



Available : <https://journal.isi-padangpanjang.ac.id/index.php/JTST>

Vol 2, No 2, Dec 2023

E-ISSN: 2962-5378

Classification of Chronic Kidney Disease based on health care records using machine learning with Support Vector Machine

Abdurrahman Niarman ¹, iswandi² Amuharnis ³

Department of Informatics Management, UIN Mahmud Yunus Batusangkar¹, Department of Informatics, Universitas Metamedia ²

aabniarman@uinmybatusangkar.ac.id, iswandi@uinmybatusangkar.ac.id,
amuharnis@metamedia.ac.id

ABSTRACT

Chronic Kidney Disease (CKD) is a global health concern with a rising prevalence that necessitates early and accurate diagnosis for effective management. This study proposes the application of Machine Learning (ML), specifically Support Vector Machine (SVM), to classify CKD based on health care records. Leveraging a comprehensive dataset of patient health records, including clinical and demographic information, the research aims to develop a predictive model that can assist in the timely identification of individuals at risk of CKD. The methodology involves preprocessing the health care records, extracting relevant features, and implementing the SVM algorithm for classification. The dataset is divided into training and testing sets to evaluate the model's performance. The SVM classification model that was developed after going through the data preprocessing process produced results that were good enough to be able to classify whether a patient was diagnosed with CKD or not with an accuracy level of 98% and a total of 400 lines of data and 25 features.

Keywords: Chronic Kidney Disease, Machine Learning, Support Vector Machine, Health Care Records, Classification, Predictive Modeling

Abdurrahman Niarman

JTSTI-Journal of Tourism Sciences, Technology and Industry

Vol 2, No 2, Dec 2023

E-ISSN: 2962-5378

INTRODUCTION

Chronic kidney failure is a progressive and slow development of kidney failure, and usually lasts for one year. The kidneys lose the ability to maintain the volume and composition of body fluids under normal food intake (Slyvia Anderson et al., 2006). Chronic Kidney Disease in the world is currently increasing and becoming a serious health problem, the results of the 2010 Global Burden of Disease research, chronic kidney disease was the 27th leading cause of death in the world in 1990 and increased to 18th in 2010. In 2013, as many as 2 per 1000 population or 499,800 Indonesians suffered from kidney failure. As many as 6 per 1000 population or 1,499,400 Indonesians suffer from kidney stones (Kemenkes, 2013).

Chronic kidney disease (CKD) arises from many heterogeneous disease pathways that alter the function and structure of the kidney irreversibly, over months or years. The diagnosis of CKD rests on establishing a chronic reduction in kidney function and structural kidney damage. The best available indicator of overall kidney function is glomerular filtration rate (GFR), which equals the total amount of fluid filtered through all of the functioning nephrons per unit of time. Generally, CKD is caused by diffuse and chronic intrinsic kidney disease. Glomerulonephritis, essential hypertension, and pyelonephritis are the most common causes of chronic renal failure, accounting for approximately 60% (Sukandar, 2006).

The burden of CKD is substantial. According to WHO global health estimates, 864 226 deaths (or 1.5% of deaths worldwide) were attributable to this condition in 2012. Ranked fourteenth in the list of leading causes of death, CKD accounted for 12.2 deaths per 100 000 people. Since 1990, only deaths from complications of HIV infection have increased at a faster rate than deaths from CKD. Projections from the Global Health Observatory suggest that although the death rate from HIV will decrease in the next 15 years, the death rate from CKD will continue to increase to reach 14 per 100 000 people by 2030. CKD is also associated with substantial morbidity. Worldwide, CKD accounted for 2 968 600 (1.1%) of disability-adjusted life-years and 2 546 700 (1.3%) of life-years lost in 2012 (figure 1) (Webster et al., 2017).

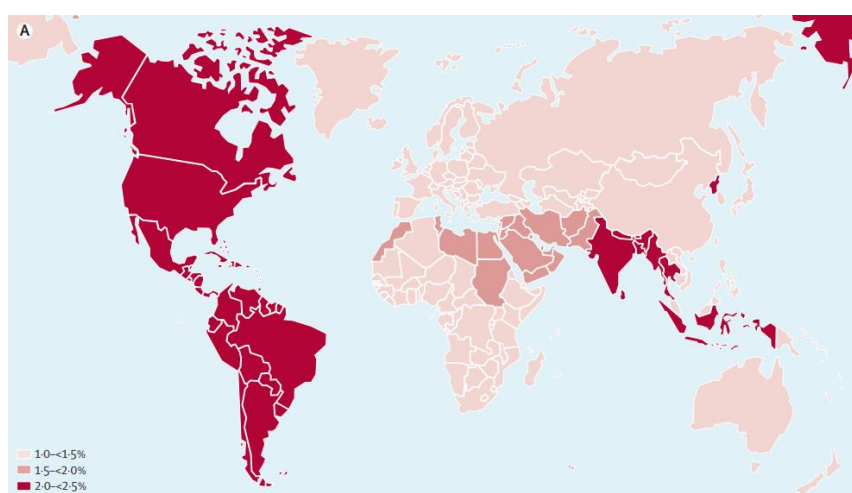


Figure 1. Proportion of total mortality attributed to kidney disease (Webster et al., 2017)

Abdurrahman Niarman

JTSTI-Journal of Tourism Sciences, Technology and Industry

Vol 2, No 2, Dec 2023

E-ISSN: 2962-5378

Kidney failure can be caused by age, gender, and a history of diseases such as diabetes, hypertension or other metabolic disorders which can cause a decrease in kidney function. In addition, misuse of analgesic drugs and NSAIDs, both over the counter and prescribed by doctors, for years can trigger the risk of papillary necrosis and chronic kidney failure. Smoking habits and the use of energy supplement drinks can also cause kidney failure (Pranandari & Supadmi, 2015).

Medical record data that has been accumulated over time to become big data can be used as a tool to develop a classification model later. Classification is one of several data mining techniques for making predictions, where the predicted value is in the form of a label (target variable) (Septhya et al., 2023). Machine learning will be used using the Support Vector Machine method to produce a model that can help classify patient illnesses based on the input medical records.

The Support Vector Machine (SVM) algorithm works to obtain test prediction results, where the results of predictions for testing are obtained from groups in the form of feature vectors (Muhammad Prasetyo et al., 2022). Support Vector Machine (SVM) is a selection technique that produces results with the highest level of classification accuracy by comparing a set of standard parameters with discrete values called the candidate set (Suhardjono et al., 2019). The advantage of Support Vector Machine (SVM) is that it is a popular method and very suitable for classification because it does not depend on the number of attributes and can solve dimensional problems. Computationally, Support Vector Machine (SVM) can carry out training quickly and also its learning techniques can deal with difficulties in doubt (Purwaningsih, 2016).

This research was carried out in order to create a classification model using the Support Vector Machine method with a good level of accuracy so that it can help doctors in diagnosing patients. The final decision remains in the hands of the doctor in determining the diagnosis that will be given to the patient, but using this model will be able to help the doctor in translating the variables in the patient's medical record into information.

RESULT AND DISCUSSION

There are several stages that must be implemented to be able to produce a model with a good level of accuracy and is quite reliable. The distribution of training data and test data is divided into 75 and 25 with a total of 400 rows of records. The next step is to create a Support Vector Machine (SVM) algorithm model which is then continued by running the model with training data. Then the test data predicts the results, the actual test data results are compared with the predicted results so as to obtain the level of accuracy of the predicted results.

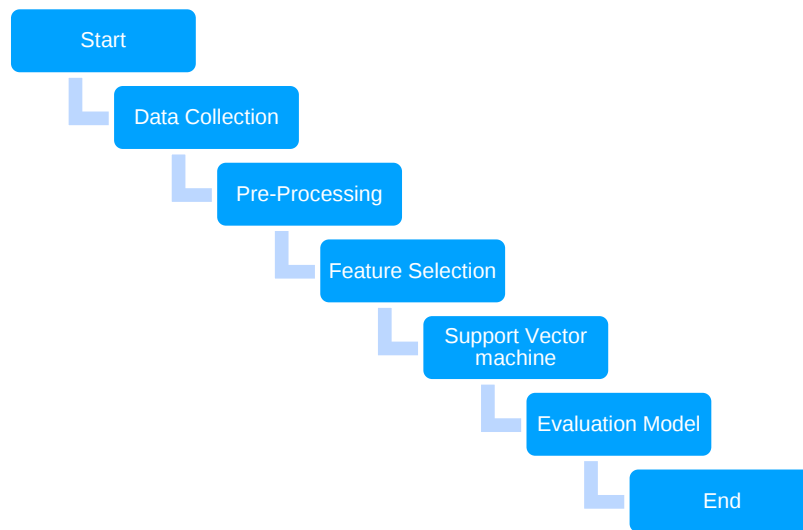


Figure 2. research flow

a) Data Collection

In this research, the main data was obtained by reviewing the chronic kidney disease data on the Kaggle.com page. The data that will be used as a data set in this research is 400 rows. The data is processed to obtain a list of people who have the potential to suffer from kidney failure.

Table 1. Chronic Kidney Disease Data set

	id	age	bp	sg	al	su	rbc	pc	pcc	ba	...	pcv	wc	rc	htn	dm	cad	appet	pe	ane	classification
395	395	55.0	80.0	1.020	0.0	0.0	normal	normal	notpresent	notpresent	...	47	6700	4.9	no	no	no	good	no	no	notckd
396	396	42.0	70.0	1.025	0.0	0.0	normal	normal	notpresent	notpresent	...	54	7800	6.2	no	no	no	good	no	no	notckd
397	397	12.0	80.0	1.020	0.0	0.0	normal	normal	notpresent	notpresent	...	49	6600	5.4	no	no	no	good	no	no	notckd
398	398	17.0	60.0	1.025	0.0	0.0	normal	normal	notpresent	notpresent	...	51	7200	5.9	no	no	no	good	no	no	notckd
399	399	58.0	80.0	1.025	0.0	0.0	normal	normal	notpresent	notpresent	...	53	6800	6.1	no	no	no	good	no	no	notckd

b) Pre-Processing Data

Data preprocessing is a process of raw data for other processing operations. The data preprocessing stage deletes data that is null or empty and changes the data to be more structured using the implementation of data cleaning and data transformation. This step is responsible for making the data to be processed more structured and making the modeling process smoother.

c) Feature Selection

Feature selection aims to select features that are influential and exclude features that have no influence in a modeling or data analysis activity. Using feature selection techniques reduces the number of features used for learning and selects

Abdurrahman Niarman

JTSTI-Journal of Tourism Sciences, Technology and Industry

Vol 2, No 2, Dec 2023

E-ISSN: 2962-5378

high discrimination features in the feature selection process. Additionally, feature selection helps improve accuracy by selecting optimal features (Lee et al., 2017).

Table 2. Feature Selection from Data Set of Chronic Kidney Disease

	specific_gravity	red_blood_cells	pus_cell	pus_cell_clumps	bacteria	blood_glucose_random	haemoglobin	packed_cell_volume	white_blood_cell_count
0	1.020	0.0	0.0	0.0	0.0	121.000000	15.4	44	7800
1	1.020	0.0	0.0	0.0	0.0	148.036517	11.3	38	6000
2	1.010	0.0	0.0	0.0	0.0	423.000000	9.6	31	7500
3	1.005	0.0	1.0	1.0	0.0	117.000000	11.2	32	6700
4	1.010	0.0	0.0	0.0	0.0	106.000000	11.6	35	7300

d) Support Vector Machine

It is a technique that makes predictions for classification or regression (Neneng et al., 2021). The classification process has two steps, namely the testing process and the training process (Chazar & Erawan, 2020). The training process is used to create a model for a test set. The basic principle of Support Vector Machine classification is linearly separating, but SVM has developed so that it can work on non-linear problems by optimizing the kernel function with a hyperplane which can maximize the distance (margin) between data classes. Hyperplane SVM is denoted:

$$f(x) = w^2 x + b$$

e) Evaluation Model

At this stage, the method is evaluated by measuring the performance of the Support Vector Machine (SVM). This method was evaluated by comparing the accuracy level of the algorithm using 75:25 data splitting. To calculate the error value of the classification method, a confusion matrix is used so that the performance of the classification method can be evaluated.

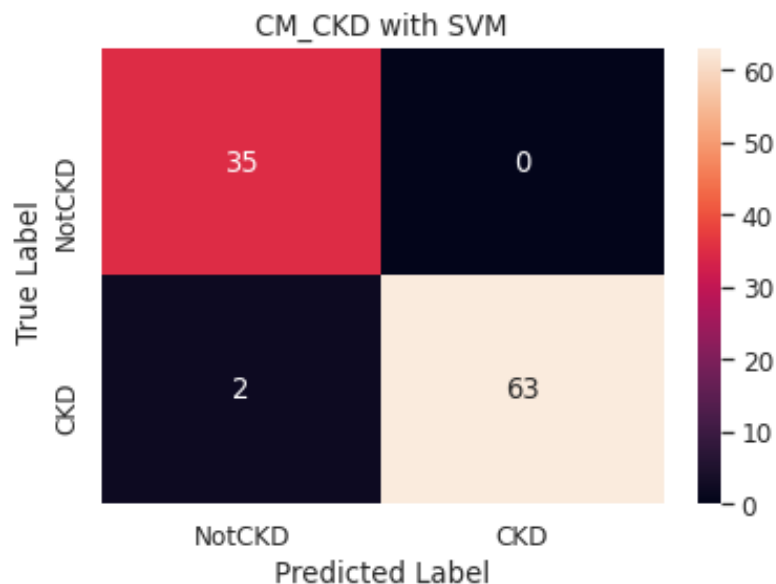


Figure 3. Confusion Matrix SVM

Abdurrahman Niarman

JTSTI-Journal of Tourism Sciences, Technology and Industry

Vol 2, No 2, Dec 2023

E-ISSN: 2962-5378

CONCLUSION

The success of the proposed Support Vector Machine algorithm in this research is measured based on the accuracy value. Based on the test results using 400 rows of data with a 75:25 division of training data and test data, it produces an accuracy value of 98%, a precision value of 100%, a recall value of 96%, and an F-measure value of 98%. The results of developing a machine learning model using the support vector machine method are considered good enough to help classify whether a patient is diagnosed with chronic kidney failure or not based on the input medical record data. This research can be developed further by using a more comprehensive data set with a larger amount of data and more complex features to ensure the model developed is truly reliable.

REFERENCES

- Chazar, C., & Erawan, B. (2020). Machine Learning Diagnosis Kanker Payudara Menggunakan Algoritma Support Vector Machine. *INFORMASI (Jurnal Informatika Dan Sistem Informasi)*, 12(1), 67–80. <https://doi.org/10.37424/informasi.v12i1.48>
- Septhya, D., Rahayu, K., Rabbani, S., Fitria, V., Irawan, Y., & Hayami, R. (2023). *MALCOM: Indonesian Journal of Machine Learning and Computer Science Implementation of Decision Tree Algorithm and Support Vector Machine for Lung Cancer Classification Implementasi Algoritma Decision Tree dan Support Vector Machine untuk Klasifikasi Penyakit Kanker Paru*. 3, 15–19.
- Kemenkes. (2013). *Laporan Riset Kesehatan Dasar (RISKESDAS) Tahun 2013*.
- Lee, J., Park, D., & Lee, C. (2017). Feature selection algorithm for intrusions detection system using sequential forward search and random forest classifier. *KSII Transactions on Internet and Information Systems*, 11(10), 5132–5148. <https://doi.org/10.3837/tiis.2017.10.024>
- Muhammad Prasetyo, T., Amrullah, A., Syahrir, S., & Nurina Sari, B. (2022). IMPLEMENTASI ALGORITMA SVM (SUPPORT VECTOR MACHINE) DALAM KLASIFIKASI PENYAKIT PARU-PARU BERDASARKAN FITUR POLA BENTUK. *Jurnal Teknologi Informasi*, 6(1).

Abdurrahman Niarman

JTSTI-Journal of Tourism Sciences, Technology and Industry

Vol 2, No 2, Dec 2023

E-ISSN: 2962-5378

- Neneng, N., Putri, N. U., & Susanto, E. R. (2021). Klasifikasi Jenis Kayu Menggunakan Support Vector Machine Berdasarkan Ciri Tekstur Local Binary Pattern. *CYBERNETICS*, 4(02). <https://doi.org/10.29406/cbn.v4i02.2324>
- Pranandari, R., & Supadmi, W. (2015). FAKTOR RISIKO GAGAL GINJAL KRONIK DI UNIT HEMODIALISIS RSUD WATES KULON PROGO RISK FACTORS CRONIC RENAL FAILURE ON HEMODIALYSIS UNIT IN RSUD WATES KULON PROGO. In *Tahun* (Vol. 11, Issue 2).
- Purwaningsih, E. (2016). Seleksi Mobil Berdasarkan Fitur Dengan Komparasi Metode Klasifikasi Neural Network, Support Vector Machine, dan Algoritma C4.5. *Jurnal Pilar Nusa Mandiri*, XII(2), 153–160.
- Slyvia Anderson, P., Wilson, L. M., & Peter, A. (2006). *Patofisiologi : konsep klinis proses-proses penyakit*. EGC.
- Suhardjono, Wijaya, G., & Hamid, A. (2019). Prediksi Waktu Kelulusan Mahasiswa Menggunakan SVM Berbasis PSO. *Bianglala Informatika*, 7(2).
- Sukandar, E. (2006). *Nefrologi Klinik* (3rd ed.). Universitas padjajaran Press.
- Webster, A. C., Nagler, E. V., Morton, R. L., & Masson, P. (2017). Chronic Kidney Disease. In *The Lancet* (Vol. 389, Issue 10075, pp. 1238–1252). Lancet Publishing Group. [https://doi.org/10.1016/S0140-6736\(16\)32064-5](https://doi.org/10.1016/S0140-6736(16)32064-5)