IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁙ Impact Factor 8.102 ⁙ Peer-reviewed / Refereed journal ⁙ Vol. 12, Issue 7, July 2023

DOI: 10.17148/IJARCCE.2023.12704

Advanced Risk Assessment for Chronic Kidney Disease using Machine Learning

Pranitha P[1], C S Swetha[2]

Dept of MCA, Bangalore Institute of Technology, Bengaluru, India[1]

Asst. Professor, Dept of MCA, Bangalore Institute of Technology, Bengaluru, India[2]

Abstract: Chronic Kidney Syndrome (CKD) is a persistent medical condition characterized by the gradual deterioration of renal function over time. It is a significant global health concern, impacting a substantial number of individuals. With advancements in technology, particularly in the field of machine learning (ML), there is an opportunity to utilize these tools for improving the detection, prediction and supervision of CKD. The objective of this scheme is to develop a extrapolative prototype for CKD and facilitate its management through the application of ML algorithms and techniques. By analyzing extensive datasets comprising patient medical records, demographic statistics, test site outcomes, and other relevant factors, this initiative aims to identify patterns, trends, and threat aspects accompanying with CKD. These insights can assist healthcare professionals in making more accurate assessments regarding CKD progression and devising personalized treatment plans. We propose the consumption of the Support Vector Machine (SVM) machine learning model to forecast CKD based on relevant clinical features. Our findings validate the effectiveness of the SVM model in accurately predicting CKD, achieving an impressive accuracy rate of 94%.

Keywords: chronic kidney infection; CKD stage identification; machine learning, support vector machine

I. INTRODUCTION

Chronic kidney disease (CKD) is a chronic condition characterized by the gradual decline in kidney function over time. The kidneys play a vital role in filtering waste products, excess fluids, and toxins from the bloodstream, along with maintaining electrolyte balance and producing hormones important for red blood cell production and bone health.CKD develops slowly and progresses through different stages, ranging from mild to severe, depending on the extent of kidney function impairment. Although high blood pressure (hypertension) and diabetes are the primary causes of CKD, supplementary aspects like certain medications, infections, immune disorders, and genetic conditions can also underwrite to its onset.

CKD can lead to complications and an elevated risk of numerous health issues, together with cardiovascular disease, anemia, bone disorders, and ultimately, kidney failure necessitating dialysis or transplantation. Detecting CKD early and managing it effectively are critical for reducing down its progression, preventing complications, and preserving kidney function.Treatment strategies for CKD often involve a multifaceted approach. Lifestyle modifications, such as maintaining a healthy diet, regular exercise, and avoiding smoking and excessive alcohol consumption, are recommended. Blood pressure control is of utmost importance, particularly for individuals with hypertension, while managing blood sugar levels is crucial for those with diabetes. Medication adjustments, dietary changes to manage protein, sodium, and potassium intake, and monitoring fluid balance are also common components of CKD management.

SVM is particularly well-suited for diagnosing CKD due to its capability to handle complex datasets and its effectiveness in dealing with high-dimensional spaces. The determination of this learning is to advance and evaluate an SVM-based machine learning model for identifying CKD. By training the model on a large dataset of patient data, including both CKD-positive and CKD-negative cases, our purpose is to generate a robust and reliable tool for the initial recognition and categorization of CKD. The SVM model will undergo rigorous training and validation processes, incorporating various techniques such as cross-validation and hyperparameter tuning to optimize its performance. Once developed, the prototype will be tested on an independent dataset to assess its applicability and accuracy in real-world scenarios. The ultimate goal of this project is to provide healthcare professionals with a authoritative decision provision tool that can assist in the early identification and management of CKD. SVM exhibits expertise in effectively handling complex datasets with high-dimensional feature spaces. By utilizing the kernel trick, SVM is capable of accurately addressing non-linear decision boundaries by transmuting the involvement statistics into a higher-dimensional feature space.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁒ Impact Factor 8.102 ⁒ Peer-reviewed / Refereed journal ⁒ Vol. 12, Issue 7, July 2023

DOI: 10.17148/IJARCCE.2023.12704

II. METHODS

Support Vector Machines (SVM) are a administered machine learning methodology that is well-suited for both organization and regression tasks. SVM's effectiveness lies in its ability to handle complex problems by finding a clear separation between different classes.The fundamental idea behind SVM is to identify an optimum hyperplane that effectively separates data points belonging to different classes. In binary classification, this hyperplane acts as the decision boundary, accurately dividing the data into two classes. The objective of SVM is to discovery the hyperplane with the extreme margin, which represents the distance between the decision boundary and the adjacent data points from every class.

The University of California, Irvine's Machine Learning Source provided the dataset. This dataset confined 400 exclusive cases. The characteristics of the dataset we used for the experiment are below Within the machine learning framework, a dataset refers to a collection of information utilized for training, evaluating, and testing machine learning models. It consists of instances or examples, each comprising a set of features or attributes associated with a corresponding label or target value in supervised learning scenarios. Datasets are of paramount importance in training machine learning models as they provide the necessary information for learning patterns and making predictions or decisions. They can vary in size, complexity, and structure, depending on the specific problem and data type. Datasets can be sourced from diverse origins such as databases, sensors, online repositories, or data generation processes. The data within datasets may encompass numerical, categorical, or textual formats and can be represented in various forms such as tabular data, images, audio files, or text documents.

Prior to utilizing a dataset for training a machine learning model, it is crucial to preprocess and prepare the data appropriately. This preprocessing comprises tasks such as cleaning statistics, handling missing values, encoding categorical variables, normalizing or scaling features, and splitting the dataset into training, validation, and test sets. Constructing a high-quality and representative dataset is essential for developing accurate and robust machine learning models. The dataset should cover a wide range of scenarios, exhibit sufficient variability, and maintain balance across different classes or categories. Moreover, larger datasets are generally preferred as they provide more examples for the model to learn from and help mitigate overfitting.

A. Attributes of ckd patient dataset

TABLE 1 DATA DESCRIPTION

Attributes	Description
age	The individual's age.
bp	Measurement of blood pressure.
al	Detection of albumin in the urine.
su	Identification of sugar in the urine.
rbc	Detection of red blood cells in the urine.
pc	Presence of pus cells in the urine.
pcc	Presence of clumps of pus cells in the urine.
ba	Presence of bacteria in the urine.
bgr	Level of random blood glucose.
bu	Measurement of blood urea level.
sc	Level of serum creatinine.
pot	Potassium level in the blood.
wc	Count of white blood cells.
htn	Indication of hypertension (high blood pressure).
dm	Indication of diabetes mellitus
cad	Indication of coronary artery disease.
pe	Presence of pedal edema (swelling in the feet and ankles).
ane	Indication of anemia.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁒ Impact Factor 8.102 ⁒ Peer-reviewed / Refereed journal ⁒ Vol. 12, Issue 7, July 2023

DOI: 10.17148/IJARCCE.2023.12704

B. Proposed Algorithm

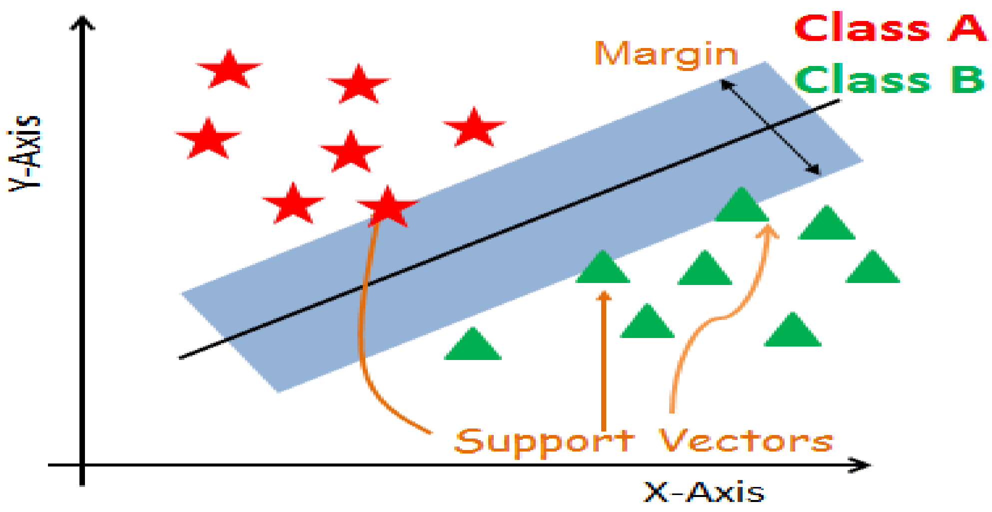

Fig. 1 Support Vector Machine

Here's a concise overview of the SVM algorithm:

1); Data Preparation: Similar to other supervised learning approaches, SVM relies on labeled training data to discern patterns and establish relationships between input features and class labels.
2); Feature Selection and Scaling: Prior to SVM model training, it is vital to carefully select relevant features and perform necessary feature scaling. This step enhances the algorithm's efficiency and overall performance.
3); Kernel Function Selection: SVM employs a kernel function to transform input features into a higher-dimensional space, facilitating the identification of a linear separation. Common kernel functions contain Linear, Polynomial, Gaussian (RBF), and Sigmoid, chosen based on specific data characteristics and problem requirements.
4); Training the SVM Model: The primary objective is to determine the hyperplane that maximizes the margin between classes. This entails formulating an optimization difficult that minimizes errors and maximizes the margin. During this process, support vectors are identified, representing data points closest to the decision boundary.
5); Model Evaluation and Tuning: Following training, the SVM model is evaluated using test data to assess its performance. Estimation metrics such as accurateness, precision, recall, or F1 score are utilized. Additionally, fine-tuning SVM hyperparameters such as the regularization parameter (C) and kernel parameters is crucial for improving the model's overall performance.

SVM offers several advantages contributing to its popularity and effectiveness:

1); Effective in High-Dimensional Spaces: SVM demonstrates proficiency in handling datasets with numerous features, making it appropriate for multifaceted problems involving a large number of input variables.
2); Robust to Overfitting: By incorporating regularization techniques, SVM mitigates the risk of overfitting, ensuring the model's ability to generalize well to unseen data.
3); Versatile: SVM is versatile, capable of addressing both classification and regression tasks. It can handle data that is linearly separable as well as non-linearly separable through the use of appropriate kernel functions.
4); Ability to Handle Outliers: SVM's focus on support vectors enables it to exhibit robustness against outliers, minimizing their impact on the model's decision boundaries.
5); Global Optimization: SVM aims to find the optimal hyperplane that maximizes the margin between classes. Its optimization process is less prone to being trapped in local optima, resulting in more reliable and consistent solutions.
6); Memory Efficiency: Once an SVM model is trained, only a subset of the training data known as support vectors is utilized for predictions. This memory-efficient approach is predominantly advantageous when allocating with large datasets.
7); Interpretable Results: SVM provides interpretable results by identifying the support vectors that contribute to defining the decision boundary. These support vectors offer valuable insights into the data and facilitate understanding of the model's decision-making process.

ISSN (O) 2278-1021, ISSN (P) 2319-5940
International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified ⁙ Impact Factor 8.102 ⁙ Peer-reviewed / Refereed journal ⁙ Vol. 12, Issue 7, July 2023
DOI: 10.17148/IJARCCE.2023.12704

III. RESULT AND DISCUSSION

In this reading, our effort was on investigating machine learning (ML) methods and conducting experiments to categorize the stages of chronic kidney disease (CKD). Our system utilizes support vector machines (SVM) within an ML framework to regulate whether a patient has CKD or not. Before applying the classification algorithm, we performed feature selection to remove precise features that were measured irrelevant or unnecessary.

A. Data Preprocessing

```
data = pd.read_csv('kidney_disease.csv')
data.info()
data['class'].unique()"
data['class'] = data['class'].replace(['ckd','ckd\t','notckd'], [1,1,0])
data.isnull().sum()
```

B. Calculating Accuracy Score

```
from sklearn.metrics import accuracy_score
y_predi = classifier.predict(X_test )
accuracy_score(y_predi,y_test)
import sklearn.metrics
print(sklearn.metrics.classification_report(y_test, y_predi))
```

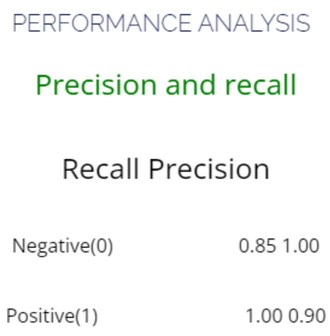

Fig. 2 Performance Analysis

C. Confusion Matrix

```
import seaborn as sns
import matplotlib.pyplot as plt
f, ax=plt.subplots(figsize=(5,5))
sns.heatmap(cm,annot=True,linewidths=0.5,linecolor="red",fmt=".0f",ax=ax)
plt.xlabel("y_pred")
plt.ylabel("y_true")
plt.show()
```

Our findings confirm the high accuracy of the SVM model in predicting chronic kidney disease (CKD), achieving an impressive correctness rate of 94%. The model also demonstrates promising performance in distinguishing between CKD and non-CKD cases, as evidenced by the precision and recall scores.

These results underscore the clinical potential of the SVM model as a decision support tool for early detection and risk stratification of CKD. Thus, this study exemplifies the practical application of the SVM machine learning model in accurately predicting CKD, offering healthcare professionals valuable assistance in identifying individuals at risk, enabling early intervention, and ultimately improving patient outcomes.

DOI: 10.17148/IJARCCE.2023.12704

IV. CONCLUSION

The organization of phases of long-lasting kidney illness patients holds noteworthy reimbursements for both patients and doctors, enabling them to make well-timed and precise clinical decisions. SVM (Support Vector Machine) education models, as efficient tools for diagnosing and predicting chronic renal illness, play a crucial role. These models employ a robust classification algorithm capable of handling complex datasets and have demonstrated success in various medical domains.

Integrating SVM models into medical training allows healthcare professionals to make informed judgments, improve patient care, and contribute to advancements in nephrology. However, it is important to recognize that machine learning models should serve as decision-support tools rather than replacing clinical expertise or judgment. They should be used in combination with the knowledge and experience of medical professionals.

In summary, SVM education models offer valuable assistance in analyzing and predicting CKD. By harnessing this technology, healthcare practitioners can optimize patient care and contribute to ongoing advancements in the field of nephrology. It is crucial to maintain a collaborative approach, combining the strengths of machine learning and clinical expertise for optimal outcomes.

To augment the diagnosis, treatment, and forecast of chronic kidney disease (CKD), the application of AI and machine learning techniques can be utilized to analyze large datasets and uncover significant patterns. These advanced technologies have the potential to predict disease progression, optimize treatment strategies, and identify potential complications, leading to improved patient outcomes. Addressing the challenge of class imbalance in CKD datasets, various techniques can be employed to improve the presentation of Support Vector Machine (SVM) models. These techniques include oversampling the minority class, undersampling the majority class, or utilizing synthetic data generation methods. By implementing these approaches, the issue of class disparity can be mitigated, thereby enhancing the effectiveness of SVM models in analyzing CKD datasets. Furthermore, efforts should be made to progress the affordability and accessibility of handling options for personalities affected by chronic kidney disease.

REFERENCES

[1] C. I. Bagnis, J. E. Heron, and D. M. Gracey, "Contemporary issues and new challenges in chronic kidney disease amongst people living with HIV," AIDS Res. Ther., vol. 17, no. 1, 2020, pp. 1–13, doi: 10.1186/s12981-020-00266-3.

[2] "A machine learning methodology for diagnosing chronic kidney disease," IEEE Access, vol. 8, no. 20991-21002, 2020; doi: 10.1109/ACCESS.2019.2963053. Y. Liu, C. Liu, J. Qin, L. Chen, C. Feng, and B. Chen.

[3] E. H. A. Rady and A. S. Anwar, "Prediction of kidney disease stages using data mining algorithms," Informatics Med. Unlocked, vol. 15, no. March, p. 100178, 2019, doi: 10.1016/j.imu.2019.100178.

[4] M. N. Amin, A. Al Imran, and F. T. Johora, "Classification of Chronic Kidney Disease using Logistic Regression, Feedforward Neural Network, and Wide Deep Learning," 2018 International Conference on Innovative Engineering Technologies, ICIET 2018, pp. 1-6, 2019, doi: 10.1109/CIET.2018.8660844. S. D. Sudarsan, N. Chetty, and K. S. Vaisla, "Role of attributes selection in classification of Chronic Kidney Disease patients," pp. 1-6, 2016, doi: 10.1109/cccs.2015.7374193.

[6] E. Perumal and P. Arulanthu, "Predicting the Chronic Kidney Disease using Various Classifiers," 4th Int. Conf. Electr. Electron. Commun. Computer Technol. Optim. Tech. ICEECCOT 2019, pp. 70–75, 2019, doi: 10.1109/ICEECCOT46775.2019.9114653.

[7] K. Shankar, G. Devika, P. Manickam, and M. Ilayaraja, "Optimal Feature Selection for Chronic Kidney Disease Classification using Deep Learning Classifier," 2018 IEEE Int. Conf. Computer Intelligence and Computer Research (ICCIC), pp. 1–5, 2018, doi: 10.1109/ICCIC.2018.8782340.

[8] "Chronic Kidney Disease Prediction and Recommendation of Suitable Diet Plan by Using Machine Learning," by R. Shinde, S. John, R. Jadhav, A. Maurya, R. Wable, and R. Dakshayani. ICNTE 2019 - Proc., no. Icnte, pp. 1-4, 2019, doi: 10.1109/ICNTE44896.2019.8946029. 2019 Int. Conf. Nascent Technol. Eng.

[9] "Feature selection and dimensionality reduction methods for chronic disease prediction," International Journal of Science and Technology Research, vol. 9, no. 4, 2020, pp. 2912–2918.

[10] CKD diagnosis using Multilayer Perceptron Classifier, 10th International Conference on Cloud Computing, Data Science & Engineering (Confluence), 2020. S. Saraswat, S. Vashisth, and I. Dhall.

[11] "Intelligent Diagnostic Prediction and Classification System for Chronic Kidney Disease," Sci. Rep., vol. 9, no. 1, pp. 1–14, 2019, doi: 10.1038/s41598-019-46074-2, by K. Shankar, M. Elhoseny, and J. Uthayakumar.

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 8.102 ⁜ Peer-reviewed / Refereed journal ⁜ Vol. 12, Issue 7, July 2023

DOI: 10.17148/IJARCCE.2023.12704

[12] M. H. Abd Wahab, A. A. Johari, and A. Mustapha, "Two-Class Classification: Comparative Experiments for Chronic Kidney Disease," in 2019 4th International Conference on Information and System Computing Networks, pp. 789–792, doi: 10.1109/ISCON47742.2019.9036306.

[13] Levey AS, Coresh J. Chronic kidney disease. Lancet. 2012 Jan 14;379(9811):165-80. doi: 10.1016/S0140-6736(11)60178-5.

[14] James SL, Abate D, Abate KH, Abay SM, Abbafati C, Abbasi N, et al. Global, regional, and national incidence, prevalence, and years lived with disability for 354 diseases and injuries for 195 countries and territories, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017. Lancet. 2018 Nov 10;392(10159):1789-858. doi: 10.1016/S0140-6736(18)32279-7.

[15] Mills KT, Xu Y, Zhang W, Bundy JD, Chen CS, Kelly TN, et al. A systematic analysis of worldwide population-based data on the global burden of chronic kidney disease in 2010. Kidney Int. 2015 May;88(5):950-7. doi: 10.1038/ki.2015.230.

[16] Eckardt KU, Coresh J, Devuyst O, Johnson RJ, Köttgen A, Levey AS, et al. Evolving importance of kidney disease: from subspecialty to global health burden. Lancet. 2013 Jan 12;382(9887):158-69. doi: 10.1016/S0140-6736(13)60439-0.

[17] KDIGO 2012 Clinical Practice Guideline for the Evaluation and Management of Chronic Kidney Disease. Kidney Int Suppl (2011). 2013 Jan;3(1):1-150. doi: 10.1038/kisup.2012.73.

[18] Hill NR, Fatoba ST, Oke JL, Hirst JA, O'Callaghan CA, Lasserson DS, et al. Global Prevalence of Chronic Kidney Disease - A Systematic Review and Meta-Analysis. PLoS One. 2016 Jul 6;11(7):e0158765. doi: 10.1371/journal.pone.0158765.

[19] Jha V, Garcia-Garcia G, Iseki K, Li Z, Naicker S, Plattner B, et al. Chronic kidney disease: global dimension and perspectives. Lancet. 2013 Jul 20;382(9888):260-72. doi: 10.1016/S0140-6736(13)60687-X.

[20] National Kidney Foundation. KDOQI Clinical Practice Guideline for Diabetes and CKD: 2012 Update. Am J Kidney Dis. 2012 Nov;60(5):850-86. doi: 10.1053/j.ajkd.2012.07.005.