International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Improving the Performance of C4.5 Algorithm using Particle Swarm Optimization in Chronic Kidney Disease Diagnosis

***Muh. Masruri*[1]*, Edi Noersasongko*[2]*, Purwanto*[3]*, Casidi*[4]**

[1,2,3]*Universitas Dian Nuswantoro, Semarang, 50131*
[4] *Universitas An Nasher, Cirebon, 45611*

ABSTRACT

Chronic Kidney Disease (CKD) is a global health issue with a high mortality rate. Early detection is crucial to prevent further complications; however, traditional diagnostic methods still have accuracy limitations. Therefore, this study aims to improve CKD prediction accuracy by optimizing the C4.5 algorithm using Particle Swarm Optimization (PSO). This research utilizes the chronic kidney disease (CKD) dataset from the UCI repository, consisting of 400 patient records with 24 attributes. The model was developed using K-Fold cross-validation to ensure result reliability. Experiments were conducted by comparing the C4.5 algorithm before and after optimization with PSO. The results show that model accuracy increased from 96.00% to 98.06%, while the Area Under Curve (AUC) value improved from 0.965 to 0.985. These findings demonstrate that PSO effectively enhances the performance of the C4.5 algorithm in CKD classification, making it more effective for early diagnosis. This research contributes to the development of decision support systems in healthcare, which can be implemented in data-driven diagnostics to improve disease detection accuracy.

Keywords: Chronic Kidney Disease, C4.5, Particle Swarm Optimization, Prediction, Data Mining

1. Introduction

Disease Kidney Chronic (CKD) is one of the problem global health with level high morbidity and mortality. PGK often called as a "silent disease" because the symptoms No easy detected at stage early, so that Lots patient new realize the condition after reaching an advanced stage (Pratama & Widiartha, 2021). Causes The main PGK is hypertension and diabetes mellitus that are not controlled, which can cause damage kidney in a way progressive (Kalantar-Zadeh et al., 2021). According to Indonesian Ministry of Health report, PGK is one of the reason main deaths in developing countries and have impact significant economy to financing health national (Ministry of Health of the Republic of Indonesia, 2023). Therefore that, effort detection Early detection of PGK is very important for increase effectiveness treatment and prevention complications more carry on.

Prediction method based on data mining has Lots used in field health for increase accuracy of diagnosis of various diseases, including PGK (Amalia, 2018). The C4.5 algorithm is one of the technique popular Classification in medical data analysis Because his ability in handling categorical data as well as generate tree model decisions that can be made interpreted with good (Yunus et al., 2023). However, the method This Still own limitations in handle attribute numeric as well as prone to against overfitting when dealing with datasets with complex attributes (Pratama & Widiartha, 2021). For overcome problem Here, Particle Swarm Optimization (PSO) is used as technique optimization to improve selection feature in C4.5 algorithm, so that can increase accuracy prediction (Rostami et al., 2021).

Study This aiming For optimize accuracy C4.5 algorithm in predict PGK with implementing PSO as method optimization Features. Dataset used originate from UCI repository with a total of 400 patient data and 24 attributes (Pratama & Widiartha, 2021). Model validation was carried out use K-Fold Cross Validation technique for ensure reliability results prediction (Yunus et al., 2023). Research results show that combination C4.5 algorithm with PSO improves accuracy from 96.00% to 98.06% and the Area Under Curve (AUC) value from 0.965 to 0.985. The research This contribute in development system Supporter decisions in the field health, which can help power medical in perform a more detailed diagnosis fast and accurate (Roy et al., 2021).

PSO itself is an optimization algorithm inspired by the social behavior of a group of birds or fish in search of food. PSO works by modeling potential solutions as particles swimming in the solution space to find the best position based on its function. PSO has proven effective in finding optimal solutions in various classification and clustering problems because of its ability to explore the solution space efficiently (Kuehn, 2022; Tyastama et al., 2021). In the context of this research, PSO is used to optimize the process of discretizing numerical data before being classified using the C4.5 algorithm.

Based on the description above, and to broaden scientific insight, the author will apply a classification algorithm model with C4.5 and incorporates Particle Swarm Optimization (PSO) which is useful for increase optimization AUC and Accuray values. The dataset is taken from public access and there are already several researchers who use it from the UCI repository, named chronic kidney diseases (CKD), the total attributes in this dataset are 24 attributes, one attribute is the ckd and notckd class and the attributes are nominal and numeric.

1.1 Formulation of the problem

Classification methods such as C4.5, SVM and SVM-NN PSO can still be improvised to improve the optimization of the resulting accuracy value, with the combination of the C4.5 classification algorithm method and Particle Swarm Optimization (PSO) expected to be able to optimize the AUC and accuracy value.

1.2 Scope of Problem

This research uses problem limitations in the application of the algorithm:

1. The combination of C4.5 and C4.5- PSO algorithms is compared with KNN and KNN-PSO algorithms.
2. Evaluation of AUC and Accuracy values.
3. Rapid Miner and Python tools.

2. Methods

This study is a quantitative experimental study that aims to compare the performance of the C4.5 algorithm with and without Particle Swarm Optimization (PSO) optimization in predicting chronic kidney disease (CKD).

2.1 Method of collecting data

The purpose of this process is to collect data used in this study, the data was obtained by examining the records of patients with chronic kidney disease (CKD) obtained from the Machine Learning Database of California University, Irvine (UCI). The 400 datasets consist of 250 ckd and 150 notckd. With a total of 24 attributes and 1 class attribute.

2.2 Data Processing Method

In data processing, data analysis and testing were conducted using experimental methods using computers for the project development process, evaluation and problem solving, including data processing and calculation of the proposed model. In addition, this study also used the Cross Industry Standard Process for Data Mining (CRISP-DM) model which consists of six steps, namely business understanding, data understanding, data preparation, modeling, evaluation, and implementation.

1. **Business Understanding stage**

Based on a study of a total of 400 CKD patients obtained from the UCI Machine Learning Repository of Apollo Hospitals in India, patients should be screened for their disease so that it can be detected early to reduce the risk of positive CKD. Using a classification technique with a high level of prediction and accuracy can help solve this problem and make diagnostic results easier, faster and more accurate. To achieve high prediction and accuracy results in this study, the implementation of the C4.5 optimization algorithm was carried out with attribute selection using particle swarm optimization (PSO).

2. **Data Understanding Stage (Data Understanding)**

The data used is secondary data obtained from research results on patients with chronic kidney disease (CKD) or chronic kidney disease. This data comes from senior nephrologists at Apollo Hospitals in India and is publicly available through the UCI Machine Learning Repository. The data consists of 24 predictive attributes and 1 outcome attribute that can identify the status of ckd and not ckd patients.

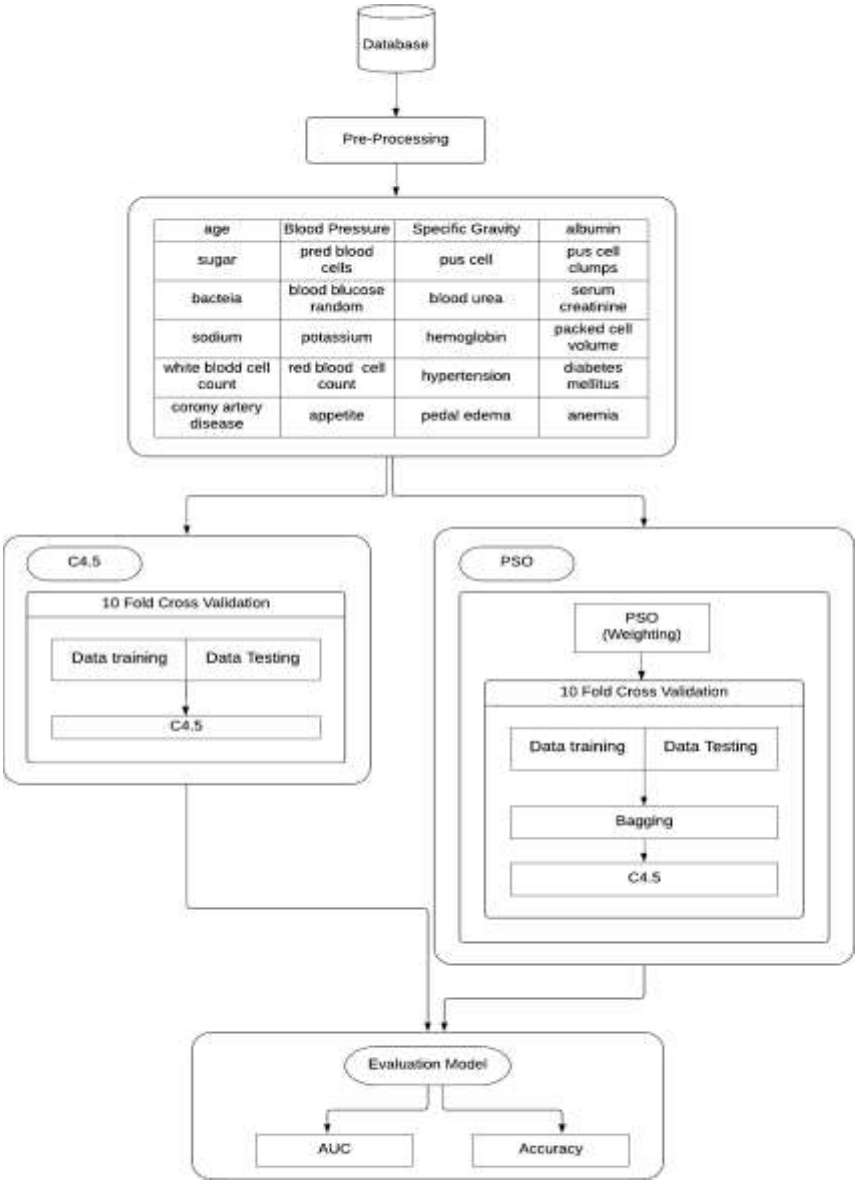
3. **Data Preparation Stage**

The data set collected in this study includes 400 data sets, both positive CKD syndrome patients and non-positive CKD patients, but the data still contains overlapping and anomalies or data inconsistencies. Therefore, this step is needed to obtain high-quality data and build a more effective and efficient model.

4. **Modeling Stage**

At this stage, the model is made with the C4.5 optimization algorithm and attribute selection using particle swarm optimization. Dataset processing is carried out to obtain selected variables using particle swarm optimization and bagging techniques are used to overcome noisy data due to the classification process using decision tree.

Then, the 10-fold cross-validation method is used to test the data and training data. The optimized attributes of these two data are used in the next step. Similar attributes are used to generate correct information. The training data is tested with the C4.5 algorithm to create a new method for the prediction process, while data validation will produce an estimation model measured by accuracy and AUC values shown in Figure 2.1.



age	Blood Pressure	Specific Gravity	albumin
sugar	pred blood cells	pus cell	pus cell clumps
bacteia	blood blucose random	blood urea	serum creatinine
sodium	potassium	hemoglobin	packed cell volume
white blodd cell count	red blood cell count	hypertension	diabetes mellitus
corony artery disease	appetite	pedal edema	anemia

Figure 1. Proposed model

The explanation for the framework in Figure 2.1 is:

1) Preprocessing is a technique used to change raw data.

a. Data Cleaning

At this stage, the data is cleaned through several processes such as filling in missing values, smoothing noisy data, and resolving any incoherences found.

b. Data Integration

In this phase, data with different representations is combined and any conflicts that arise are resolved.

c. Data Transformation

At this stage the data will be normalized and generalized, self-normalization.

d. Data Reduction

At this stage there are techniques for processing large amounts of data, where processing large amounts of data can complicate the data analysis process.

2) Using the C4.5 classification algorithm by dividing the dataset into training and testing data.

3) Added PSO optimization to process training and testing data to get performance results. More optimal accuracy.

4) Classification model results based on AUC and Accuracy values.

5. Evaluation Stage

In this phase, the resulting model is tested to obtain accurate model data. Evaluation and validation are carried out using the Confusion Matrix and ROC Curve methods. The testing mechanism for the method is carried out by evaluating the accuracy and AUC results of all experiments between the C4.5 algorithm and the C4.5 algorithm with PSO to compare the prediction results with other algorithms. The value and validation of the prediction model are considered the most optimal. The higher the accuracy score, the better the classification method and algorithm used.

6. Implementation Phase (Deployment)

In this phase, the resulting model is implemented to obtain accurate model data in predicting chronic kidney disease. Model implementation and model performance monitoring are carried out to ensure that the model works well.

3. Result and Discussion

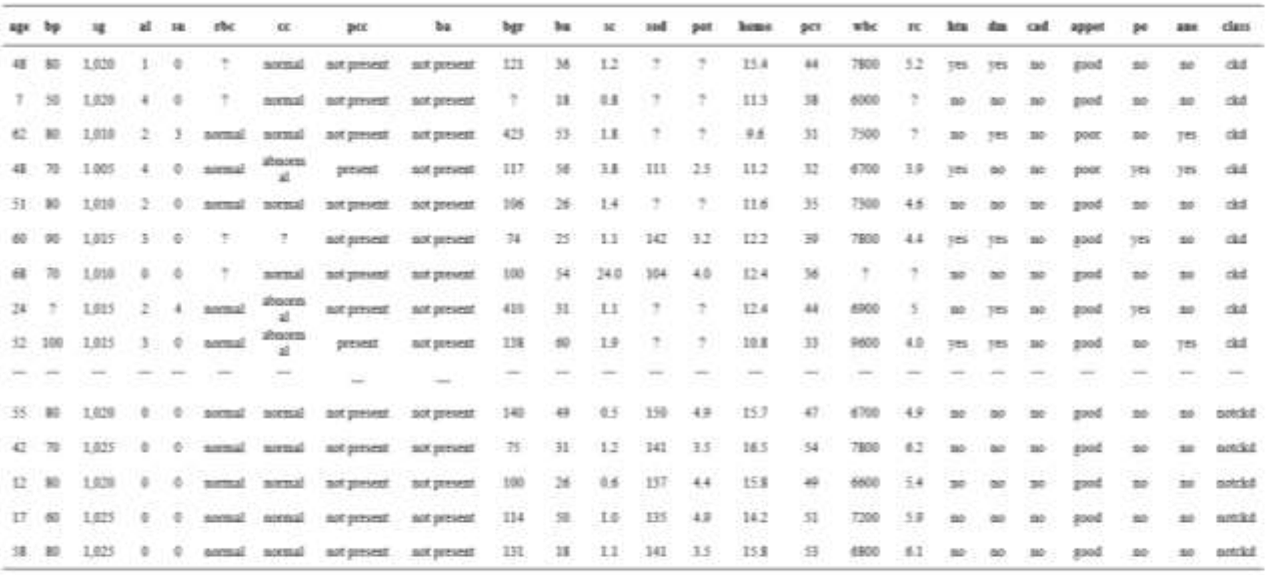
3.1 Dataset Analysis

This study aims to compare the selected C4.5 model with the C4.5 algorithm optimized by attribute selection using Particle Swarm Optimization (PSO). Also compared with other algorithms, in predicting chronic kidney disease (CKD). This test was conducted on a dataset obtained from https://archive.ics.uci.edu/dataset/336/chronic+kidney+disease.

This chronic kidney disease dataset contains 24 attributes and 1 class attribute and has 400 records consisting of 1 class **ckd** and **notckd** as well as nominal and numeric type attributes. The data set is displayed with the help of the RapidMiner tool as shown in Table 3.1.

Table 1

CKD Dataset



age	bp	sg	al	su	rbc	pc	pcc	ba	bgr	bu	sc	sod	pot	hemo	pcv	wbc	rc	htn	dm	cad	appet	pe	ane	class
48	80	1,020	1	0	?	normal	not present	not present	121	36	1.2	?	?	15.4	44	7800	5.2	yes	yes	no	good	no	no	ckd
7	50	1,020	4	0	?	normal	not present	not present	?	18	0.8	?	?	11.3	38	6000	?	no	no	no	good	no	no	ckd
62	80	1,010	2	3	normal	normal	not present	not present	423	53	1.8	?	?	9.6	31	7500	?	no	yes	no	poor	no	yes	ckd
48	70	1,005	4	0	normal	abnormal	present	not present	117	56	3.8	111	2.5	11.2	32	6700	3.9	yes	no	no	poor	yes	yes	ckd
51	80	1,010	2	0	normal	normal	not present	not present	106	26	1.4	?	?	11.6	35	7300	4.6	no	no	no	good	no	no	ckd
60	90	1,015	3	0	?	?	not present	not present	74	25	1.1	142	3.2	12.2	39	7800	4.4	yes	yes	no	good	yes	no	ckd
68	70	1,010	0	0	?	normal	not present	not present	100	54	24.0	104	4.0	12.4	36	?	?	no	no	no	good	no	no	ckd
24	?	1,015	2	4	normal	abnormal	not present	not present	410	31	1.1	?	?	12.4	44	6900	5	no	yes	no	good	yes	no	ckd
52	100	1,015	3	0	normal	abnormal	present	not present	138	60	1.9	?	?	10.8	33	9600	4.0	yes	yes	no	good	no	yes	ckd
—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
55	80	1,020	0	0	normal	normal	not present	not present	140	49	0.5	150	4.9	15.7	47	6700	4.9	no	no	no	good	no	no	notckd
42	70	1,025	0	0	normal	normal	not present	not present	75	31	1.2	141	3.5	16.5	54	7800	6.2	no	no	no	good	no	no	notckd
12	80	1,020	0	0	normal	normal	not present	not present	100	26	0.6	137	4.4	15.8	49	6600	5.4	no	no	no	good	no	no	notckd
17	60	1,025	0	0	normal	normal	not present	not present	114	50	1.0	135	4.9	14.2	51	7200	5.9	no	no	no	good	no	no	notckd
58	80	1,025	0	0	normal	normal	not present	not present	131	18	1.1	141	3.5	15.8	53	6800	6.1	no	no	no	good	no	no	notckd

Before being processed, the dataset still contains many empty values and there are several values with the character "?', in the attributes Age, Bp, Sg, for that, data pre-processing is required first before processing in rapidminer to obtain consistent data, the following are attributes that contain missing values and several characters.

3.2 Dataset Preprocessing

Before the data is processed, preprocessing is the initial step where the data is cleaned from values that are still inconsistent, such as noisy redundancy, by preprocessing the dataset first. The stages carried out are as follows with the help of the RapidMiner tool:

1) Prepare the ckd dataset then format the attributes and determine the data type according to its value, such as int, binominal, nominal, numeric, real and other data types.

2) Determine the label or class in the dataset, in this case the classification column becomes a label containing ckd and notckd with a ratio of ckd totaling 250, and notckd totaling 150 with a total of 400 records.

3) Missing Value data cleaning aims to clean up data that is not appropriate in the dataset, there are some data that contain missing values, each data is cleaned using the replace missing value operator in RapidMiner. The selected parameters are Attribute = All and default value average, this method aims to eliminate all inappropriate values in all attributes.

Table 2

Dataset attribute with missing value

No	Name	Type	Missing/character
1	Age	Integer	9
2	Mr.	Integer	12
3	Sg	Real	47
4	Al	Integer	46
5	Rbc	Polynomial	152
6	Pcv	Polynomial	?
7	Rc	Polynomial	?

4) Data Reduction to eliminate data that is difficult to assume by using the filter operator Examples by applying the condition class no_missing_attribute.

5) The Detect Outlier (Distances) operator is used to detect outliers in a dataset using a distance-based method and calculates the distance between each data point and other data points.

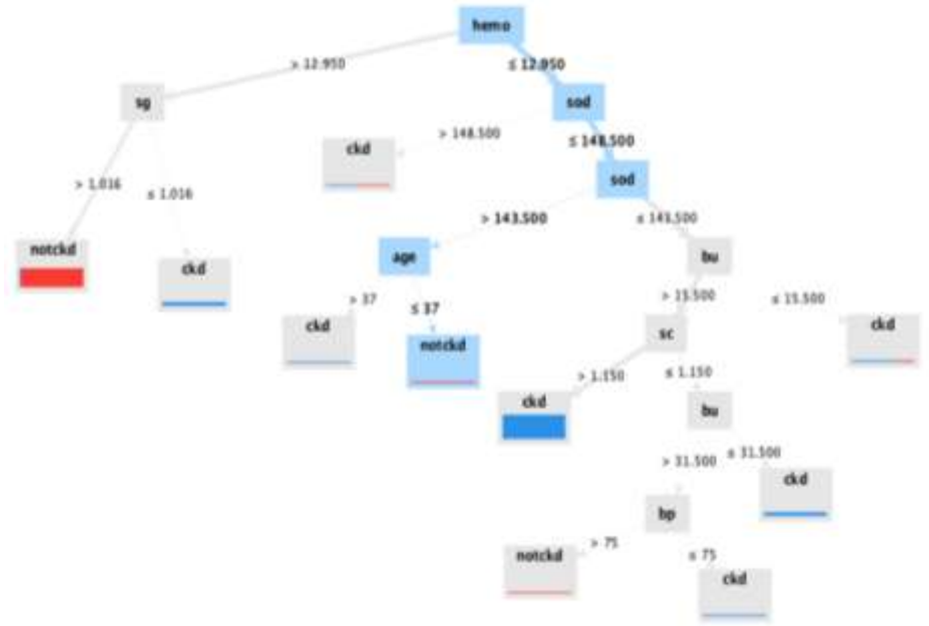
6) Duplicate data by using the remove duplicates operator with the all filter attribute parameter.

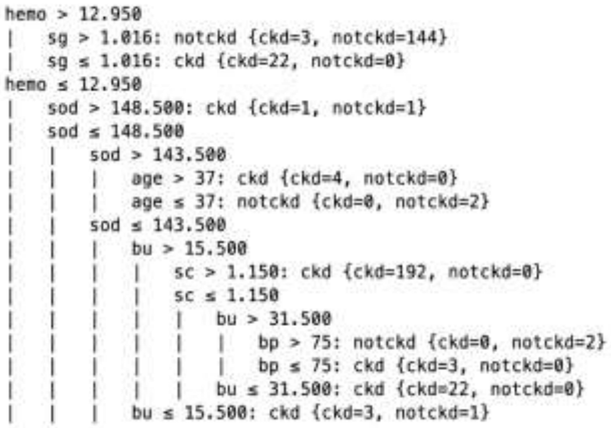
3.3 Validation with K-Fold Cross Validation

Before the data is predicted, it is first validated using k-fold cross-validation , meaning the data is divided into 10 parts randomly. The data testing and data training sections can also be called 10-fold cross-validation. In this study, testing was carried out using 10-fold cross-validation, namely by separating the data into two parts, using 400 data sets and 24 attributes with test and training data. This operation was carried out 3, 5, 10 times, and the final result was obtained, namely the accuracy value of 3 experiments, the average was formed.

3.4 Experiments with the C4.5 Algorithm

The C4.5 algorithm model is generated from a dataset consisting of 24 attributes presented in Table 4.1 to represent the attributes of chronic kidney disease (CKD) diagnosis. In general, the C4.5 decision tree construction algorithm takes one attribute as the root and creates branches for each value. The selection of attributes as the root is based on the highest validation score based on the existing attributes . After the data is processed to test the dataset on rapidminer using the cross validation operator by applying the parameter k = 10, and using the C4.5 algorithm produces an accuracy value of 97.25% and an AUC value of 0.970 and produces a decision tree where hemoglobin is the root with the highest gain value as can be seen in Figure 2.





```
hemo > 12.950
|   sg > 1.016: notckd {ckd=3, notckd=144}
|   sg ≤ 1.016: ckd {ckd=22, notckd=0}
hemo ≤ 12.950
|   sod > 148.500: ckd {ckd=1, notckd=1}
|   sod ≤ 148.500
|   |   sod > 143.500
|   |   |   age > 37: ckd {ckd=4, notckd=0}
|   |   |   age ≤ 37: notckd {ckd=0, notckd=2}
|   |   sod ≤ 143.500
|   |   |   bu > 15.500
|   |   |   |   sc > 1.150: ckd {ckd=192, notckd=0}
|   |   |   |   sc ≤ 1.150
|   |   |   |   |   bu > 31.500
|   |   |   |   |   |   bp > 75: notckd {ckd=0, notckd=2}
|   |   |   |   |   |   bp ≤ 75: ckd {ckd=3, notckd=0}
|   |   |   |   |   bu ≤ 31.500: ckd {ckd=22, notckd=0}
|   |   |   bu ≤ 15.500: ckd {ckd=3, notckd=1}
```

Figure 2. Decision Tree and Role Rules C4.5

Based on the generated rules, 10 rules were formed from the decision tree, where a total of 7 rules were for the CKD class and 3 rules were for the NOTCKD class in predicting chronic kidney disease.

3.5 Confusion Matrix Model Result

Confusion matrix model will create a matrix consisting of true positive or positive tuples and true negative or negative tuples . Then, enter the test data that has been provided in the confusion matrix so that the results are obtained as in table 3.

Table 3

Confusion Matrix Classification C4.5 Algorithm

Accuracy: 96.00%			
	True ckd	*True notckd*	*Class precision*
Pred ckd	242	8	96.80%
Pred notckd	8	142	94.67%
Class recall	96.80%	94.67%	

From 400 tuples , there are data on the number of true positive (TP) 242, false negative (FN) 8, false positive (FP) 8, and true negative (TN) 142. Based on these data, the accuracy, specificity, sensitivity, positive predictive value (PPV) and negative predictive value (NPV) values are calculated. Table 4.4. shows the predicted values of the calculation results.

Table 4

Values of Accuracy, Sensitivity, Specificity, PPV NPV

Prediction	**Mark**
Accuracy	0.960
Sensitivity	0.955
Specificity	0.966
PPV	0.968
NPV	0.946

3.6 Evaluation with ROC Curve

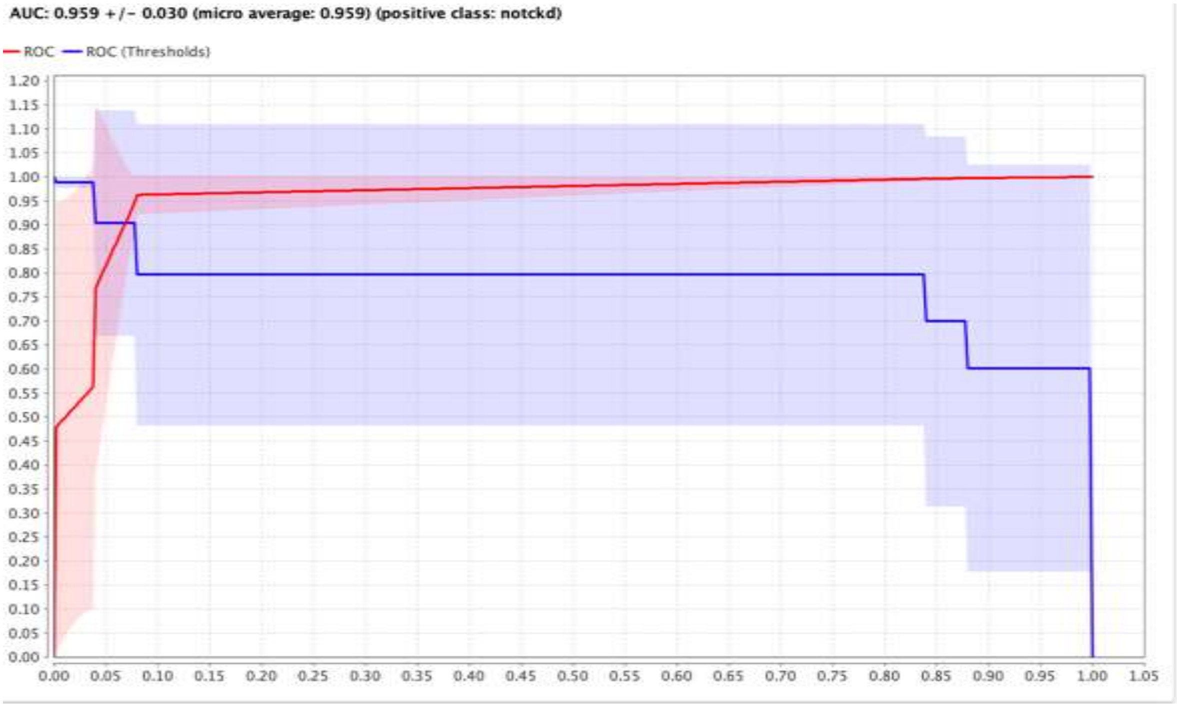


Figure 3 AUC value of C4.5 algorithm in the graph

Based on Figure 4.2, the AUC value obtained is 0.965 so that its accuracy can be classified into a very good classification.

3.7 Testing the C4.5 Model with PSO

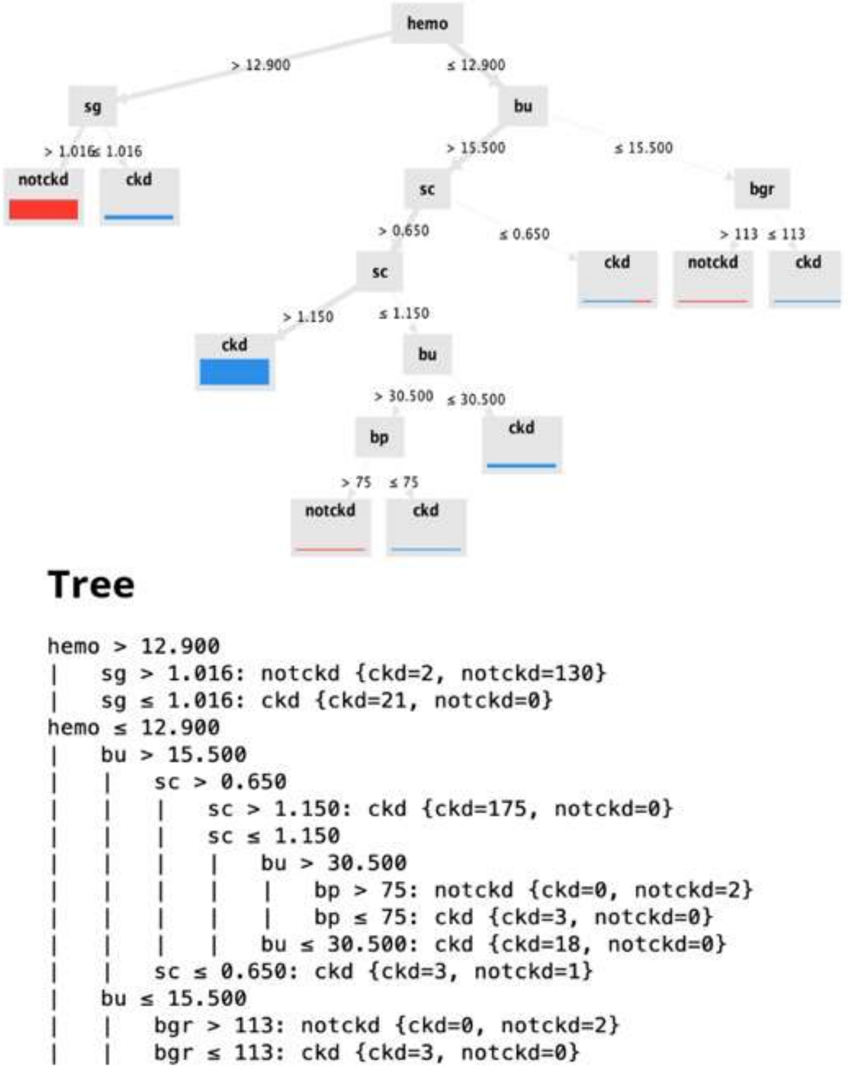
To perfect the calculation results of the C4.5 model, it will be added with a weight optimization operator to get better accuracy results, the steps are as follows, the data that has been preprocessed will then be divided into 2, namely for testing and training datasets with a division of 10% to 20% for testing data, with the split data operator, before the next process enters the weight optimization operator by adding cross validation using the parameter k = 10, then the C4.5 algorithm is added to process the dataset, the results of the combination of weight optimization operators (PSO) the accuracy value is 98.06% and the AUC value is 0.985, from the algorithm model The results of the attribute selection are as follows:

Table 5

0of attribute slection with PSO

Attributes	*Weight*
Age	1
Blood Pressure (BP)	0.182
Specific Gravity	0.489
Albumin	0.441
Sugar	0
Red Blood Cells	0
Cell Center	0.868
Pus Cell Clumps	0.873
Bacteria	0.275
Blood Blucose Random	0.897
Blood Urea	1
Serum Creatinine	0.301
Sodium	0.281
Potassium	0.917
Hemoglobin	0.939
Packed Cell Volume	0
White Blood Cell Count	0.390
Red Blood Cell Count	0.168
Hypertension	1
Diabetes Mellitus	0
Coronary Artery Disease	0.431
Appetite	1
Pedal Edema	0
Anemia	1

The decision tree produced by the C4.5 and PSO algorithms can be seen in Figure 5.



Tree

```
hemo > 12.900
|   sg > 1.016: notckd {ckd=2, notckd=130}
|   sg ≤ 1.016: ckd {ckd=21, notckd=0}
hemo ≤ 12.900
|   bu > 15.500
|   |   sc > 0.650
|   |   |   sc > 1.150: ckd {ckd=175, notckd=0}
|   |   |   sc ≤ 1.150
|   |   |   |   bu > 30.500
|   |   |   |   |   bp > 75: notckd {ckd=0, notckd=2}
|   |   |   |   |   bp ≤ 75: ckd {ckd=3, notckd=0}
|   |   |   |   bu ≤ 30.500: ckd {ckd=18, notckd=0}
|   |   sc ≤ 0.650: ckd {ckd=3, notckd=1}
|   bu ≤ 15.500
|   |   bgr > 113: notckd {ckd=0, notckd=2}
|   |   bgr ≤ 113: ckd {ckd=3, notckd=0}
```

Figure 4 Decision Tree and Role Rules with PSO

Based on the rules obtained, 9 rules were formed from the decision tree, which included a total of 6 CKD classes and 3 rules for the NOTCKD class in predicting chronic kidney disease

1) Evaluation of C4.5 and PSO Algorithm Models with Confusion Matrix

1C4.5 1 C4.5 and PSO algorithm in the graph

It mean that the AUC value obtained is 0.985 so that its accuracy can be classified into a very good classification in distinguishing between positive and negative classes. AUC ranges from 0 to 1, where a value of 0.5 indicates random performance, and a value of 1 indicates perfect classification . The Receiver Operating Characteristic (ROC) graph illustrates the trade-off between True Positive Rate (TPR) and False Positive Rate (FPR) at various thresholds. An increasing line indicates that the model is able to identify more true positives. positives when the FPR remains low. This graph also shows how the model performs at various thresholds. By looking at the points along the curve, we can determine the optimal threshold for a particular application, based on the need for a balance between TPR and FPR.

In other words, this graph shows that the analyzed classification model is very effective in separating the positive class from the negative class, with an AUC close to 1. This indicates that the model is reliable in applications that require accurate classification

3.8 Model Evaluation and Validation Analysis

From the test results, it can be concluded that there are 24 attributes that affect the test . The comparison between the test results of the C4.5 model without attribute selection and the C4.5 model with attribute selection using PSO is shown in Table 8.

Table 8

Testing of C4.5, KNN and C4.5-PSO, KNN-PSO Algorithms

Testing	***Accuracy***	***AUC***
C4.5	96.00%	0.965
KNN	86.00%	0.930
C4.5-PSO	98.06%	0.985
KNN-PSO	93.61%	0.971

Based on the test results by evaluating the confusion matrix and ROC curve, it is proven that the test carried out by optimizing the C4.5 algorithm with PSO has a higher accuracy score than using the C4.5 and KNN algorithms alone.

Score generated for the C4.5 algorithm model is 96.00 % with an AUC of 0.965, and the accuracy score for the C4.5 model with attribute selection using PSO is 98.06 % with an AUC of 0.985. From these scores, the accuracy difference is 0.206% while the AUC difference is 0.20. When compared with the KNN and KNN-PSO algorithms, the C4.5 and C4.5-PSO models show better accuracy values.

4. Conclusion

The accuracy value of the results of chronic kidney disease (CKD) research accuracy using the C4.5 classification algorithm is 96.00%, while using PSO attribute selection and bagging is 98.06%, so the difference in accuracy is 206%. The results of the ROC curve evaluation obtained the value of the C4.5 classification algorithm of 0.965, while the value of the C4.5 algorithm with PSO attribute selection and packaging is 0.985, so the difference in value is 0.20, when using the KNN classification algorithm the accuracy value is 85.50%, the AUC value is 0.930, while when using KNN with PSO attribute selection the value is 93.61%. and the AUC value is 0.971.

Based on these results, the Particle Swarm Optimization (PSO) optimization technique can be applied to select attributes in the C4.5 algorithm so that it can... obtained a higher accuracy value compared to the C4.5 algorithm method. Furthermore, if a comparison is made between the C4.5 algorithm and other classification algorithms such as KNN and KNN P article S warm Optimization (PSO), then it is obtained better accuracy value of C4.5 algorithm.

Based on the testing process of several algorithms and conclusions, the following are suggestions for further research to obtain more optimal results, including: Explore other optimization methods, such as Genetic Algorithm, Bee Colony, or other swarm-based optimization methods to compare their effectiveness with PSO, Consider using larger and more diverse datasets to improve model generalization and prediction accuracy, and perform comparisons with other classification algorithms such as Random Forest, XGBoost, or Neural Networks to evaluate relative performance in the context of kidney disease prediction.

References

Amalia, H. (2018). Comparison of SVM and NN Data Mining Methods for Chronic Kidney Disease Classification. PILAR Nusa Mandiri Journal , 14 (1), 1–6.

Debal, D.A., & Sitote, T.M. (2022). Chronic kidney disease prediction using machine learning techniques. Journal of Big Data , 9 (1), 109. https://doi.org/10.1186/s40537-022-00657-5

Hendry, MA, Gumanof, MI, Mulya, FR, & Meidelfi, D. (2021). Web-Based Kidney Disease Expert System. Journal of Information and Communication Technology , 11 (1), 80–88. http://dx.doi.org/10.35585%2Finspir.v11i1.2609

Hoar, MFE, Nababan, D., Risald, R., & Ulu, HH (2023). EXPERT SYSTEM FOR DIAGNOSIS OF KIDNEY DISEASE USING BACKWARD CHAINING METHOD. Jurnal Teknik Informatika Kaputama , 7 (2), 297–303. https://doi.org/10.59697/jtik.v7i2.92

Kalantar-Zadeh, K., Jafar, T.H., Nitsch, D., Neuen, B.L., & Perkovic, V. (2021). Chronic kidney disease. The Lancet , 398 (10302), 786–802. https://doi.org/10.1016/S0140-6736(21)00519-5

Kaur, G., & Sharma, A. (2017). Predict chronic kidney disease using data mining algorithms in Hadoop. 2017 International Conference on Inventive Computing and Informatics (ICICI) , 973–979. https://doi.org/10.1109/ICICI.2017.8365283

Kuehn, B. M. (2022). Targeted Efforts Needed to Combat Kidney Failure in Native Hawaiians, Pacific Islanders . 14 (9).

Performance Report for Semester I of 2023, Directorate General of P2P, Ministry of Health . (nd).

Pratama, IGAM, & Widiartha, IM (2021). IMPLEMENTATION OF C4.5 AND K-MEANS ALGORITHMS IN CHRONIC KIDNEY DISEASE DIAGNOSIS. Journal of Information and Computer Technology , 7 (4). https://doi.org/10.36002/jutik.v7i4.1535

Rostami, M., Berahmand, K., Nasiri, E., & Forouzandeh, S. (2021). Review of swarm intelligence-based feature selection methods. Engineering Applications of Artificial Intelligence , 100 , 104210. https://doi.org/10.1016/j.engappai.2021.104210

Roy, M.S., Ghosh, R., Goswami, D., & Karthik, R. (2021). Comparative Analysis of Machine Learning Methods to Detect Chronic Kidney Disease. Journal of Physics: Conference Series , 1911 (1), 012005. https://doi.org/10.1088/1742-6596/1911/1/012005

Sanmarchi, F., Fanconi, C., Golinelli, D., Gori, D., Hernandez-Boussard, T., & Capodici, A. (2023). Predict, diagnose, and treat chronic kidney disease with machine learning: A systematic literature review. Journal of Nephrology , 36 (4), 1101–1117. https://doi.org/10.1007/s40620-023-01573-4

Schröer, C., Kruse, F., & Gómez, J. M. (2021). A Systematic Literature Review on Applying CRISP-DM Process Model. Procedia Computer Science , 181 , 526–534. https://doi.org/10.1016/j.procs.2021.01.199

Tyastama, SA, Laksana, TG, & Arifa, AB (2021). Prediction of Chronic Kidney Disease Using Hybrid Backpropagation Neural Network with Particle Swarm Optimization. Journal of Innovation Information Technology and Application (JINITA) , 3 (1), 9–16. https://doi.org/10.35970/jinita.v3i1.588

Verbakel, J.Y., Steyerberg, E.W., Uno, H., De Cock, B., Wynants, L., Collins, G.S., & Van Calster, B. (2020). ROC curves for clinical prediction models part 1. ROC plots showed no added value above the AUC when evaluating the performance of clinical prediction models. Journal of Clinical Epidemiology , 126 , 207–216. https://doi.org/10.1016/j.jclinepi.2020.01.028

Verdiansah, V. (2016). Kidney Function Examination. Mirror of the World of Medicine , 43 (2), 148–154.

Xiao, W., Jing, L., Xu, Y., Zheng, S., Gan, Y., & Wen, C. (2021). Different Data Mining Approaches Based on Medical Text Data. Journal of Healthcare Engineering , 2021 , 1–11. https://doi.org/10.1155/2021/1285167

Yunus, M., Biddinika, MK, & Fadlil, A. (2023). Classification of Stunting in Children Using the C4.5 Algorithm. Online Journal of Informatics , 8 (1), 99–106. https://doi.org/10.15575/join.v8i1.1062