

Improvement of Kernel SVM to Enhance Accuracy in Chronic Kidney Disease

Ganda Wijaya

Universitas Nusa Mandiri, Indonesia
ganda.gws@nusamandiri.ac.id

Submitted : Oct 25, 2023 | **Accepted** : Nov 1, 2023 | **Published** : Jan 1, 2024

Abstract: Chronic Kidney Disease (CKD) is a very serious health problem, affecting millions of people throughout the world. Early diagnosis and accurate prediction of chronic failure disease is the main key to success in treatment. Problems that occur if not diagnosed early will endanger people's lives, that's why there is a lot of research on CKD. However, of the several methods that have been used, one approach to diagnosing this disease is to use the Support Vector Machine (SVM) algorithm with initial kernel selection and various kernels used for testing. However, the problem that occurs is that the accuracy of the SVM method with the kernels that have been tested is still less than optimal and can be improved further for better diagnosis. Therefore, the aim of the current research is to increase the accuracy of SVM kernels by combining the Particle Swarm Optimization (PSO) algorithm. The PSO method can help to optimize parameters and improve the kernel in SVM. The results of this study show that using the PSO method to improve the SVM kernel can significantly increase accuracy by 01.50% in CKD diagnosis compared with the conventional SVM approach. The research results have the potential to help medical practitioners in early diagnosis and better management of CKD, which in turn can improve the patient's prognosis and quality of life.

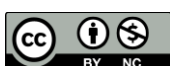
Keywords: Chronic kidney disease, Particle swarm optimization, SVM accuracy improvement, SVM kernel improvement.

INTRODUCTION

Chronic Kidney Disease (CKD) is a global health problem that affects millions of people throughout the world. CKD is characterized by a progressive decline in kidney function and can be life threatening if not diagnosed and managed properly. Early and accurate diagnosis and identification of chronic kidney disease is very crucial in efforts to prevent and treat this disease. Therefore, the development of sophisticated and accurate diagnostic methods is very important in the medical field, especially in CKD.

Many researchers have carried out predictions or classification of PGK data, namely by comparing the KNN, SVM and NB methods with experimental results showing that the highest method is KNN and the lowest method is SVM (Akben, 2018). After that, the research was carried out again using the KNN, J48, NN, NB and SVM methods which were assisted by using the FS method with the same results from the previous year that KNN had the highest accuracy and the lowest (Zeynu, 2018). In the same year 2018, PGK data testing was carried out again using the DT, LR, NB and ANN methods with the highest accuracy results being the NB and ANN methods (Murshid, Parvez, Fezal, Azaz, & Asif, 2019). After that, testing on SVM was carried out again by comparing the MLP, RBF, and SVM methods with the results on the data which were surprising: SVM was far superior to other methods with an accuracy of 87.00% (Rady & Anwar, 2019). From the tests that have been carried out, SVM already has superior accuracy compared to other methods. So the next researcher directly used SVM on PGK

*name of corresponding author



data by improving the performance of the kernels on SVM with Feature Selection, and the results of the test obtained results that were very close to perfect with an accuracy of 99.50% (Purwaningsih, 2022).

From several studies that have been carried out, in this research, we face a problem where the PGK data is almost close to perfect prediction, but the SVM method can be improved by using an optimization method, namely using Particle Swarm Optimization (PSO) to improve the SVM kernel. Where the SVM method is a method that has been proven effective in diagnosing disease (Riyanto, Hamid, & Ridwansyah, 2019). SVM is a machine learning method that has been widely used in various applications, including in the medical field. However, in the context of CKD, the challenge faced is how to improve the accuracy of SVM predictions for diagnosing CKD which has been carried out by previous research, and SVM diagnoses when the symptoms of the disease may not be clearly visible.

The aim of this research is to deal with this problem by increasing the accuracy of SVM in diagnosing CKD. By working to improve the SVM kernel parameters. The kernel type parameters used are dot kernel, polynomial kernel and radial kernel. By proposing the use of the PSO algorithm to improve this performance (Indah Ariyati, Ridwansyah, & Suhardjono, 2018). PSO is an optimization algorithm inspired by the behavior of flocks of birds which has been proven effective in finding optimal solutions in various optimization problems (I. Ariyati et al., 2020). Integrating PSO into SVM, with the hope of improving SVM performance in recognizing early signs of CKD (Ridwansyah, Ariyati, & Faizah, 2018).

If the PGK data is tested using the PSO method in improving the SVM kernels, it is hoped that it can improve the accuracy of the existing PGK data. So that later it can be used for the benefit of hospitals, especially the public, in knowing whether CKD is present or not in the body from an early age.

The main contribution of this research is to create an SVM kernel model that is more accurate in diagnosing CKD as a more reliable tool for CKD diagnosis, which can help medical personnel make more appropriate decisions in the treatment and care of CKD patients. Thus, it can improve the patient's quality of life and reduce the overall negative impact of CKD. Another contribution is highlighting the potential for using the PSO method in improving the performance of other Machine Learning algorithms for similar medical problems.

LITERATURE REVIEW

In the research on Chronic Kidney Disease (CKD), several researchers have conducted studies, including:

(Akben, 2018) conducted CKD research with 400 records using urine test results, blood tests, and medical history. Testing was done using the K-Means method for initial clustering, followed by classification using K-Nearest Neighbor (KNN), SVM, and NB. Specific gravity, albumin, sugar, and red blood cells together were the highest diagnostic attributes. The results for automatic CKD diagnosis were useful in detecting CKD to assist medical experts in decision-making.

(Rady & Anwar, 2019) made predictions about CKD with 25 attributes and 361 records, where 39 records were not used in the study. The research aimed to uncover and extract hidden information from clinical laboratory data. Data testing was done using the Probabilistic Neural Networks (PNN), MLP, Support Vector Machine (SVM), and Radial Basis Function (RBF) methods. Among these methods, PNN produced better classification and prediction results, which were valuable for improving accuracy in disease severity identification.

(Rezayi, Maghooli, & Saedi, 2021) used eight computation-based classification algorithms to classify and diagnose CKD using matrices. The methods included Random Forest (RF) with 100 trees, Deep Learning Network with 5 hidden layers, and NN with a training rate of 0.02 and 100 cycles. The results showed that the RF method outperformed other methods.

(Purwaningsih, 2022) conducted CKD research due to its potential to affect kidney health. The goal was to prevent and control CKD. The SVM method with various kernels was tested, and Feature Selection (FS) methods were used to improve SVM kernel accuracy. Among the various kernels, SVM with the radial kernel FS had the highest accuracy.

In summary, these research studies on CKD have used various attributes and different methods, as summarized in Table 1.

*name of corresponding author



Table 1. Chronic kidney disease research

Peneliti	Metode	Accuracy
(Rady & Anwar, 2019)	Probabilistic Neural Networks	96,70%
	Multilayer Perceptron	60,70%
	Support Vector Machine	87,00%
	Radial Basis Function	51,50%
(Akben, 2018)	K-means	86,00%
	KNN	78,75%
	SVM	73,75%
	Naïve Bayes	78,00%
(Purwaningsih, 2022)	SVM (dot)+FS	99,50%
	SVM (polynomial)+FS	95,50%
	SVM (radial)+FS	99,75%
(Rezayi et al., 2021)	Random Forest	99,09%
	Neural Network	96,52%
	Naive Bayes	96,14%
	Support Vector Machine	93,69%
	Deep Learning	98,04%
	KNN	71,56%
	DT	95,42%
	Multilayer Perceptron	96,22%
	KNN+FS	99,00%
	J48+ FS	98,75%
(Saha, Gourisaria, & Harshvardhan, 2022)	ANN+ FS	99,50%
	NB + FS	99,00%
	logistic regression	<99.08%
	support vector machine	<99.08%
	K-nearest neighbors	<99.08%
	naïve Bayes	<99.08%
	decision tree	<99.08%
	Random forest	99.08%
	support vector machine	<98.81%
	K-nearest neighbors	<98.81%
(Pramanik, Khare, & Gourisaria, 2021)	naïve Bayes	<98.81%
	decision tree	<98.81%
	Random forest	98.81%
	support vector machine	97.75%
	K-nearest neighbors	98.50%
	naïve Bayes	94.50%
	decision tree	96.75%
	JRip	96.00%
	AdaBoost	99.00%
	LogitBoost	99.75%
(Arif-Ul-Islam & Ripon, 2019)	J48	99.00%
	Ant-Miner	99.50%
	SVM	61.25%
	NB	56.50%

METHOD

The focus of this research is to build a prediction model for chronic kidney disease by comparing kernel performance in the SVM and PSO methods. This research aims to develop a model algorithm

*name of corresponding author



that is often used to predict chronic kidney disease. By comparing the accuracy between the SVM and PSO methods, this research also aims to analyze the impact of PSO on the performance of the SVM model by considering the kernel parameters implemented in the SVM, as well as presenting the accuracy test results obtained.

Data about CKD is obtained from public data called the UCI Repository with the data name chronic_kidney_disease, after which the data can be directly downloaded and used as research data. The data that has been downloaded can be processed into a datamining processing program with software called Rapid Miner, which in this test uses version 5.3. Data is processed using the SVM method.

SVM is a method that has minimal risk and a statistical structure theory by mapping high-dimensional features using existing kernel functions (Purnama, Nawawi, Rosyida, Ridwansyah, & Risandar, 2020). Some of the kernel functions used in SVM include radial (RBF), linear (point), and polynomial, with the C parameter having a range between 0 to 100. The best way to measure accuracy is to use a method called 10 k-fold cross validation, where the data in the test is re-divided in each iteration.

In conducting PGK research, a research flow is used which can be seen in Figure 1.

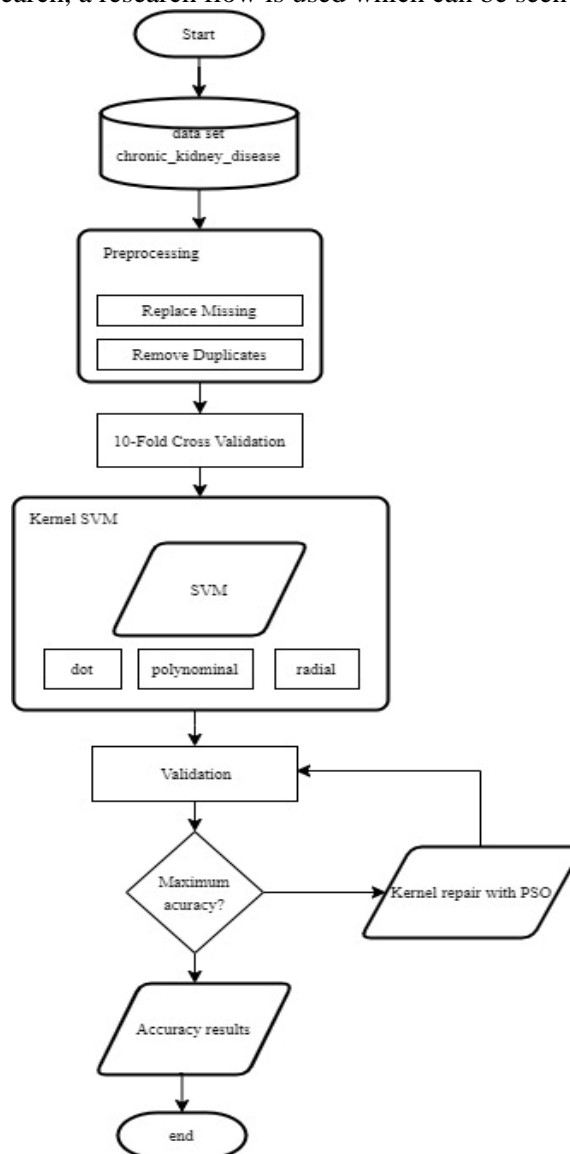


Fig 1. SVM kernel improvement research flow

From Figure 1 it can be explained that the research flow starts from collecting clinical data on CKD in the form of secondary data obtained from the UCI Repository public data which consists of data on 400 patients with and without CKD. The data consists of 25 clinically relevant features and includes

*name of corresponding author

predictor features. The data collected was carried out in the preprocessing stage using the Rapid Miner 5.3 tool. In the preprocessing stage the data is analyzed and selected to overcome filling in missing values with average values, deleting similar data and other processing. So that irrelevant features are removed. After that, a cross validation stage was carried out with K-fold to measure model performance. We measure evaluation metrics such as accuracy by conducting experiments on different types of SVM kernels including linear kernels, polynomial kernels, and radial kernels. To increase accuracy, we make improvements to the SVM kernel by optimizing the parameters in the PSO method, if the SVM kernel accuracy results are not optimal. After optimization, you will get maximum accuracy results.

RESULT

Based on the research flowchart depicted in Figure 1, the data used in this study was collected from secondary data sources originating from the UCI Repository. This data underwent experiments using the SVM method and an optimized SVM kernel improvement method using PSO. Prior to this, the data underwent a process of replacing missing data with mean values, and duplicate data that were not used were removed, ensuring data validity. The results of this process can be seen in Table 2, which displays the sample data used in this study.

Table 2. Sample research data on chronic kidney disease

age	bp	sg	albumin	serum	creatinine	phosphorus	hemoglobin	hematocrit	hemoglobin	creatinine	blood urea nitrogen	serum calcium	potassium	hemoglobin	hematocrit	creatinine	blood urea nitrogen	serum calcium	hemoglobin	hematocrit
69	70	1.0	4	3	1	0	1	1	21	96	6.3	12	3.	9.4	28	115	3.	1	1	
48	110	1.0	3	0	0	1	1	0	10	21	15.	12	5.	8.6	26	500	2.	1	0	
59	70	1.0	1	3	0	0	0	0	42	55	1.7	13	4.	12.	37	102	4.	1	1	
56	90	1.0	4	1	1	0	1	0	17	30	13.	12	6.	3.1	9	540	2.	1	1	
40	80	1.0	0	0	1	1	0	0	14	10	1.2	13	5.	15.	48	104	4.	0	0	
23	80	1.0	0	0	1	1	0	0	70	36	1.0	15	4.	17.	52	980	5.	0	0	
45	80	1.0	0	0	1	1	0	0	82	49	0.6	14	4.	15.	46	910	4.	0	0	
57	80	1.0	0	0	1	1	0	0	11	17	1.2	13	4.	15.	42	620	6.	0	0	
51	60	1.0	0	0	1	1	0	0	99	38	0.8	13	3.	13.	49	830	5.	0	0	

cad	appet	pe	ane	class
1	1	1	1	1
1	1	0	1	1
1	1	0	0	1
0	0	1	1	1
0	1	0	0	0
0	1	0	0	0
0	1	0	0	0
0	1	0	0	0
0	1	0	0	0

In table 2, data processing will be carried out by adopting the SVM method to improve the performance of the SVM kernel through the PSO approach. The main objective of this research is to evaluate the level of accuracy of applying various kernel methods to SVM and to determine whether optimizing the SVM kernel with the PSO approach is more effective in solving problems related to

*name of corresponding author



chronic kidney disease. This research was carried out in two experimental stages. The first stage involved testing PGK data using the SVM method with dot, polynomial and radial kernel comparisons. Meanwhile, the second stage involves applying optimization techniques using the PSO approach to SVM with the three types of kernels.

After conducting the testing of the chronic kidney disease dataset using various SVM kernels in the first stage, the research will yield results as shown in Table 3.

Table 3. PGK accuracy results with SVM kernel

Algorithm	TP	TN	FP	FN	Accuracy	AUC
SVM (Kernel dot)	243	150	0	7	98,25%	1,000
SVM (Kernel polynominal)	238	149	1	12	96,75%	0,986
SVM (Kernel radial)	250	118	0	32	92,00%	1,000

The accuracy results can be seen in table 3 that the SVM method with a dot kernel has the highest accuracy with a value of 98.25% and an AUC with a value of 1,000 and the SVM method with a radial kernel has a low accuracy but has the highest AUC with a value of 1,000. And the AUC value obtained from the SVM method will produce the curve in Figure 2

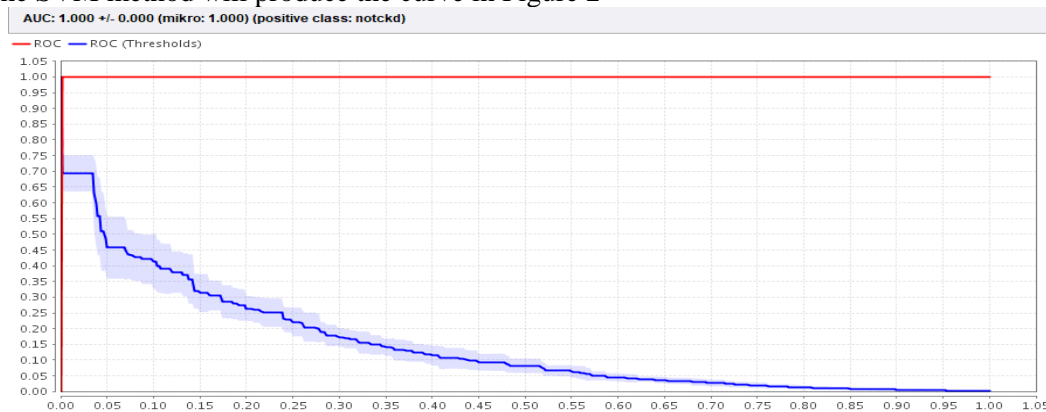


Fig 2. AUC SVM dot and radial kernels

In Figure 2, the AUC value obtained from the SVM method with dot kernels is very high and perfect with a value of 1,000, which is a very good curve. After the first stage is carried out, the second stage will be carried out by improving the SVM method kernel with PSO optimization.

Table 4. PGK accuracy results with PSO optimization SVM kernel improvements

Algoritma	TP	TN	FP	FN	Accuracy	AUC
SVM (Kernel dot) + PSO	249	150	0	1	99,75%	1,000
SVM (Kernel polynominal)+PSO	244	150	0	6	98,50%	0,988
SVM (Kernel radial)+PSO	249	150	0	1	99,75%	0,999

The accuracy results in Table 4 show that PSO optimization of the SVM method with the dot kernel has the same high accuracy as the radial kernel, both with a value of 99.75%. However, the AUC value for the dot kernel remains the highest at 1.000. The AUC value obtained from the PSO-optimized SVM method will result in a curve as shown in Figure 3.

*name of corresponding author

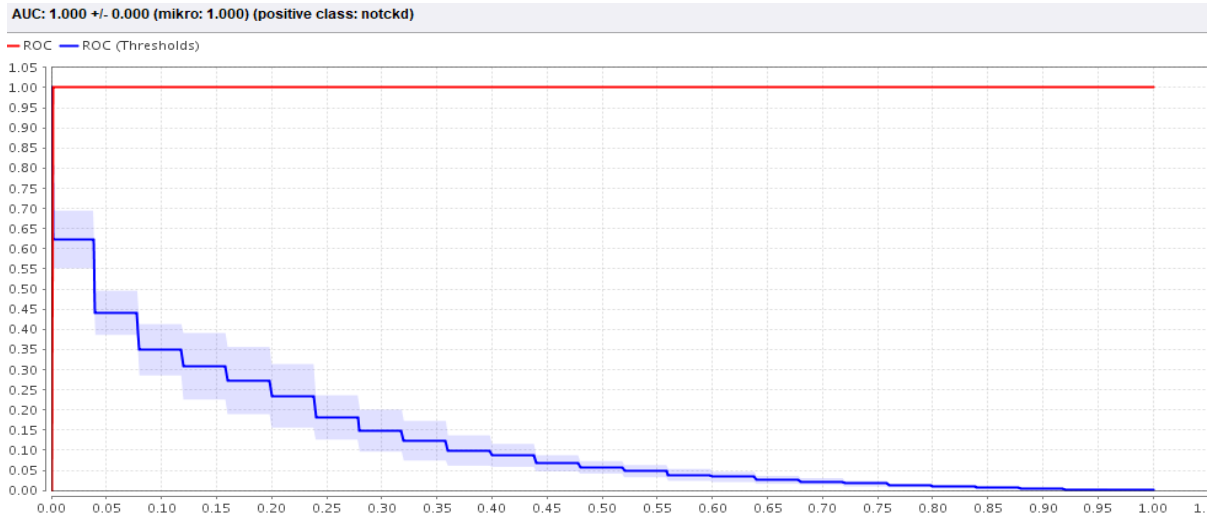


Fig 3. AUC SVM kernel dot with PSO improvements

Figure 2 and Figure 3 have the same curve of the dot SVM kernel for both those without kernel repair and those with kernel repair using the PSO method. However, the accuracy value is different, there is an increase in the kernel dot. In the curve in Figure 3, the AUC value of the SVM method with PSO improvements has not decreased and is still a very good curve.

The accuracy results for both methods, comparing SVM with or without PSO, can be observed in the graph in Figure 4.

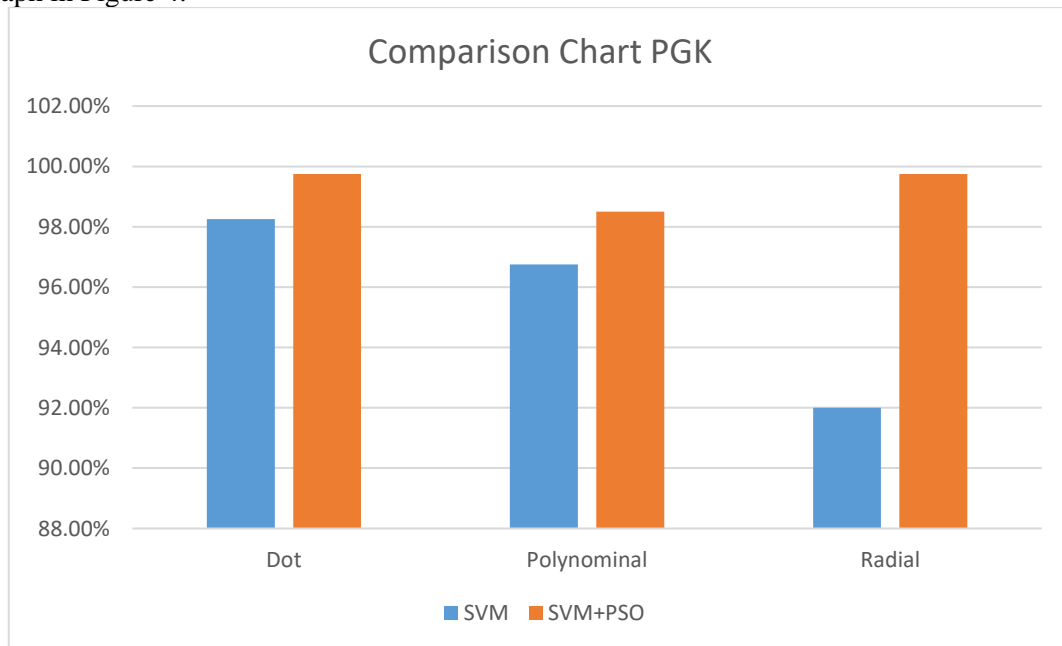


Fig 4. Comparison of SVM and SVM+PSO kernel accuracy values

Experimental results show that the use of improved SVM kernels, including customized custom kernels, results in significant improvements in CKD diagnosis accuracy. The SVM kernel with optimized parameters is significantly better than previous models. It can be seen in Figure 4. Where there is an increase in the PSO method for SVM method kernels compared to the SVM method without the PSO method. These results state that the PSO method can improve SVM kernels and increase SVM accuracy.

DISCUSSIONS

*name of corresponding author



The PGK data used has been carried out in the data preprocessing stage by deleting duplicate data and replacing missing data originating from the average value of each attribute and by carrying out testing methods on SVM method kernels and producing high accuracy on dot kernels with the value was 98.25% and the data was retested by improving the kernels on the SVM using PSO and the accuracy results obtained were 99.75% with an AUC value of 1,000.

The results of this experiment reveal a significant increase in accuracy values in the diagnosis of Chronic Kidney Disease (CKD) using the Particle Swarm Optimization (PSO) method to optimize the Support Vector Machine (SVM) kernel. These results provide strong support for the effectiveness of PSO to improve SVM performance in classifying CKD data. So that these results can provide a more precise diagnosis and enable better early treatment and avoidance of serious complications. Therefore, the results of this study have great potential to provide significant benefits to health professionals working in patient management.

One of the main contributions of this research is the use of a custom kernel adapted to the characteristics of PGK data. This kernel allows SVM to better describe complex relationships in data. Increasing accuracy in the diagnosis of CKD is an important step in clinical practice

CONCLUSION

In conclusion, this study succeeded in showing that the use of the PSO method to improve the SVM kernel is an effective approach to increase accuracy in CKD diagnosis. The results have positive implications in clinical practice and have the potential to help in better management of patients with CKD. After testing the CKD data, the results obtained can help the public to identify CKD disease early and can avoid death and reduce the death rate. With research limitations only for people who do not know about CKD in themselves.

This research provides a strong foundation for future research in this area. Although this research has provided promising results, there are still several future research opportunities that can be pursued. Some of these include the use of larger and more diverse datasets, the application of other optimization methods that may be more efficient, as well as further exploration of other factors that may influence CKD predictions.

Thus, this research makes a significant contribution in efforts to improve early diagnosis of CKD using the kernel SVM approach enhanced with the PSO method. This increase in accuracy has great potential to help health professionals manage and care for CKD patients more effectively

REFERENCES

- Akben, S. B. (2018). Early Stage Chronic Kidney Disease Diagnosis by Applying Data Mining Methods to Urinalysis, Blood Analysis and Disease History. *Irbm*, 39(5), 353–358. <https://doi.org/10.1016/j.irbm.2018.09.004>
- Alaiad, A., Najadat, H., Mohsen, B., & Balhaf, K. (2020). Classification and Association Rule Mining Technique for Predicting Chronic Kidney Disease. *Journal of Information and Knowledge Management*, 19(1). <https://doi.org/10.1142/S0219649220400158>
- Arif-UI-Islam, & Ripon, S. H. (2019). Rule Induction and Prediction of Chronic Kidney Disease Using Boosting Classifiers, Ant-Miner and J48 Decision Tree. *2nd International Conference on Electrical, Computer and Communication Engineering, ECCE 2019*, 1–6. <https://doi.org/10.1109/ECACE.2019.8679388>
- Ariyati, I., Rosyida, S., Ramanda, K., Riyanto, V., Faizah, S., & Ridwansyah. (2020). Optimization of the Decision Tree Algorithm Used Particle Swarm Optimization in the Selection of Digital Payments. *Journal of Physics: Conference Series*, 1641(1). <https://doi.org/10.1088/1742-6596/1641/1/012090>
- Ariyati, Indah, Ridwansyah, & Suhardjono. (2018). Implementasi Particle Swarm Optimization untuk Optimalisasi Data Mining Dalam Evaluasi Kinerja Asisten Dosen. *JIKO (Jurnal Informatika Dan Komputer) STMIK AKAKOM*, 3(2), 70–75.
- Murshid, G., Parvez, T., Fezal, N., Azaz, L., & Asif, M. (2019). Data Mining Techniques to Predict Chronic Kidney Disease. *International Journal of Scientific Research in Computer Science, Engineering and Information Technology*, 5(2).

*name of corresponding author



- <https://doi.org/https://doi.org/10.32628/CSEIT1952331>
- Pramanik, R., Khare, S., & Gourisaria, M. K. (2021). Inferring the Occurrence of Chronic Kidney Failure: A Data Mining Solution. *Proceedings of Second Doctoral Symposium on Computational Intelligence*.
- Purnama, J. J., Nawawi, H. M., Rosyida, S., Ridwansyah, & Risandar. (2020). Klasifikasi Mahasiswa Her Berbasis Algoritma Svm Dan Decision Classification of Her Students Based on Svm Algorithm and. *Jurnal Teknologi Informasi Dan Ilmu Komputer*, 7(6), 1253–1260. <https://doi.org/10.25126/jtiik.202073080>
- Purwaningsih, E. (2022). Improving the Performance of Support Vector Machine With Forward Selection for Prediction of Chronic Kidney Disease. *JITK (Jurnal Ilmu Pengetahuan Dan Teknologi Komputer)*, 8(1), 18–24. <https://doi.org/10.33480/jitk.v8i1.3327>
- Rady, E. H. A., & Anwar, A. S. (2019). Prediction of kidney disease stages using data mining algorithms. *Informatics in Medicine Unlocked*, 15(December 2018), 100178. <https://doi.org/10.1016/j.imu.2019.100178>
- Rezayi, S., Maghooli, K., & Saeedi, S. (2021). Applying Data Mining Approaches for Chronic Kidney Disease Diagnosis. *International Journal of INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING*.
- Ridwansyah, Ariyati, I., & Faizah, S. (2018). PARTICLE SWARM OPTIMIZATION BERBASIS CO-EVOLUSIONER DALAM EVALUASI KINERJA ASISTEN DOSEN. *Jurnal SAINTEKOM*, 9(2), 166–177. <https://doi.org/https://doi.org/10.33020/saintekom.v9i2.96>
- Riyanto, V., Hamid, A., & Ridwansyah. (2019). Prediction of Student Graduation Time Using the Best Algorithm. *Indonesian Journal of Artificial Intelligence and Data Mining*, 2(2), 1–9. <https://doi.org/http://dx.doi.org/10.24014/ijaidm.v2i1.6424>
- Saha, I., Gourisaria, M. K., & Harshvardhan, G. M. (2022). Classification System for Prediction of Chronic Kidney Disease Using Data Mining Techniques. *Advances in Data and Information Sciences*.
- Zeynu, S. (2018). Prediction of Chronic Kidney Disease Using Data Mining Feature Selection and Ensemble Method. *WSEAS TRANSACTIONS on INFORMATION SCIENCE and APPLICATION*, 15, 168–176.

*name of corresponding author



This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.