VOL. 8. NO. 1 AUGUST 2022
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480 /jitk.v8i1.3327

JITK (JURNAL ILMU PENGETAHUAN DAN TEKNOLOGI KOMPUTER)

IMPROVING THE PERFORMANCE OF SUPPORT VECTOR MACHINE WITH FORWARD SELECTION FOR PREDICTION OF CHRONIC KIDNEY DISEASE

Esty Purwaningsih

Sistem Informasi
Universitas Bina Sarana Informatika
www.bsi.ac.id
esty.epw@bsi.ac.id

Abstract—*Chronic kidney disease is a disorder that affects the kidneys and arises due to various factors, usually occurs slowly and is chronic. For prevention and control, proper treatment is needed, so that detection of this disease can play a very important role. This study aims to determine the level of accuracy in predicting chronic kidney disease through SVM based on forward selection and to determine the performance of Feature Selection which is applied to the SVM method in solving problems in chronic kidney disease. The test uses 10 k-fold cross validation by dividing the training and testing data from each experiment. In this study, experiments were carried out on the SVM method using various kernels and it was seen that SVM with the dot kernel was 98.50% with AUC 1,000 which was superior to the polynominal kernel and RBF. However, when the experiment was carried out again by applying FS to SVM, it was found that SVM+FS with the RBF kernel outperformed the other kernels by 99.75% with AUC 1,000. So it can be concluded that the Forward Selection on SVM has succeeded in improving its performance, especially in this case, namely the prediction of chronic kidney disease.*

Keywords: *chronic kidney, forward selection, SVM, features*

Intisari—Penyakit ginjal kronis merupakan kelainan yang mengenai organ ginjal dan timbul akibat berbagai faktor, biasanya timbul secara perlahan dan sifatnyamenahun. Untuk pencegahan dan penanggulangannya, diperlukan penanganan yang tepat, sehingga deteksi pada penyakit ini dapat berperan sangat penting. Penelitian ini bertujuan untuk mengetahui tingkat akurasi pada prediksi penyakit ginjal kronis melalui SVM berbasis forward selection serta mengetahui kinerja Feature Selection yang diterapkan kedalam metode SVM dalam memecahkan masalah pada penyakit ginjal kronis. Pengujian menggunakan 10 k-fold cross validation dengan membagi data training maupun testing dari setiap eksperimen. Pada penelitian ini dilakukan percobaan pada metode SVM dengan menggunakan berbagai macam kernel dan terlihat SVM dengan kernel dot yaitu 98,50% dengan AUC 1,000 lebih unggul dari pada kernel polynominal dan RBF. Namun ketika dilakukan percobaan kembali dengan menerapkan FS pada SVM, didapat bahwa SVM+FS dengan kernel RBF mengungguli dari kernel lainnya yaitu sebesar 99,75% dengan AUC 1,000. Sehingga dapat disimpulkan bahwa Forward Selection pada SVM berhasil meningkatkan kinerjanya khususnya pada kasus ini yaitu prediksi penyakit ginjal kronis.

Kata Kunci : ginjal kronis, forward_selection, SVM, fitur.

INTRODUCTION

A vital organ in the body that looks like a legume is located behind the stomach or abdomen, functions as an excretory organ, it is located on the right side and the left side of the spine is a vital organ of the kidney [1].

Kidney function will decrease if it is caused by tissue damage to the kidneys, medically described a significant level of decline in 3 (three) months and even more is called chronic kidney failure [2].

Problems that occur in terms of public health by looking at the prevalence coupled with the increasing incidence of kidney failure and with a poor prognosis with a very large level of expenditure in the world are included in chronic kidney disease or also called CKD, namely Chronic Kidney Diseases. From the results of the analysis, the increasing number of elderly people plus the presence of diseases such as diabetes and hypertension also triggers an increase in the prevalence of kidney failure [3] [4].

Accredited Rank 3 (Sinta 3) based on the Decree of the Dirjen Penguatan RisBang Kemenristekdikti No.28/E/KPT/2019, September 26, 2019. Published by LPPM Universitas Nusa Mandiri

VOL. 8. NO. 1 AUGUST 2022
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v8i1.3327

From the results of the World Health Organization (WHO) report, diseases that are included in the top 10 (ten) diseases in the world that cause death, especially in high-income countries are chronic kidney disease [5].

Riskesdas Results [6], also shows an increase in prevalence rates with increasing age and prevalence in men, which is shown in Figure 1 as follows:

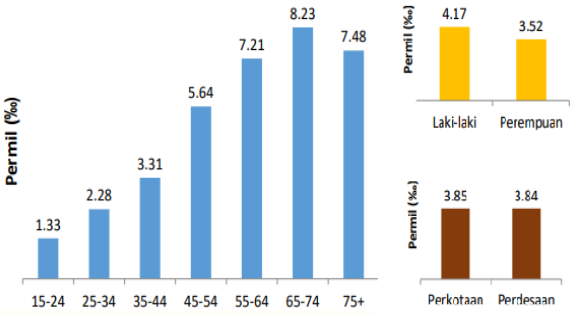

Figure 1. Prevalence (Permil) of 'Chronic Kidney' Disease Based on Doctor's Diagnosis 'At Age 15 Years According to Characteristics

To be able to prevent and overcome this chronic kidney disease, proper treatment is needed, because the number of deaths is expected to continue to increase over time. So the detection of this disease can play a very important role.

There are several studies conducted using machine learning techniques that aim to solve cases of chronic kidney disease, including using the Artificial Neural Network and Support Vector Machine (SVM) methods [7], where in this study aims to assist in the prevention of Chronic Kidney Disease (CKD) by utilizing machine learning techniques for early diagnosis. The classification techniques used in this research include Artificial Neural Network (ANN) and Support Vector Machine (SVM) methods. The empirical results of the experiment show that ANN performs better than SVM, with an accuracy of 99.75% and 97.75% respectively.

Furthermore, in the K-Nearest Neighbor (KNN), SVM and Soft Independent Modeling of Class Analogy (SIMCA) methods [8], The authors labeled patients as CKD or non-CKD with greater than 93% accuracy. SVM processes noise disturbance from composite datasets better and has the highest accuracy in predicting CKD compared to the other two models because the accuracy reaches 99%.

Naive Bayes method, LDA, KNN, Decision tree, Forest Random [9], it was found that K-Nearest Neighbor (KNN) with random subspace classifier has a prediction accuracy of 94%. It is recommended to use kernel or neural based classifier instead of KNN with random subspace classifier.

On the C4.5 Algorithm [10], The accuracy of the algorithm was found to be 98.25%. Artificial Neural Network (ANN) Method, Naive Bayes and Decision Tree [11], This system was implemented using data from Prince Hamzah Hospital in Jordan. The performance of machine learning techniques is analyzed by calculating sensitivity, specificity, and accuracy. Decision Tree, is the most accurate technique in predicting chronic renal failure (CRF). It is recommended to apply more machine learning techniques, including Support Vector Machine (SVM), to predict CRF.

SVM, Decision tree, Naive Bayes and KNN [12], Different machine learning techniques were applied to the data set, and their accuracy was calculated using the root mean square error and the mean absolute error and by plotting the receiver operating characteristic curve. The Decision Tree algorithm was found to be the most accurate with an accuracy of 99.75%, while the second accurate SVM with an accuracy of 97.75%. So it is recommended to implement ANN and fuzzy logic to the dataset and analyze the accuracy.

KNN, SVM and Naive Bayes [13], The accuracy rate of 97.8% is the highest level, with the proposed method, namely the urine test attribute. Accuracy increases in the age group of 35 years and over using the same attributes. Different combinations of dataset attributes result in different accuracy rates ranging from 83.75% to 97.8%.

SVM [14], The study compares the results of implementing SVM on all features and implementing SVM on Feature Selection Methods. Features are selected using wrapping and filter techniques.

PSO based C4.5 algorithm with bagging technique [5], This study was conducted to find out which model gives the best results in detecting chronic kidney disease. The method used in this case study is able to select attributes so that it can increase the accuracy value better with a result of 99.70% compared to the C4.5 algorithm model which produces an accuracy of 91.72%.

So many researchers have conducted research on chronic kidney disease, but in this study applied a Support Vector Machine (SVM) classification model based on forward selection. Where the SVM model has the advantage of a high level of accuracy and can overcome classification problems in both linear and nonlinear regression. However, according to Ilhan and Tezel in [15] of the advantages of SVM, it also has drawbacks where it is difficult to choose features for optimal input.

So that forward selection plays a role in selecting features that can improve machine learning performance, especially in predicting chronic

kidney disease. From the advantages of each model and its algorithm, this study aims to determine the level of accuracy in predicting chronic kidney disease through SVM based on forward selection and to determine whether the performance of forward selection is more effectively applied in the SVM method to solve problems in chronic kidney disease.

MATERIALS AND METHODS

Data mining has an interest in society and the world, especially in the field of information systems, because a number of big data can create useful information according to needs and desires. [16].

According to witten [17] that in data mining there are several applications that focus on forecasting which predicts an event in the future in a new situation in which there is new data that comes from the history of data that has happened before.

The focus of this study is to build a predictive model for chronic kidney disease by comparing the performance of the SVM method with the forward selection algorithm. The purpose of this study is to develop an algorithm training model that is used to predict chronic kidney disease, to compare the performance value of accuracy between the SVM method and forward selection, analyze the effect of kernel parameter values on model performance, and feature subsets from the results of the forward selection feature selection.

The training process of the SVM method is first carried out with normalization [18]. Normalization used in the form of minimum and maximum values of normalization. minimum and maximum (Min-max) normalization is done by changing the original data into linear data with a value range of +1 (plus 1) and -1 (minus 1).

The data is processed by being entered into the SVM training system. Several SVM kernel functions are used including radial (RBF), dot (linear), and polynomial with parameter C between values of 0.0-100. The best test uses the method of calculating accuracy through a process step called 10 k-fold cross validation by dividing the training and testing data from each experiment.

This research was conducted in several stages, as shown in the following figure:

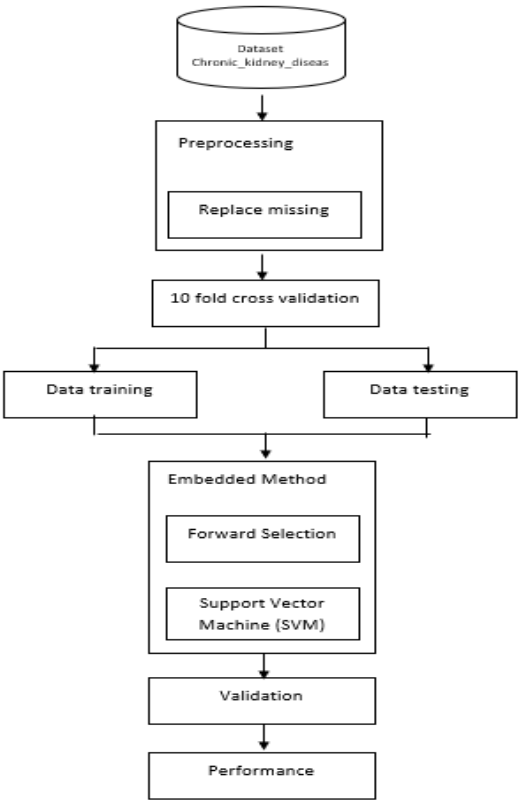

Figure 2. Research model schema

The dataset used in this study is chronic_kidney_disease data which was downloaded from the UCI machine learning repository on the page https://archive.ics.uci.edu/ml/machine-learning-databases/00336/.

In the chronic_kidney_disease (CDK) dataset or in other words chronic kidney disease, there are 25 attributes, including 24 free attributes and 1 label. The 24 free attributes consist of Age, Blood Pressure (mmHg), Specific Gravity, Albumin, Sugar, Red Blood Cells, Pus Cell, Pus Cell Clumps, Bacteria, Blood Glucose Random (mgs/d), Blood Urea (mgs/d). dl), Serum Creatinine (mgs/dl), Sodium (mEq/L), Potassium (mEq/L), Hemoglobin (gms), Packed Cell Volume, White Blood Cell Count (cells/cumm), Red Blood Cell Count (millions / cmm), Hypertension, Diabetes Mellitus, Corony Artery Disease, Appetiete Pedal Edema and Anemia. While the attribute label consists of 1 = CDK and 2 = No CDK.

The number of records in the chronic_kidney_disease dataset is 400 records. Can be seen in Table 1.

Accredited Rank 3 (Sinta 3) based on the Decree of the Dirjen Penguatan RisBang Kemenristekdikti No.28/E/KPT/2019, September 26, 2019. Published by LPPM Universitas Nusa Mandiri

VOL. 8. NO. 1 AUGUST 2022
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v8i1.3327

Table 1. Sample Dataset (before preprocessing)

age	bp	sg	al	su	rbc	pc	pcc
48	80	1.020	1	0	?	1	2
7	50	1.020	4	0	?	1	2
62	80	1.010	2	3	1	1	2
48	70	1.005	4	0	1	2	1
51	80	1.010	2	0	1	1	2
60	90	1.015	3	0	?	?	2
68	70	1.010	0	0	?	1	2
24	?	1.015	2	4	1	2	2
52	100	1.015	3	0	1	2	1
53	90	1.020	2	0	2	2	1
50	60	1.010	2	4	?	2	1
63	70	1.010	3	0	2	2	1

ba	bgr	bu	sc	sod	pot	hemo	pcv
2	121	36	1.2	?	?	15.4	44
2	?	18	0.8	?	?	11.3	38
2	423	53	1.8	?	?	9.6	31
2	117	56	3.8	111	2.5	11.2	32
2	106	26	1.4	?	?	11.6	35
2	74	25	1.1	142	3.2	12.2	39
2	100	54	24.0	104	4.0	12.4	36
2	410	31	1.1	?	?	12.4	44
2	138	60	1.9	?	?	10.8	33
2	70	107	7.2	114	3.7	9.5	29
2	490	55	4.0	?	?	9.4	28
2	380	60	2.7	131	4.2	10.8	32

wbcc	rbcc	htn	dm	cad	appet	pe	ane	class
7800	5.2	1	1	2	1	2	2	1
6000	?	2	2	2	1	2	2	1
7500	?	2	1	2	2	2	1	1
6700	3.9	1	2	2	2	1	1	1
7300	4.6	2	2	2	1	2	2	1
7800	4.4	1	1	2	1	1	2	1
?	?	2	2	2	1	2	2	1
6900	5	2	1	2	1	1	2	1
9600	4.0	1	1	2	1	2	1	1
12100	3.7	1	1	2	2	2	1	1
?	?	1	1	2	1	2	1	1
4500	3.8	1	1	2	2	1	2	1

To make the data valid and of good quality, initial data processing or also called **preprocessing** is carried out, namely identifying data and cleaning up problematic data by replacing missing. As for changing records on attributes rbc (normal to 1, abnormal to 2), pc (normal to 1, abnormal to 2), pcc (present to 1, notpresent to 2), htn (yes to 1, no to 2), dm (yes becomes 1, no becomes 2), cad (yes becomes 1, no becomes 2), appet (good becomes 1, poor becomes 2), pe (yes becomes 1, no becomes 2) and ane (yes becomes 1, no be 2).

Table 2. Sample Dataset (after preprocessing)

age	bp	sg	al	su	rbc	pc	pcc	ba
48	80	1020	1	0	1	1	2	2
7	50	1020	4	0	1	1	2	2
62	80	1010	2	3	1	1	2	2
48	70	1005	4	0	1	2	1	2
51	80	1010	2	0	1	1	2	2
60	90	1015	3	0	1	1	2	2
68	70	1010	0	0	1	1	2	2
24	76	1015	2	4	1	2	2	2
52	100	1015	3	0	1	2	1	2
53	90	1020	2	0	2	2	1	2

bgr	bu	sc	sod	pot	hemo	pcv	wbcc	rbcc
121	36	1	138	5	15	44	7800	5
148	18	1	138	5	11	38	6000	5
423	53	2	138	5	10	31	7500	5
117	56	4	111	3	11	32	6700	4
106	26	1	138	5	12	35	7300	5
74	25	1	142	3	12	39	7800	4
100	54	24	104	4	12	36	8406	5
410	31	1	138	5	12	44	6900	5
138	60	2	138	5	11	33	9600	4
70	107	7	114	4	10	29	12100	4

htn	dm	cad	appet	pe	ane	class
1	1	2	1	2	2	ckd
2	2	2	1	2	2	ckd
2	1	2	2	2	1	ckd
1	2	2	2	1	1	ckd
2	2	2	1	2	2	ckd
1	1	2	1	1	2	ckd
2	2	2	1	2	2	ckd
2	1	2	1	1	2	ckd
1	1	2	1	2	1	ckd
1	1	2	2	2	1	ckd

After preprocessing, the dataset is divided into training data and testing data using **10-fold cross validation**, after which the data will be

selected through the **embedded feature selection method**, according to Zhu & Song[15] the feature search process is embedded into the classification algorithm, and the learning process with the feature selection process cannot be separated which in this study is applied to SVM.

Support Vector Machine is included in the prediction algorithm. Where in applying the SVM algorithm it is necessary to find the distance from a dataset. While in reality many cannot be separated in a linear way [19]. It is called a non-linear dataset, where dataset conditions cannot be separated by fields that separate data between classes from one another. So that this problem can be solved through Feature Space. Where is a method of converting the input space (dot product) which can only separate linear data into high dimensional form (feature space). Feature space or also known as kernel trick then changed to K Kernel (xi, xj) [19]. Kernel functions that can be used include radial (RBF), polynomial, and dot (linear). The following is the formulation of the three kernels:

Table 3. Kernel SVM Testing

Kernel Name	Kernel $\boldsymbol{K}$ $(\boldsymbol{x}, \boldsymbol{y})$, $\boldsymbol{i}$=1, 2,.....,$\boldsymbol{N}$
Dot (Linear)	$K(x, y) = xTy + c$
Polynomial	$K(x,y) = (axTy + c)d$
Radial (RBF)	$K(x, y) = exp(-g \|\|x - y\|\|2)$

Attribute selection which is commonly referred to as Feature Selection used in data mining is a process to find a subset of feature selection results from a dataset [19] yang bertujuan untuk menghilangkan atribut yang tidak relevan. which aims to eliminate irrelevant attributes. There are 3 feature selections [19], including:

1. Based Feature Selection
   Use a statistical method to assess each attribute. Attributes that will be ranked will be made a matrix first in order to identify the relevant or irrelevant attributes that will be removed from the dataset.
2. Embedded Feature Selection
   Run feature selection during the training process from parameters to get optimal values. In this type of feature selection, it adopts the principle of regularization methods which have a limiting penalty factor.
3. Wrapper Feature Selection
   Perform a search process from a combination of attribute groups by comparing one attribute with other attributes.

The attribute search process using forward selection is the first time with an empty model, then the variables are inputted until the attribute model combination is met properly. Below is the pseudo code of the forward selection:

1. Create an empty set: $Yk = \{\emptyset\}$, $k = 0$ .....(1)
2. Choose the best features:
   $X+ = \arg maxx+\in Yk\ [J(Yk + X+)]$ .....(2)
3. If $J((Yk + x+) > J((Yk)$ .....(3)
   a. Update $Yk+1 = Yk + x+$
   b. $k = k + 1$
   c. Back to step -2

Next, **validation** is carried out using the confusion matrix which can be seen in table 4 so that the accuracy value can be obtained, as follows:

Table 4. Confusion Matrix

Real Condition	Prediction Result	
	+	-
+	*True Positives*	*False Negatives*
-	*False Positives*	*True Negatives*

True positive, the prediction engine can predict a positive value as a positive value, meaning that a true positive represents the correct value, which is in accordance with actual conditions. Same goes for true negatives.

False negative, the machine predicts a positive value as a negative value. This means that the negative value generated by the prediction engine is wrong or can be called a false negative value, because the actual value should be positive.

Accuracy formula is obtained by the following equation:

$$\text{Accuracy} = \frac{(\text{TN}+\text{TP})}{(\text{TN}+\text{FN}+\text{TP}+\text{FP})} \text{ x } 100\% \text{.....(4)}$$

The value of the confusion matrix will be 'used to calculate' predictive **performance** [20]. The performance measurement consists of True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN) variables.

RESULTS AND DISCUSSION

The dataset used is chronic_kidney_disease data obtained from the UCI machine learning repository on the page https://archive.ics.uci.edu/ml/machine-learning-databases/00336/. This research uses the Support Vector Machine (SVM) method based on forward selection. The results of this study aim to determine the level of accuracy of the method applied in this study, as well as to determine whether the forward selection performance is more effectively applied in the SVM method to solve problems in chronic kidney disease.

After the dataset is preprocessed, then the data is processed by applying 10-fold cross validation by dividing the data into 90% training data and 10%

Accredited Rank 3 (Sinta 3) based on the Decree of the Dirjen Penguatan RisBang Kemenristekdikti No.28/E/KPT/2019, September 26, 2019. Published by LPPM Universitas Nusa Mandiri

VOL. 8. NO. 1 AUGUST 2022
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v8i1.3327

testing data. After that, the first experiment was carried out by entering training data into the Support Vector Machine (SVM) method by comparing the dot, polynomial, and RBF kernels.Percobaan kedua, dengan menerapkan teknis seleksi fitur *embedded method* yaitu forward selection pada SVM dengan kernel *dot, polynominal*, dan RBF.

From two experiments, namely SVM (dot, polynomial, RBF) and SVM + FS (dot, polynomial, RBF), the accuracy values can be obtained as shown in table 5 below:

Table 5. Support Vector Machine (SVM) Accuracy Value

Algorithm	TP	TN	FP	FN	Accuracy	AUC
SVM (Dot)	244	150	0	6	98,50%	1,000
SVM (Polynominal)	239	150	0	11	97,25%	0,986
SVM (RBF)	250	119	31	0	92,25%	0,999

Table 5 shows that SVM with dot kernel has higher accuracy than SVM with polynominal kernel and RBF which is 98,505 with AUC 1,000. While in Table IV is a table display of the comparison of the accuracy values of the SVM method based on Forward Selection (FS) with the dot, polynominal and RBF kernels. It can be seen that SVM+FS with the RBF kernel outperforms the accuracy and AUC values of other kernels, with an accuracy value of 99.75% and an AUC of 1,000.

Table 6. SVM Accuracy Value With Forward Selection

Algorithm	TP	TN	FP	FN	Accuracy	AUC
SVM + FS (Dot)	248	150	0	2	99,50%	0,998
SVM + FS (Polynominal)	232	150	0	18	95,50%	0,964
SVM + FS (RBF)	250	149	1	0	99,75%	1,000

The accuracy results obtained by comparing the accuracy value and the AUC value in the SVM method with the accuracy value and the AUC value in the SVM + FS method, it has increased. It can be seen from the graphs 3 and 4 as below.

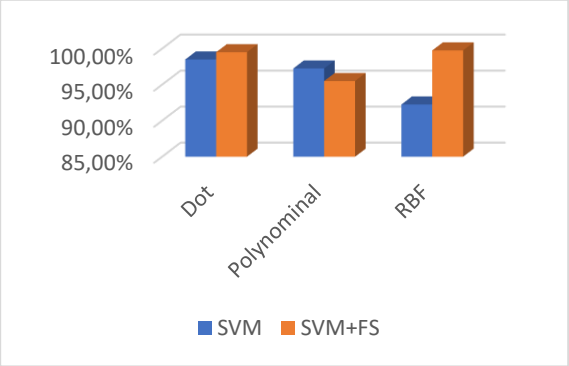

Figure 3. Comparison of Accuracy Values in the SVM Method with the SVM+FS Method

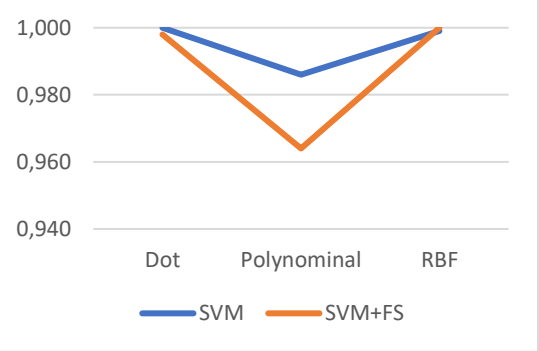

Figure 4. Comparison of AUC Value in the SVM Method with the SVM+FS method

CONCLUSION

Support Vector Machine (SVM) has the advantage of a high level of accuracy, but SVM also has the disadvantage that it is difficult to select features for optimal input, but Forward Selection can overcome the problems of SVM with feature selection to improve machine learning performance. So this study aims to determine the accuracy of the prediction of chronic kidney disease in the SVM method based on Forward Selection (FS) and to determine the performance of the FS which is applied to the SVM method in solving problems in chronic kidney disease.

From the results of the research, experiments were carried out on the SVM method using various kernels and it was seen that SVM with the dot kernel was 98.50% with AUC 1,000 which was superior to the polynominal kernel and RBF. However, when the experiment was carried out again by applying FS to SVM, it was found that SVM+FS with the RBF kernel outperformed the other kernels by 99.75% with AUC 1,000. So it can be concluded that the Forward Selection on SVM has succeeded in improving its performance, especially in this case, namely the prediction of chronic kidney disease. For

further research, it is recommended to compare the accuracy results by entering parameter values in SVM because one of the shortcomings of SVM is that it is weak in determining the optimal parameter values.

REFERENCE

[1] Aulia, "Fungsi dan Faktor Risiko Ginjal," *P2PTM Kementerian Kesehatan*, 2017. [Online]. Available: http://p2ptm.kemkes.go.id/kegiatan-p2ptm/subdit-penyakit-jantung-dan-pembuluh-darah/fungsi-dan-faktor-risiko-ginjal. [Accessed: 04-Mar-2022].

[2] E. Febria Sari, "Waspada Ginjal Kronis," *Hermina Galaxy*, 2021. [Online]. Available: https://herminahospitals.com/id/articles/waspada-gagal-ginjal-kronis.html. [Accessed: 04-Mar-2022].

[3] "Chronic Kidney Disease," *worldkidneyday.org*, 2022. [Online]. Available: https://www.worldkidneyday.org/facts/chronic-kidney-disease/. [Accessed: 04-Mar-2022].

[4] Pusdatin, "Situasi Penyakit Ginjal Kronis," *pusdatin kemkes*, Indonesia, pp. 1–4, 2017.

[5] I. Yulianti, R. Amegia Saputra, M. Sukrisno Mardiyanto, and A. Rahmawati, "Optimasi Akurasi Algoritma C4.5 Berbasis Particle Swarm Optimization dengan Teknik Bagging pada Prediksi Penyakit Ginjal Kronis Optimization of C4.5 Algorithm Based On Particle Swarm Optimization with Bagging Technique on Prediction of Chronic Kidney Dise," *Techno.COM*, vol. 19, no. 4, pp. 411–421, 2020.

[6] Kemenkes RI, "Hasil Riset Kesehatan Dasar Tahun 2018," 2018.

[7] N. A. Almansour *et al.*, "Neural network and support vector machine for the prediction of chronic kidney disease: A comparative study," *Comput. Biol. Med.*, vol. 109, no. October 2018, pp. 101–111, 2019.

[8] Z. Chen, X. Zhang, and Z. Zhang, "Clinical risk assessment of patients with chronic kidney disease by using clinical data and multivariate models," *Int. Urol. Nephrol.*, vol. 48, no. 12, pp. 2069–2075, 2016.

[9] A. Y. Al-Hyari, A. M. Al-Taee, and M. A. Al-Taee, "Clinical decision support system for diagnosis and management of Chronic Renal Failure," *2013 IEEE Jordan Conf. Appl. Electr. Eng. Comput. Technol. AEECT 2013*, 2013.

[10] A. A. Serpen, "Diagnosis Rule Extraction from Patient Data for Chronic Kidney Disease Using Machine Learning," *Int. J. Biomed. Clin. Eng.*, vol. 5, no. 2, pp. 64–72, 2016.

[11] R. Ani, G. Sasi, U. R. Sankar, and O. S. Deepa, "Decision support system for diagnosis and prediction of chronic renal failure using random subspace classification," *2016 Int. Conf. Adv. Comput. Commun. Informatics, ICACCI 2016*, pp. 1287–1292, 2016.

[12] N. Tazin, S. Lecturer, S. A. Sabab, and M. T. Chowdhury, "Effective Classification and Feature Selection," *2016 Int. Conf. Med. Eng. Heal. Informatics Technol.*, pp. 1–6, 2016.

[13] S. B. Akben, "Early Stage Chronic Kidney Disease Diagnosis by Applying Data Mining Methods to Urinalysis, Blood Analysis and Disease History," *Irbm*, vol. 39, no. 5, pp. 353–358, 2018.

[14] H. Polat, H. Danaei Mehr, and A. Cetin, "Diagnosis of Chronic Kidney Disease Based on Support Vector Machine by Feature Selection Methods," *J. Med. Syst.*, vol. 41, no. 4, 2017.

[15] H.- Harafani, "Forward Selection pada Support Vector Machine untuk Memprediksi Kanker Payudara," *J. Infortech*, vol. 1, no. 2, pp. 131–139, 2020.

[16] E. Purwaningsih, "Application of the Support Vector Machine and Neural Network Model Based on Particle Swarm Optimization for Breast Cancer Prediction," *SinkrOn*, vol. 4, no. 1, p. 66, 2019.

[17] E. Purwaningsih, "Penerapan Particle Swarm Optimization pada Metode Neural Network untuk Perawatan Penyakit Kutil melalui Immunotherapy," *J. Sist. dan Teknol. Inf.*, vol. 8, no. 2, p. 207, 2020.

[18] W. Li and Z. Liu, "A method of SVM with normalization in intrusion detection," *Procedia Environ. Sci.*, vol. 11, no. PART A, pp. 256–262, 2011.

[19] J. Han, J; Kamber, M; & Pei, *Data Mining Concepts and Techniques. In Data Mining.* 2012.

[20] E. Purwaningsih, "Seleksi Mobil Berdasarkan Fitur dengan Komparasi Metode Klasifikasi Neural Network, Support Vector Machine, dan Algoritma C4.5," *J. Pilar Nusa Mandiri*, vol. XII, no. 2, pp. 153–160, 2016.

Accredited Rank 3 (Sinta 3) based on the Decree of the Dirjen Penguatan RisBang Kemenristekdikti No.28/E/KPT/2019, September 26, 2019. Published by LPPM Universitas Nusa Mandiri