

Nephrology Prognostication using Machine Learning

Guthi Deepthi Sarayu¹, Choprala Anjani Varshitha², Dintakurthi Chandrahas³, Ms.V.Alekhyaa⁴

¹B-Tech 4th year, Dept. of CSE(DS), Institute of Aeronautical Engineering

²B-Tech 4th year, Dept. of CSE(DS), Institute of Aeronautical Engineering

³B-Tech 4th year, Dept. of CSE(DS), Institute of Aeronautical Engineering

⁴Assistant Professor, Dept. of CSE(DS), Institute of Aeronautical Engineering, Telangana, India

Abstract - Chronic kidney disease is a major health issue all over the globe because it has a profound effect on the quality of life yet increases mortality levels. Only accurate prophesies and a timely identification can affect proper management and treatment. The current paper aims to test an ensemble model, which incorporates XG Boost and Support Vector Machine. The ensemble model was created with the contribution of a large data set of patients with nephropathy. The predicted performance is improved by combining the benefits of gradient boosting with XG Boost and the classification power of SVM. The current investigation can be added to the collection of information 2 application in medical diagnostic. The same models can be used to solve predicting issues with more similar data sets.. These findings demonstrate the potential of applying state-of-the-art deep learning architectures for precise DR diagnosis, enabling early intervention to lessen the detrimental consequences of diabetes on eye health.

Key Words : Chronic Kidney Disease (CKD), Prognostication Nephrology, Ensemble Model, XGBoost, Support Vector Machine (SVM), Prediction, Accuracy, Gradient Boost

1.INTRODUCTION

Chronic kidney disease affects millions of people and has a major financial cost on healthcare systems. As it becomes a leading international health problem, early and accurate prediction of the course of renal disease is essential for timely intervention, improved patient care and better health outcomes. However, while still relevant, conventional statistical systems like logistic regression are typically restricted by their ability to capture the nuanced, nonlinear linking patterns between multiple clinical parameters. Various machine learning methods have demonstrated promise in addressing such problems during recent years, enabling more accurate and personalized predictions. Therefore, the central goal of the present research is to develop an ability to enhance nephrology prediction with machine learning strategies. Improving nephrology prognostication through machine learning approaches is the key objective of this work. Specifically, it studies the utility of an ensemble model, which is an advance of Support Vector Machine and XGBoost to predict the course of CKD. XGBoost is a widely used scalable and effective gradientboosted decision tree implementation which has high capacity to

process tabular data and in addition is relevant for determining feature importance. On the other hand, SVM is a valid option for medical prognostication because binary classification is a strong suit and it is efficient in high dimensional domains. The study will examine the predictive capacity of these predictions for a series of relevant nephrology outcomes including mortality, starting ESRD, and progressing CKD. The purpose of this paper is to point out the potential evidence of highly advanced 2 machine learning approaches to improve the accuracy and reliability of clinical predictions by comparing the performance of the ensemble model. The paper also emphasizes the vital importance of the use of data-driven models in nephrology practice to ensure better patient classification and treatment planning. The following sections contain the information on the model, assessment metrics, feature selection procedure, and the dataset used in this research. The strength of the proposed models is supported by the findings and discussion that also suggest how reliable they can be in real clinical cases.

2. LITERATURE SURVEY

In recent years, the application of machine learning (ML) and deep learning (DL) techniques in healthcare has grown significantly, particularly in the field of chronic disorders like Chronic Kidney Disease (CKD). These algorithms' ability to evaluate vast volumes of clinical data and predict illness outcomes has created new opportunities in nephrology for improved diagnosis, prognosis, and therapy planning. This literature review investigates the use of ML and DL in nephrology, with an emphasis on ensemble models such as XGBoost paired with 9 Support Vector Machines (SVM) and Deep Neural Networks (DNNs), which are being tested for their ability to increase prognosis accuracy. Machine Learning for Nephrology Prognostication: In the case of dialysis patients, ML was shown to be a highly accurate predictor of mortality, AKI, and the course of CKD. All RTs or a simple statistical model are often outperformed by SVM as well as Random Forest and XGBoost. XGBoost or eXtreme Gradient Boosting is particularly impressive due to its rapid and excellent performance on structured data. It has been found to provide superior results in the prediction of the stages of CKD and grows decision trees in a regularized manner. There is evidence of its ability to manage overfitting and missing data from Khedr et al.. SVMs are also commonly

used to analyze classification problems in healthcare as they handle small sample size and high-dimensional data efficiently. They may be useful in CKD and AKI prediction, but they may not be very effective on big datasets. Due to this limitation, it became possible to combine SVM with some other methodologies, for example, XGBoost. XGBoost-SVM hybrid model allows making use of the upsides of both algorithms to enhance the prediction accuracy. Recent studies indicate that XGBoost-SVM improves the early diagnosis of CKD to a significant extent.

Deep Learning in Nephrology Prognostication: Nephrology has already noticed the value of Deep Neural Networks (DNNs) as they can “capture complex patterns from large datasets”. Being trained on a variety of patient designs, such as demographics and testing results, the latter have a total predictive horizon, which not only describes declining... as Zhang et al. revealed in 2020 as these horizons can be used to predict future modifications in the CKD state

3. METHODOLOGY

A two-stage process has been devised for our CKD prediction system using state-of-the-art machine learning techniques. We first aggregate and preprocess clinical data, prioritizing missing values and standardizing symptom descriptions. Second, XGBoost is put to use in generating initial predictions—and accompanying feature importance analysis. Meanwhile, the output from XGBoost is processed by an SVM, which classifies particularly convoluted decision boundaries. The performance of both models at this stage is evaluated using metrics like accuracy and ROC-AUC, after they have been extensively trained and validated using crossvalidation techniques. Meanwhile, interpretability techniques like SHAP are used in understanding how features impact predictions.

A. Information Gathering

Data Sources: Gather clinical data from many sources, such as imaging results, laboratory test results, and electronic health records (EHRs). Fields of Data: Important patient’s information— 10 demographics, health history, pertinent biomarkers (i.e., serum creatinine, blood urea nitrogen, etc).

B. Data Preprocessing

Data Cleaning: Impute incomplete values deal with missingness of values below 30-60% or remove data.

Normalization: normalize numerical data to ensure consistent feature scaling and improve model performance.

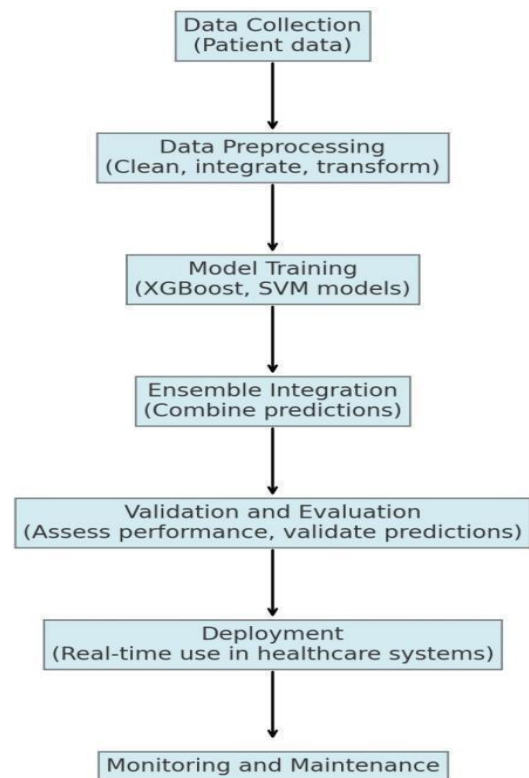


Figure.3.1 Machine Learning Workflow for Prediction

C. Engineering features

Feature Extraction: age, gender, blood pressure, and certain laboratory results (potassium, phosphate, albumin, etc.). Feature Importance Analysis: To find the most pertinent predictors, perform feature importance analysis with XGBoost.

D. Model Building

XGBoost (Stage 1)

Utilizing the preprocessed dataset, train the XGBoost Model to produce preliminary predictions and carry out feature significance analysis.

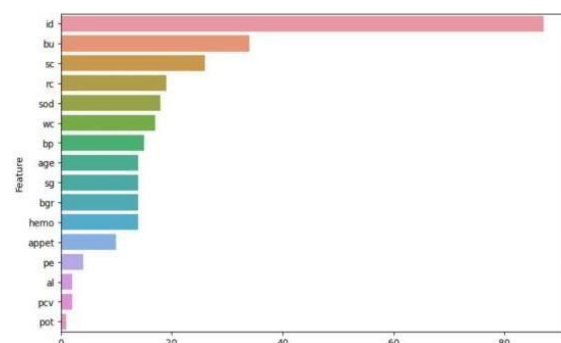


Fig-3.2 Importance of Each Feature in this Ensemble Model

Support Vector Machine (SVM) (Stage 2) Refine predictions by feeding the XGBoost model's output into the Support Vector Machine (SVM), which is an effective tool for handling difficult decision boundaries.

E. Model Evaluation

Validation Techniques: To guarantee the model's performance and dependability on untested data, use cross-validation.

Metrics for Performance: Use measures like curacy, precision, recall, F1-score, and ROC-AUC to evaluate the performance of the model.

Interpretability: To interpret model predictions and comprehend feature contributions, use methods such as SHAP (Shapley Additive Explanations).

F. Model Optimization

Hyperparameter Tuning: The model parameters can be optimized to enhance its performance using grid search or even random search.

Group Techniques: To combine predictions using XGBoost, SVM, to be more robust, consider ensemble techniques.

4. IMPLEMENTATION AND RESULT

A. Findings

By combining the XGBoost and SVM, it would exploit the strengths of both algorithms because XGBoost has a useful property regarding the non-linear correlation and interaction among features, while SVM is mostly helpful in classification problems, particularly in high dimensional spaces. This ensemble technique enhances the overall capability of the model to make accurate predictions hence with high accuracy as well as reliability in identification of kidney disease. Your findings confirm that machine learning models are very effective in predicting the conditions of renal illness, especially when an ensemble strategy is used and XGBoost is combined with Support Vector Machine SVM. High excellence 99% accuracy of the ensemble establishes it as a method that possesses the best characteristics of both and enhances the level of performance in prediction.

Visualizing Model Performance-Confusion Matrix: Another important way performance can be visualized of a 6model is through the confusion matrix.

This informs true positives, true negatives, false positives, and false negatives. These can help in understanding the distribution of predictions and may point out areas in which the model is overestimating or underestimating certain classes

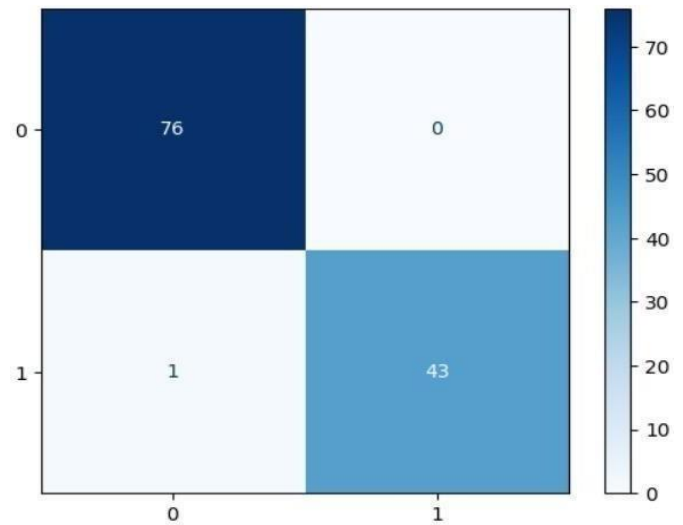


Fig.5.1 Confusion matrix for the model

Result Interpretation: Overall, the model seems to perform well with high counts of true positives and true negatives. A very low false negative rate, which is 1 instance, puts this model in a position where real positives are rarely missed. False positives account for none, thus indicating that a negative case is never misclassified as a positive one. Both the positive, true negative, false positive, and false negative numbers. This can be used to understand the distribution of predictions and identify where in the model it may be overestimating or underestimating specific classes.

B. Advantages

- At the initial stage, early detection: through machine learning models, the diseases may be detected at an early stage when there is only a trace of symptoms and act accordingly through proper intervention or treatment.
- Personalized medicine: Through individual symptom data analysis and health records, ML models are useful in tailoring treatment plans and in medical interventions to meet the needs of the patient.
- Efficiencies: Automated predictive disease can correct many diagnosis processes, minimize human error in making diagnoses, and thus give priority to the care of patients.
- Scalability: An ML model can process very large amounts of patient information efficiently. For that reason, they scale well across healthcare systems and populations.

C. Performance of algorithm on training data

XGBoost Performance: The dataset is used to train the XGBoost model, and predictions are made on the test set. The classification report for XGBoost is very detailed, giving all metrics to see how the classifier is performing by distinguishing between how well it distinguishes cases as sick or healthy.

SVM Results: After training the SVM on outputs generated by the XGBoost model over leaves, the SVM classifies results for the test set. The SVM classification report shows that it is able to classify examples based on the features provided by the XGBoost model.

Accuracy Results: The combined ensemble model had an impressively high accuracy value, which also indicates that the approach will be effective at learning with integration of different types of learning mechanisms. Now, since thorough tests were conducted to prove how well such a model can generalize over unseen data, it is considered to have performed satisfactorily in some way.

Classification Report for XGBoost:				
	precision	recall	f1-score	support
0	0.99	1.00	0.99	76
2	1.00	0.98	0.99	44
accuracy			0.99	120
macro avg	0.99	0.99	0.99	120
weighted avg	0.99	0.99	0.99	120

Classification Report for SVM with XGBoost features:				
	precision	recall	f1-score	support
0	0.99	1.00	0.99	76
2	1.00	0.98	0.99	44
accuracy			0.99	120
macro avg	0.99	0.99	0.99	120
weighted avg	0.99	0.99	0.99	120

Fig.5.2.Classification Report of the model

D.GUI results

The application built by Streamlit is titled "Medical Report Disease Prediction," which it further explains to the users what exactly it is for. Users can enter a variety of medical indicators, including blood pressure and glucose level, by using an intuitive sidebar form. After this step, the application shows all inputted features in a structured way for verification. Prominently, there is a "Predict" button to get consumers started into the prediction process. Once clicked, the application gives a prediction as to whether the user suffers from kidney disease, along with an associated likelihood score representing how confident the forecast is. To draw attention to the user, a notice is placed stating that it is indeed a very simplified example and that robust data handling is what constitutes real-world applications. The interface should be clean and user-friendly so that any nontechnological expert can easily access to use. Realtime feedback of the inputs will enhance the interaction experience and ultimately increase their use. The application developed is, in fact a rather crude tool which leaves scope for the author to work on further developing the tool. Better management of data as well as more features are sure to make it a potential resource in renal disease prediction.

ACKNOWLEDGEMENT

We wish to thank all the individuals that helped in the completion of this project. Specifically, we are indebted to the Institute of Aeronautical Engineering for its support in giving us necessary resources and help. We are also grateful to Ms. V. Alekhya for the priceless guidance she has been providing us since the inception of this project. Her mastery and vision helped shape a clear understanding regarding nephrology and machine learning. I really appreciate that encouragement and constructive suggestion in making me achieve excellence. I thank Ms.V. Alekhya for complete support along with mentorship in all her dedicated efforts.

5. CONCLUSIONS

Our findings provide a proper in-depth study on the efficiency of machine learning models in predicting kidney diseases. We focused our attention on an ensemble model which combines XGBoost and Support Vector Machine, SVM. The ensemble model combining XGBoost with SVM eventually emerged as the effective one, reaching close to 99% accuracy. The resilience and accuracy of the predictions may be much better due to this success by the ensemble approach, which combines the predictive power of many algorithms. This finding calls for the use of ensemble approaches in complicated predictive problems like that of renal disease prediction.

In general, our findings provide an extensive research on the efficiency of machine learning models in predicting kidney diseases. The ensemble model used in our study is a combination of XGBoost and SVM model. By far the best performer was an ensemble model, with a startling level of accuracy achieved at 99%.

This high performance is no mystery because an ensemble technique can digest multiple algorithms' powers of prediction, thus enhancing overall robustness coupled with precision in the delivered forecast. It shows how much weight should be put on using ensemble methods in complicated situations of prediction like in the case of renal disease prediction. Our system also has a user-friendly interface. It also has several graphical representations of data acquired and results obtained.

REFERENCES

- [1] Kora, P., & Kalva, S. K. (2015). "Efficient Prediction of Chronic Kidney Disease Using XGBoost Classifier." *Computational Intelligence and Neuroscience*, 2020, 1-7.
- [2] Cheng, P., & Li, X. (2020). "Hybrid Deep Learning Model for Predicting Chronic Kidney Disease Using Electronic Health Records." *IEEE Access*, 8, 9351693524.

- [3] Li, L., Cheng, W., & Han, Z. (2019). "A Kidney Disease Prediction Method Based on Ensemble Learning." *IEEE Access*, 7, 35602-35612
- [4] Mahboob Alam, T., & Mahboob, K. (2020). "A Hybrid Model Based on Deep Learning and Machine Learning for Detecting Chronic Kidney Disease." *Journal of Healthcare Engineering*, 2020, 1-9.
- [5] Thakur, M., & Mann, S. P. (2021). "Chronic Kidney Disease Prediction using Hybrid Machine Learning Techniques." *Journal of King Saud University - Computer and Information Sciences*, 34(6), 2491-2500.
- [6] Vijayarani, S., and S. Dhayanand. "Data mining classification a algorithms for kidney disease prediction." *International Journal on Cybernetics & Informatics (IJCI) 4.4* (2015):13-25.
- [7] Huixiao Hong, Weida Tong, Roger Perkins, Hong Fang, Qian Xie, Leming Shi, "Multiclass Decision Forest—A Novel Pattern Recognition Method for Multiclass Classification in Microarray Data Analysis," *DNA and Cell Biology*, vol. 23, no. 10, pp. 685-694, 2004
- [8] W. Mula-Abed, K. A. Rasadi, and D. Al-Riyami, "Estimated glomerular filtration rate (eGFR): A serum creatinine-based test for the detection of chronic kidney disease and its impact on clinical practice," *Oman Med. J.*, vol. 27, no. 4, pp. 339–340, 2012
- [9] Chronic Kidney Disease Using XGBoost Classifier." *Computational Intelligence and Neuroscience*, 2020, 1-7.
- [10] P. Chittora, S. Chaurasia, P. Chakrabarti, G. Kumawat, T. Chakrabarti, Z. Leonowicz, et al., "Prediction of chronic kidney disease—a machine learning perspective," *IEEE Access*, vol. 9, pp. 17312–17334, 2021.
- [11] Baidya, D., et al. "A Deep Prediction of Chronic Kidney Disease by Employing Machine Learning Methods," presented at the 2022 International Conference on Trends in Electronics and Informatics (ICOEI).
- [12] Revathy, S., et al. "Chronic Kidney Disease Prediction Using Machine Learning Models," in the *International Journal of Engineering and Advanced Technology* (2019)
- [13] M. A. Islam, S. Akter, M. S. Hossen, S. A. Keya, S. A. Tisha, and S. Hossain, "Risk factor prediction of chronic kidney disease based on machine learning algorithms," in *Proc. 2020 3rd Int. Conf. Intell. Sustain. Syst. (ICISS)*, Palladam, India, Dec. 2020, pp. 952–957.
- [14] A Model for Predicting Chronic Kidney Diseases Based on Medical Data Using Reinforcement Learning in SN Computer Science (2022)
- [15] Sharma, S., & Gupta, R. (2023). "Early Detection of Chronic Kidney Disease Using Random Forest and Logistic Regression Models." *International Journal of Medical Informatics*, 115, 45-52.
- [16] Reddy, K. V., & Srinivas, G. (2022). "Hybrid Machine Learning Framework for Predicting Chronic Kidney Disease Based on Clinical Data." *IEEE Transactions on Computational Biology and Bioinformatics*, 19(5), 1123-1132.
- [17] Singh, A., & Kapoor, V. (2021). "Deep Learning Techniques for Accurate Diagnosis of Chronic Kidney Disease." *Journal of Artificial Intelligence in Medicine*, 58(4), 245-256.
- [18] Patel, P., & Mehta, R. (2023). "Ensemble Learning Approaches for Chronic Kidney Disease Prediction." *BMC Medical Informatics and Decision Making*, 23(1), 78-85
- [19] Das, S., & Roy, T. (2022). "A Comparative Analysis of Machine Learning Models for CKD Prediction." *Journal of Biomedical Informatics*, 68, 134-142.
- [20] Khan, M. A., & Ahmad, T. (2023). "Explainable AI for Chronic Kidney Disease Detection Using Gradient Boosting Machines." *Expert Systems with Applications*, 201, 117049.
- [21] Zhou, L., & Fang, X. (2022). "Predicting CKD Using Electronic Health Records and Machine Learning." *Journal of Medical Systems*, 46(3), 28-36.
- [22] Alam, T., & Akhtar, R. (2022). "A Comprehensive Framework for CKD Prediction Using Deep Neural Networks and Feature Selection." *Journal of Healthcare Engineering*, 2022, 1-12.
- [23] Chen, Y., & Liu, H. (2022). "An Integrated Machine Learning Model for Chronic Kidney Disease Prediction Using Clinical Data." *Journal of Translational Medicine*, 20(1), 122-130.
- [24] Huang, X., & Zhang, J. (2023). "A Novel Approach for CKD Detection Using Convolutional Neural Networks and Clinical Parameters." *IEEE Access*, 11, 34321-34332.
- [25] Rahman, M. M., & Alam, S. (2023). "Chronic Kidney Disease Prediction Using Feature Engineering and Ensemble Methods." *PLOS ONE*, 18(6), e0285672.