the NB algorithm, where all metrics reached 100%. This value is considered the highest in this article, and through this value, we have significantly improved chronic nephrotic and predicted a significant improvement in classification performance. It is significantly superior to the previous works, and this makes and assures us that this paper is successful and gave excellent results because chronic nephrotic is a nasty and difficult disease and a danger to humans. Our research will improve the classification performance and prove that our proposal is well and will predict chronic nephrotic and improve his performance.

In future work, we will work on other ways to find the best solutions, for example, clustering and association. We will greatly predict kidney disease and other diseases, our only work is to search for knowledge and to promote such diseases with this research that will make diseases more improved and more predictable.

References

[1] Asgarnezhad R. and Ali Mohsin Alhameedawi K., "MVO-Autism: An Effective Pre-treatment with High Performance for Improving Diagnosis of Autism Mellitus", Journal of Electrical and Computer Engineering Innovations (JECEI), 2021.

[2] Asgarnezhad R., Monadjemi S. A., and SoltanAghaei M., "A new hierarchy framework for feature engineering through multi- objective evolutionary algorithm in text classification", Concurrency and Computation: Practice and Experience, 2021. https://doi.org/10.1002/cpe.6594

[3] Asgarnezhad R., Monadjemi S. A., and SoltanAghaei M., "An application of MOGW optimization for feature selection in text classification", The Journal of Supercomputing, 2021, 77: 5806-5839.

[4] Dulhare U.N. and Ayesha M., "Extraction of action rules for chronic kidney disease using Naïve bayes classifier", in 2016 IEEE International Conference on Computational Intelligence and Computing Research (ICCIC), 2016: 1-5.

[5] Amirgaliyev Y., Shamiluulu S., and Serek A., "Analysis of chronic kidney disease dataset by applying machine learning methods", in 2018 IEEE 12th International Conference on Application of Information and Communication Technologies (AICT), 2018: 1-4.

[6] Bhutani H., Smith V., Rahbari-Oskoui F., Mittal A., Grantham J.J., Torres V.E., et al., "A comparison of ultrasound and magnetic resonance imaging shows that kidney length predicts chronic kidney disease in autosomal dominant polycystic kidney disease", Kidney international, 2015, 88: 146-151.

[7] Ani R., Sasi G., Sankar U.R., and Deepa O., "Decision support system for diagnosis and prediction of chronic renal failure using random subspace classification", in 2016 International Conference on Advances in Computing, Communications and Informatics (ICACCI), 2016: 1287-1292.

[8] Wickramasinghe M., Perera D., and Kahandawaarachchi K., "Dietary prediction for patients with Chronic Kidney Disease (CKD) by considering blood potassium level using machine learning algorithms", in 2017 IEEE Life Sciences Conference (LSC), 2017: 300-303.

[9] A. Al Imran, M. N. Amin, and F. T. Johora, "Classification of chronic kidney disease using logistic regression, feedforward neural network and wide & deep learning," in 2018 International Conference on Innovation in Engineering and Technology (ICIET), 2018: 1-6.

[10] Suresh C., Pani B.C., Swatisri C., Priya R., and Rohith R., "A Neural Network based Model for Predicting Chronic Kidney Diseases", in 2020 Second International Conference on Inventive Research in Computing Applications (ICIRCA), 2020: 157-162.

[11] E. M. Senan, M. H. Al-Adhaileh, F. W. Alsaade, T. H. Aldhyani, A. A. Alqarni, N. Alsharif, et al., "Diagnosis of Chronic Kidney Disease Using Effective Classification Algorithms and Recursive Feature Elimination Techniques," Journal of Healthcare Engineering, 2021.

[12] Khan B., Naseem R., Muhammad F., Abbas G., and Kim S., "An empirical evaluation of machine learning techniques for chronic kidney disease prophecy", IEEE Access, 2020, 8: 55012-55022.

[13] Alaiad A., Najadat H., Mohsen B., and Balhaf K., "Classification and association rule mining technique for predicting chronic kidney disease", Journal of Information & Knowledge Management, 2020, 19.

[14] Parthiban R., Usharani S., Saravanan D., Jayakumar D., Palani D.U., StalinDavid D.D., et al., "Prognosis of chronic kidney disease (CKD) using hybrid filter wrapper embedded feature selection method", European Journal of Molecular & Clinical Medicine, 2021, 7: 2511-2530.

[15] Nishat M.M., Dip R.R., Faisal F., Nasrullah S.M., Ahsan R., Shikder M.F., et al., "A Comprehensive Analysis on Detecting Chronic Kidney Disease by Employing Machine Learning Algorithms", EAI Endorsed Transactions on Pervasive Health and Technology, 2021, 18.

[16] Asgarnezhad R., Monadjemi A., and Soltanaghaei M., "NSE-PSO: Toward an Effective Model Using Optimization Algorithm and Sampling Methods for Text Classification", Journal of Electrical and Computer Engineering Innovations (JECEI), 2020, 8: 183-192. https://doi.org/10.22061/jecei.2020.7295.379

[17] Asgarnezhad R., Monadjemi A., and Soltanaghaei M., "A High-Performance Model based on Ensembles for Twitter

Sentiment Classification", Journal of Electrical and Computer Engineering Innovations (JECEI), 2020, 8: 41-52. https://doi.org/10.22061/jecei.2020.7100.357

[18] Asgarnezhad A., Monadjemi S., and Soltanaghaei M., "FAHPBEP: A fuzzy Analytic Hierarchy Process framework in text classification", Majlesi Journal of Electrical Engineering, 2020, 14. https://doi.org/10.29252/mjee.14.3.14

[19] Asgarnezhad R. and Monadjemi S. A., "NB VS. SVM: A contrastive study for sentiment classification on two text domains", Journal of applied intelligent systems & information sciences, 2021, 2(1): 1-12. https://doi.org/10.22034/jaisis.2021.279225.1025

[20] Asgarnezhad R. and Monadjemi S. A., "NSE: An effective model for investigating the role of pre-processing using ensembles in sentiment clssification", Journal of advances in computer research, 2022, 12(3): 27-41.