iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 12 **Issue:** I **Month of publication:** January 2024

DOI: https://doi.org/10.22214/ijraset.2024.58102

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 12 Issue I Jan 2024- Available at www.ijraset.com

Predictive Analysis in Nephrology: Evaluating Machine Learning Models for Chronic Kidney Disease Prediction

Metun[1], Geethanjali P[2], Ajay V[3]
[1]*Graduate Student Researcher, Department of Computer Science, New York University, Brooklyn, New York - 11201*
[2]*Associate Software Engineer, Bosch Global Software Technologies, Bangalore, India*
[3]*Senior Associate Engineer, Price Waterhouse Coopers, Bangalore, India*

Abstract: ***In the realm of healthcare, early detection and accurate diagnosis are critical for effective treatment, especially for diseases that are often undiagnosed until later stages. This research presents an Automated Classification System focused on Chronic Kidney Disease (CKD), a prevalent condition frequently undetected in its early stages. Leveraging advanced machine learning techniques, the study aims to overcome the traditional reliance on doctors' intuition and experience by providing a data-driven approach for CKD diagnosis. Utilizing a comprehensive dataset, the research explores various data pre-processing methods, including innovative approaches for handling missing values and data integrity, which surpass the conventional mean and mode-based imputation methods prevalent in existing literature. Through exploratory data analysis and the application of diverse machine learning algorithms, the study seeks to identify the most effective model for predicting CKD risk. The findings of this research have significant implications for enhancing clinical decisions, offering a path towards more accessible, efficient, and accurate diagnosis of Chronic Kidney Disease, ultimately contributing to improved patient outcomes and healthcare practices.***

I. INTRODUCTION

Chronic Kidney Disease (CKD) presents a critical global health challenge, particularly due to its often asymptomatic nature in initial stages, which results in a significant number of undiagnosed cases. The National Library of Medicine highlights a concerning statistic: approximately 90% of individuals with early-stage CKD are unaware of their condition. This unawareness, combined with restricted access to contemporary digital healthcare solutions, further compounds the problem. Historically, clinical decision-making in this area has predominantly depended on the expertise and intuition of healthcare professionals, as opposed to detailed data-driven analysis.

The progression of CKD is marked by gradual deterioration of kidney function, leading to the inability of the kidneys to filter blood effectively. This impairment results in the buildup of waste materials in the body. The slow development of the disease categorizes it as chronic and it affects individuals worldwide. Various health issues, such as diabetes, hypertension, and heart diseases, are known to contribute to the onset of CKD. Additional risk factors include age and gender. Symptoms like back and abdominal pain, gastrointestinal issues, fever, bleeding disorders, skin eruptions, and vomiting may manifest as the condition worsens. Diabetes and hypertension are identified as the leading causes of irreversible kidney damage, underscoring the importance of managing these conditions to prevent the onset of CKD. The deceptive nature of CKD often leads to late diagnosis, typically at an advanced stage of the disease.

A. Stages of CKD:

1) *Initial Stages:* CKD initially may not show any symptoms, allowing the body to compensate for the loss of kidney function. As a result, early-stage CKD is often undetected unless routine health check-ups for other conditions reveal kidney-related anomalies. Early detection allows for interventions such as medication and regular check-ups to prevent the disease from progressing.
2) *Advanced Stages:* As CKD progresses, various symptoms indicating deteriorating kidney function may appear.
3) *End-Stage:* The final stage of CKD, or kidney failure, may require treatments like dialysis or a kidney transplant.

©IJRASET: All Rights are Reserved

B. Diagnostic Tools for CKD:

CKD arises from ongoing kidney damage, often secondary to other health conditions. Hospital admissions for CKD have been increasing, despite a stable global death rate. Key diagnostic tests for CKD include:

1) *Estimated Glomerular Filtration Rate (eGFR):* Measures kidney filtration efficiency. An eGFR above 90 indicates healthy kidney function, while below 60 suggests CKD.
2) *Urine Analysis:* Evaluates kidney function. Blood and protein in urine indicate abnormal kidney activity.
3) *Blood Pressure:* Assesses heart function and its impact on the kidneys. An eGFR below 15 indicates end-stage renal disease.
4) *Additional Testing:* Methods like ultrasounds, MRI, or CT scans, and kidney biopsies are utilized for further assessment of kidney damage and diagnosis of specific kidney conditions.

The application of advanced analytical techniques, such as data mining, in medicine plays a crucial role in extracting hidden insights from extensive patient medical records and treatment data. This approach enhances the accuracy of symptom analysis and treatment planning.The primary objective of this research is to develop an Automated Classification System for Chronic Kidney Disease using machine learning algorithms. The goal is to enhance clinical decision-making by employing data-driven methods for the early detection and accurate diagnosis of CKD. This involves:

- Utilizing a comprehensive dataset to identify and understand the key factors influencing CKD.
- Implementing advanced data pre-processing techniques to manage missing values and ensure data integrity.
- Exploring and comparing various machine learning models to determine the most effective algorithm for predicting the risk of CKD in patients.

II. LITERATURE REVIEW

Table 1 shows an overview of the literature on chronic kidney disease prediction using several machine learning algorithms.

Table 1 Literature Survey

Title	Author and year	Proposed solution	Outcome	Benefits	Drawbacks
A Machine Learning Methodology for Diagnosing Chronic Kidney Disease[1]	Jiongming Qin1, Lin Chen2, Yuhua Liu1, Chuanjun Liu2, Chang Hao Feng1, Bin Chen1 Year - December 2019	Train and test the CKD data on different classification ML models.[1]	Random Forest – 99.75%	Due to data pre-processing and feature engineering, the system was designed in less time.	Because of the restricted dataset, the model's generalization ability may be limited, and it cannot predict the severity of CKD (due to only 2 columns CKD and not CKD)
Comparative Study of Classifier for Chronic Kidney Disease prediction using Naive Bayes, KNN and Random Forest [2]	Devika R, Sai Vaishnavi Avilala, V. Subramaniyaswamy Year- March 2019	Use 3 different classification techniques - Naive Bayes, KNN, and random forest to predict the presence of chronic renal failure, and performance is calculated based on their accuracy, F1, and precision.	1)Random Forest – 99.84% 2)Naive Bayes - 99.635%	The 3 classifiers used gave more accuracy than other classifiers and the more accurate result was achieved due to the use of statistical mining.	The model could not predict the severity of chronic renal failure.
A Predictive Analysis of	Mafizur Rahman, Jannatul Ferdous	ML models used - naive Bayes, KNN,	SVM, Extra Tree Classifier – 98%	In this work, we used PCA	There should be new models

Chronic Kidney Disease by Exploring Important Features [3]	Sorna, Masud Rana, Linta Islam, Malika Zannat Tazim, Syada Tasmia Alvi Year - 2020	SVM, and Logistic Regression.		methods to minimize data dimensionality and improve renal illness prediction.	available that disclose more information about this condition with higher precision.
A diagnostic prediction model for chronic kidney disease in the internet of things platform [4]	Mehdi Hosseinzadeh, Jalil Koohpayehzadeh, Ahmed Omar Bali, Parvaneh Asghari, Alireza Souri, Ali Mazaherinezhad, Mahdi Bohlouli, Reza Rawassizadeh, Year- 2020	Using sensor technology in the IoT platform, they suggested this model.	Decision tree classifier (J48) – 97%	In addition, when compared to other datasets with various characteristics, the feature set employed can enhance the execution time.	Instances can be increased in the future to make the model more effective.

III. METHODOLOGY

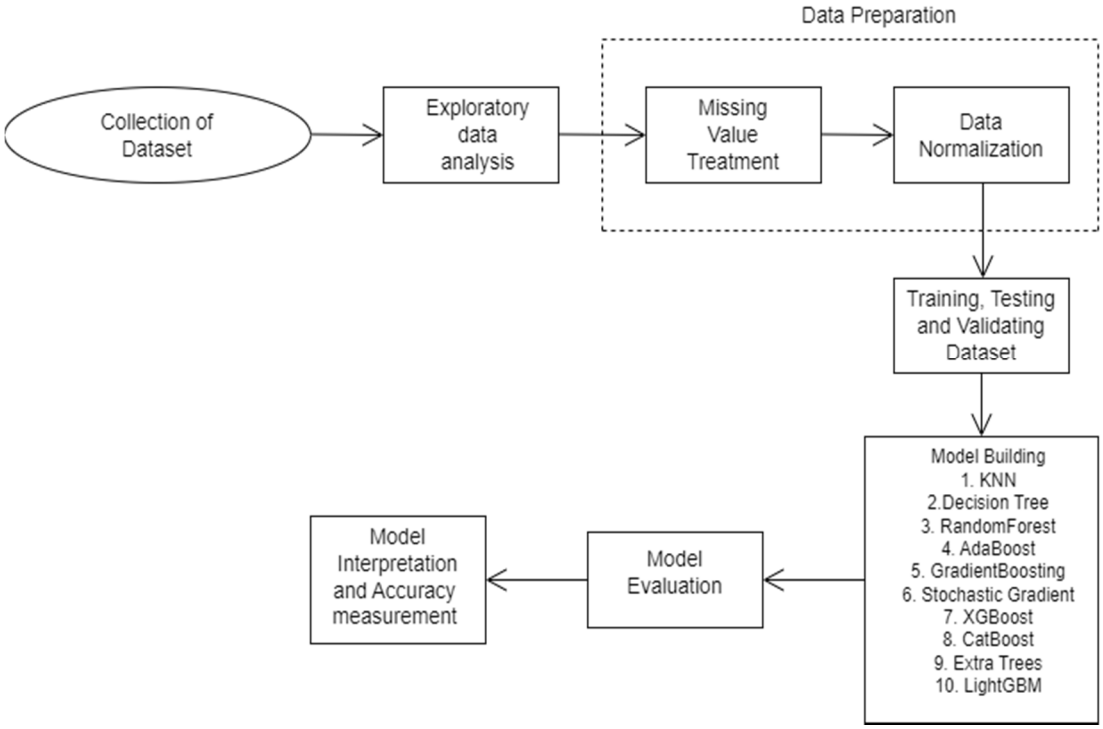


Figure 1 Proposed methodology of Chronic Kidney Disease

The proposed methodology for prediction of the Chronic Kidney Disease is shown in the Figure 1. The key steps are as follows:

A. Collection of Dataset

The study utilizes a dataset specifically tailored for Chronic Kidney Disease, which includes a wide range of variables pertinent to CKD diagnosis. The dataset is sourced from a reliable medical database, Kidney disease UCI dataset ensuring the validity and relevance of the data for this research[5]. The dataset contains 25 attributes among which 14 attributes are nominal and 11 attributes are numeric which is shown in Table 2.

Table 2 Description of Dataset

Sl. No	Attributes	Description	Type	Values
1	age	Patient's Age	Numerical	in years
2	bp	Blood Pressure	Numerical	in mm/Hg
3	sg	Specific Gravity	Nominal	(1.005,1.010,1.015,1.020,1.025)
4	al	Albumin	Nominal	(0,1,2,3,4,5)
5	su	Sugar	Nominal	(0,1,2,3,4,5)
6	rbc	Red Blood Cells	Nominal	(normal,abnormal)
7	pc	Pus Cell	Nominal	(normal,abnormal)
8	pcc	Pus Cell clumps	Nominal	(present,notpresent)
9	ba	Bacteria	Nominal	(present,notpresent)
10	bgr	Blood Glucose Random	Numerical	in mgs/dl
11	bu	Blood Urea	Numerical	in mgs/dl
12	sc	Serum Creatinine	Numerical	in mgs/dl
13	sod	Sodium	Numerical	in mEq/L
14	pot	Potassium	Numerical	in mEq/L
15	hemo	Hemoglobin	Numerical	in gms
16	pcv	Packed Cell Volume	Numerical	-
17	wc	White Blood Cell Count	Numerical	in cells/cumm
18	rc	Red Blood Cell Count	Numerical	in millions/cmm
19	htn	Hypertension	Nominal	(yes,no)
20	dm	Diabetes Mellitus	Nominal	(yes,no)
21	cad	Coronary Artery Disease	Nominal	(yes,no)
22	appet	Appetite	Nominal	(good,poor)
23	pe	Pedal Edema	Nominal	(yes,no)
24	ane	Anemia	Nominal	(yes,no)
25	class	Class	Nominal	(ckd,notckd)

B. Exploratory Data Analysis

Before delving into model building, an extensive exploratory data analysis was conducted. This included identifying unique values in categorical columns and visualizing key patterns and correlations within the data, thereby gaining valuable insights into the factors influencing CKD[6].

1) *Categorical Feature Distribution Graphs:* These bar charts displayed in Figure 2 the distribution of categorical variables like pus cell clumps, bacteria, hypertension, diabetes mellitus, and others within the dataset, offering insights into the frequency and distribution of these features.

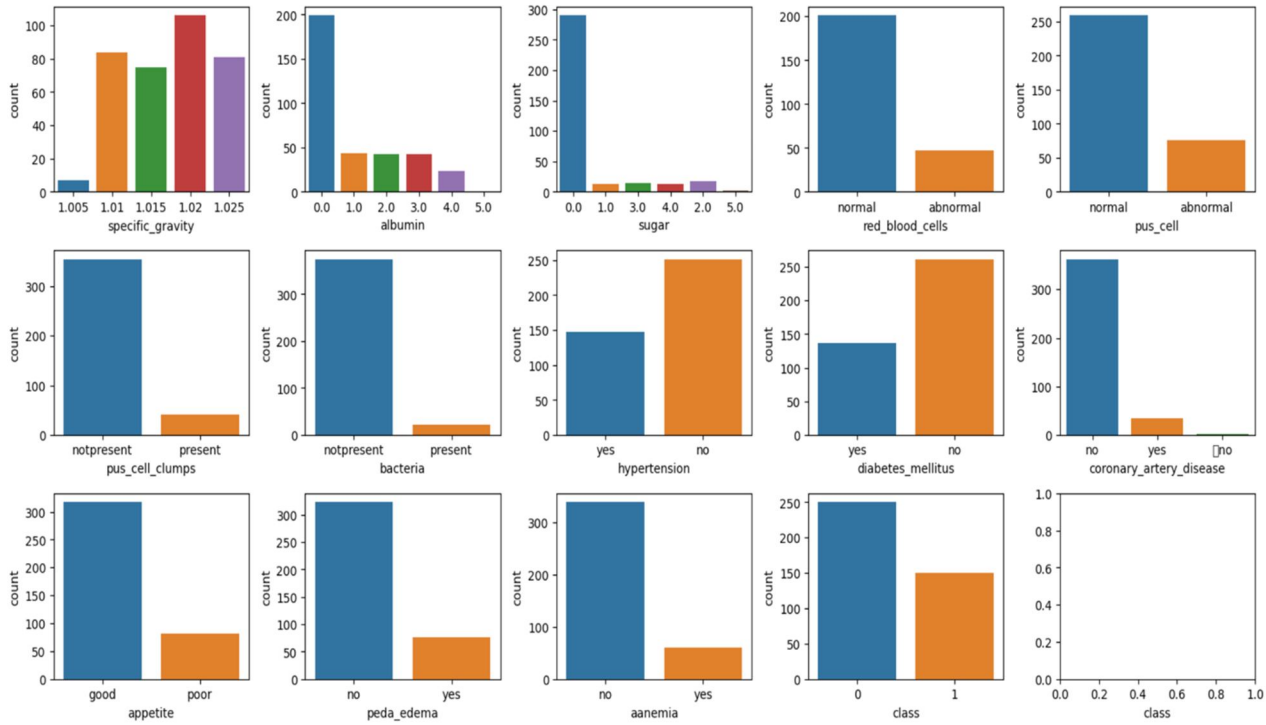


Figure 2 Graphical representation of Distribution of Categorical columns

2) *Numerical Feature Distribution Graphs:* Kernel Density Estimation (KDE) plots for numerical variables such as age, blood pressure, blood urea, serum creatinine, and others provide a visual representation of the data distribution, which is essential for identifying patterns and potential outliers as shown in Figure 3.

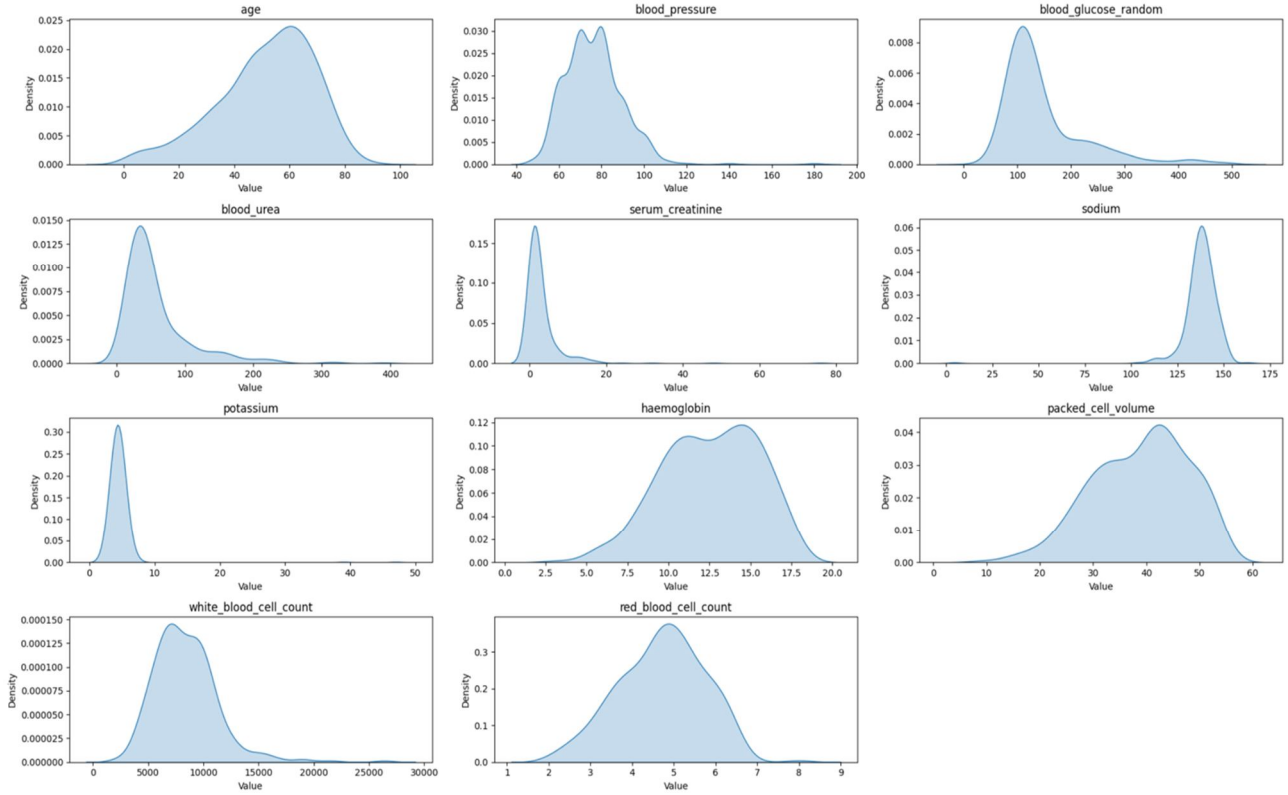


Figure 3 Graphical representation of Distribution of Numerical columns

3) *Correlation Heatmap:* The heatmap of correlation coefficients shown in Figure 4 among the features offers insights into potential multicollinearity and the strength of linear relationships between feature pairs, which is pivotal in understanding the interdependencies within the dataset[7].

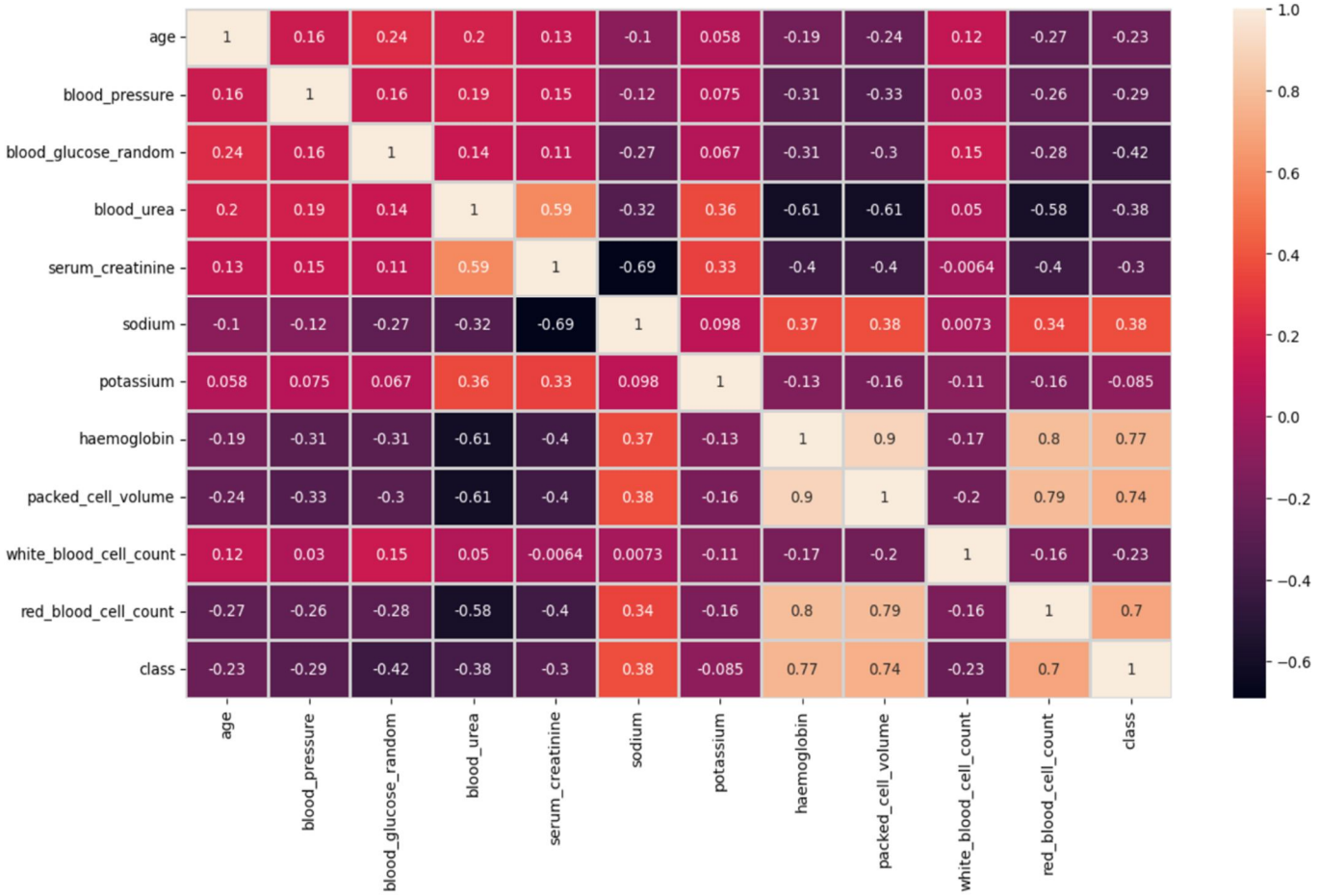


Figure 4 Correlation Heatmap

C. *Data Preparation*

1) *Data Cleaning and Pre-processing***:** The initial step involved cleaning the dataset by removing irrelevant columns, such as 'id', and renaming the remaining columns for better readability and understanding. Data types of various columns were adjusted for appropriate analysis, with specific focus on converting certain columns to categorical types and handling numerical columns with special care to address errors and inconsistencies.
2) *Handling Missing Values:* A critical aspect of the pre-processing phase involved developing a novel approach to manage missing values in the dataset. Moving beyond the traditional mean and mode-based imputation methods, the study employs more advanced techniques to maintain the integrity and usefulness of the data for machine learning models.[8]

D. *Model Training and Validation*:

In the model training and validation phase, the `train_test_split` method from scikit-learn is utilized to bifurcate the dataset into training and testing subsets. The training subset, constituting 70% of the original data, was employed to train the models, allowing them to learn and adapt to the data's underlying patterns. The remaining 30%, marked for testing, served as a benchmark to impartially assess the predictive prowess of the models. A `random_state` parameter was set to 45, which guaranteed reproducibility in the results, ensuring that the random partitioning of the data remains consistent across different executions. This rigorous approach allowed us to train these models thoroughly while maintaining an objective lens for validation, thereby ensuring an accurate evaluation of their performance.

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 12 Issue I Jan 2024- Available at www.ijraset.com

E. Model Development and Evaluation

1) *Selection of Machine Learning Algorithms:* The study explores a range of machine learning algorithms to identify the most effective model for CKD prediction. This includes both classical algorithms and more advanced techniques, ensuring a comprehensive evaluation of potential models.

a) *K-Nearest Neighbors (KNN):* The KNN method is a straightforward supervised learning technique frequently utilized to address classification and regression problems. In KNN, the decision-making process involves computing the Euclidean distances between a given query and every data point. It then selects the value of k, representing the nearest neighbors to the query, and either picks the most frequent label for classification or calculates the label's average for regression. Nevertheless, this approach is not suitable for dynamic web mining due to its reliance on lazy learning and the necessity of a well-chosen k value. In the context of dynamic web mining, the value of k is automatically determined to enhance the accuracy of the KNN algorithm.[9].

b) *Decision Tree:* When it comes to addressing classification problems, one of the most effective and widely utilized techniques in supervised machine learning is the decision tree. A decision tree is a type of hierarchical structure akin to a flowchart. In this structure, each internal node represents a test applied to a feature, each branch corresponds to the result of that test, and each leaf node contains a class label. The decision tree begins at the root node, evaluates various variables' values, and proceed along branches until it reaches a leaf node. In the context of classification tasks, the decision tree poses inquiries and based on the responses, subdivides the data into subsequent branches. It employs several methods to analyze the population's divisions and parameters that lead to the creation of the most homogeneous subsets[10].

c) *Random Forest:* The Random Forest classifier is an ensemble machine learning model known for its robustness in classification tasks. It combines multiple decision trees, each trained on a random subset of the dataset. During tree construction, a subset of features is considered at each node, preventing bias towards dominant features. The model's predictions are derived through a majority vote in classification tasks, ensuring high accuracy and reducing overfitting. Random Forests also provide insights into feature importance, aiding in feature selection and interpretability, making them versatile and valuable in various applications[11].

d) *AdaBoost Classifier:* The AdaBoost (Adaptive Boosting) classifier is an ensemble machine learning algorithm that combines the predictions of multiple weak learners to create a strong and accurate classifier. It assigns weights to data points, with a focus on those previously misclassified, and sequentially trains weak learners to minimize weighted classification errors. AdaBoost's final prediction is a weighted majority vote in classification tasks, making it robust and less prone to overfitting. It finds applications in various fields, including face detection and text classification, due to its ability to handle complex datasets and achieve high accuracy.

e) *Gradient Boosting Classifier:* Gradient Boosting is a machine learning technique that sequentially builds an ensemble of decision trees by minimizing errors from previous iterations. It focuses on improving predictions for misclassified data points, making it highly accurate. This method uses gradient descent optimization to adjust tree predictions and combines them to create a robust model. It includes regularization techniques to prevent overfitting and is widely used in various applications due to its versatility and strong predictive capabilities[12].

f) *Stochastic Gradient Boosting:* Stochastic Gradient Boosting (SGB) stands as a robust machine learning technique, harnessing the power of both gradient boosting and stochastic processes. This approach uniquely integrates random subsampling of the training data, not unlike techniques found in bagging, with the sequential addition of decision trees, a hallmark of gradient boosting. SGB's strength lies in its ability to mitigate overfitting, a common challenge in traditional boosting, by introducing randomness into the model training process. This randomness ensures a more generalized and reliable predictive performance on unseen data[13]. The adaptability of SGB allows it to excel in a variety of tasks, from classification to regression, making it a versatile tool in the machine learning arsenal. Its efficacy is particularly noted in scenarios where the data is noisy and complex, showcasing its robustness in real-world applications.

g) *XGBoost Classifier:* XGBoost, or Extreme Gradient Boosting, represents a highly efficient and scalable implementation of gradient boosting. This model has gained widespread popularity due to its performance and speed, especially in structured or tabular data competitions. XGBoost distinguishes itself with its ability to handle sparse data and its efficient use of system resources, making it both powerful and practical for large datasets. The algorithm employs a unique tree learning technique and regularization, which helps in reducing overfitting and improving overall model performance. It is particularly lauded for its versatility, being able to solve a wide range of problems, from regression to classification.

©IJRASET: All Rights are Reserved

Moreover, XGBoost's compatibility with numerous programming languages and its integration with machine learning frameworks like scikit-learn make it a go-to choice for many practitioners in the field.

h) *CatBoost Classifier:* CatBoost, an acronym for Categorical Boosting, is a novel machine learning algorithm that excels in dealing with categorical data. Developed by Yandex, it automates the process of category encoding, a significant advantage over traditional models that require extensive pre-processing. CatBoost is distinguished by its use of oblivious decision trees, which results in more balanced tree structures than classical gradient boosting models. This unique feature enhances the model's generalization capabilities and reduces the likelihood of overfitting. CatBoost also boasts an impressive speed and accuracy, particularly in datasets with a high proportion of categorical variables. Its user-friendly design and compatibility with multiple programming languages make it accessible for both beginners and experts in the field. Furthermore, CatBoost's advanced handling of categorical data without significant loss of time efficiency positions it as a cutting-edge tool in predictive modelling[14].

i) *Extra Trees Classifier:* Extra Trees, short for Extremely Randomized Trees, is a powerful ensemble learning technique that builds on the concept of random forests. This model differentiates itself by the way it splits nodes in the decision trees, using random thresholds for each feature rather than searching for the best possible thresholds like traditional decision trees. This randomness not only reduces the computational burden significantly but also helps in increasing the diversity among the individual trees in the model, which can lead to improved model robustness and generalization. Extra Trees are particularly effective in preventing overfitting, a common challenge in complex models. They are well-suited for both regression and classification tasks, delivering strong performance even with minimal hyperparameter tuning. Moreover, the simplicity and efficiency of Extra Trees make them a popular choice for tackling high-dimensional data, where they often outperform more complex and computationally intensive models.[15]

j) *LightGBM Classifier:* LightGBM, short for Light Gradient Boosting Machine, is a highly efficient gradient boosting framework designed for speed and performance. Developed by Microsoft, it stands out due to its novel approach of using a histogram-based algorithm for decision tree learning, which significantly reduces memory usage and increases the speed of computation. LightGBM is particularly adept at handling large-scale data, offering a scalable solution that doesn't compromise on accuracy. One of its key features is the Gradient-based One-Side Sampling (GOSS) and Exclusive Feature Bundling (EFB), which together reduce the number of data instances and features without significant loss of information. This results in faster training times compared to traditional gradient-boosting models. LightGBM's ability to handle sparse data and support for parallel and GPU learning make it a versatile tool in various domains, particularly for applications requiring quick model iteration and deployment. Its effectiveness in predictive accuracy, coupled with its efficiency, makes LightGBM a favored choice among data science practitioners.

2) *Feature Engineering:* Significant effort was dedicated to feature engineering, where relevant features were identified and transformed to enhance the predictive power of the models[16]. This step is crucial for improving model accuracy and ensuring the models are well-tuned to the specificities of CKD data.

F. Model Evaluation

1) *Performance Metrics:* The models were assessed using appropriate performance metrics, which might include accuracy, precision, recall, and F1-score.[17-19]. These metrics provide a comprehensive understanding of each model's effectiveness in predicting CKD.

$$\text{Accuracy} = \frac{TP+TN}{TP+FN+FP+TN} \quad (1)$$

$$\text{Recall (Sensitivity)} = \frac{TP}{TP+FN} \quad (2)$$

$$\text{Precision} = \frac{TP}{TP+FP} \quad (3)$$

$$F1 \text{ score} = 2 \times \frac{\text{precision} \times \text{recall}}{\text{precision} + \text{recall}} \quad (4)$$

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 12 Issue I Jan 2024- Available at www.ijraset.com

2) *Cross-Validation:* To ensure the robustness and reliability of the results, cross-validation techniques were employed. This method helps in mitigating any overfitting issues and validates the model's performance across different subsets of the dataset.[20]

The methodology outlined in this study represents a meticulous and comprehensive approach to developing a machine learning-based solution for the early detection of Chronic Kidney Disease, ensuring accuracy, reliability, and ethical compliance.

IV. RESULTS & DISCUSSION

The evaluation of various machine learning models for the detection of Chronic Kidney Disease (CKD) yielded the following results in table 3:

Table 3 Performance Metrices of the model.

Models	Training Accuracy	Test Accuracy	Precision	Recall	F1-score
KNN	76.7%	60.83%	69%	56%	62%
Decision Tree	100%	98.3%	99%	99%	99%
Random Forest	99.64%	100%	100%	100%	100%
Ada Boost Classifier	100%	98.3%	99%	99%	99%
Gradient Boosting Classifier	100%	99.16%	100%	99%	99%
Stochastic Gradient Boosting Classifier	100%	99.16%	100%	99%	99%
XGB Classifier	100%	99.16%	100%	99%	99%
Cat Boost Classifier	100%	99.16%	100%	99%	99%
Extra Boost Classifier	100%	100%	100%	100%	100%
LGBM Classifier	100%	99.16%	100%	99%	99%

From the analysis of model performance, it is notable that the KNN classifier isn't very effective; it has much lower accuracy in tests compared to other models, indicating it struggles with complex medical data. However, models like Decision Tree, AdaBoost, Gradient Boosting, Stochastic Gradient Boosting, XGBoost, CatBoost, and LightGBM perform exceptionally well, with accuracies around or above 99%. This high accuracy shows they predict well, but there's a risk they might be too tailored to the training data, a problem known as overfitting. The Extra Trees classifier is a clear example of overfitting, achieving 100% accuracy in both training and tests. While perfect scores seem good, they're not usually practical and might not work well on new, unseen data.

On the other hand, the Random Forest model also achieves very high accuracy, but not always 100%, sometimes dropping to 99%. This variation suggests that Random Forest is more balanced and reliable, possibly better for real-world applications because it's not overly adjusted to just the training data.

V. CONCLUSION

The study embarked on developing an automated classification system for Chronic Kidney Disease (CKD) detection, utilizing a range of machine learning models. The research aimed to address the critical gap in early CKD diagnosis, leveraging data-driven approaches to potentially replace the traditional reliance on intuition and experience in clinical decision-making.

The findings reveal that ensemble methods, particularly the Random Forest and Extra Trees classifiers, performed with remarkable accuracy, achieving perfect scores in test scenarios. However, these results, while promising, warrant caution due to the potential risk of overfitting, which could diminish the models' effectiveness in real-world clinical settings. The significant disparity in the KNN model's performance underscores the necessity of model selection tailored to the specific characteristics of medical data.

Through exploratory data analysis and careful preprocessing, the study also highlighted the importance of understanding the underlying distribution and correlations within the dataset, which are critical factors in model training and feature engineering. The visualizations provided key insights into the data structure, aiding in the identification of influential features and informing the direction of model development.

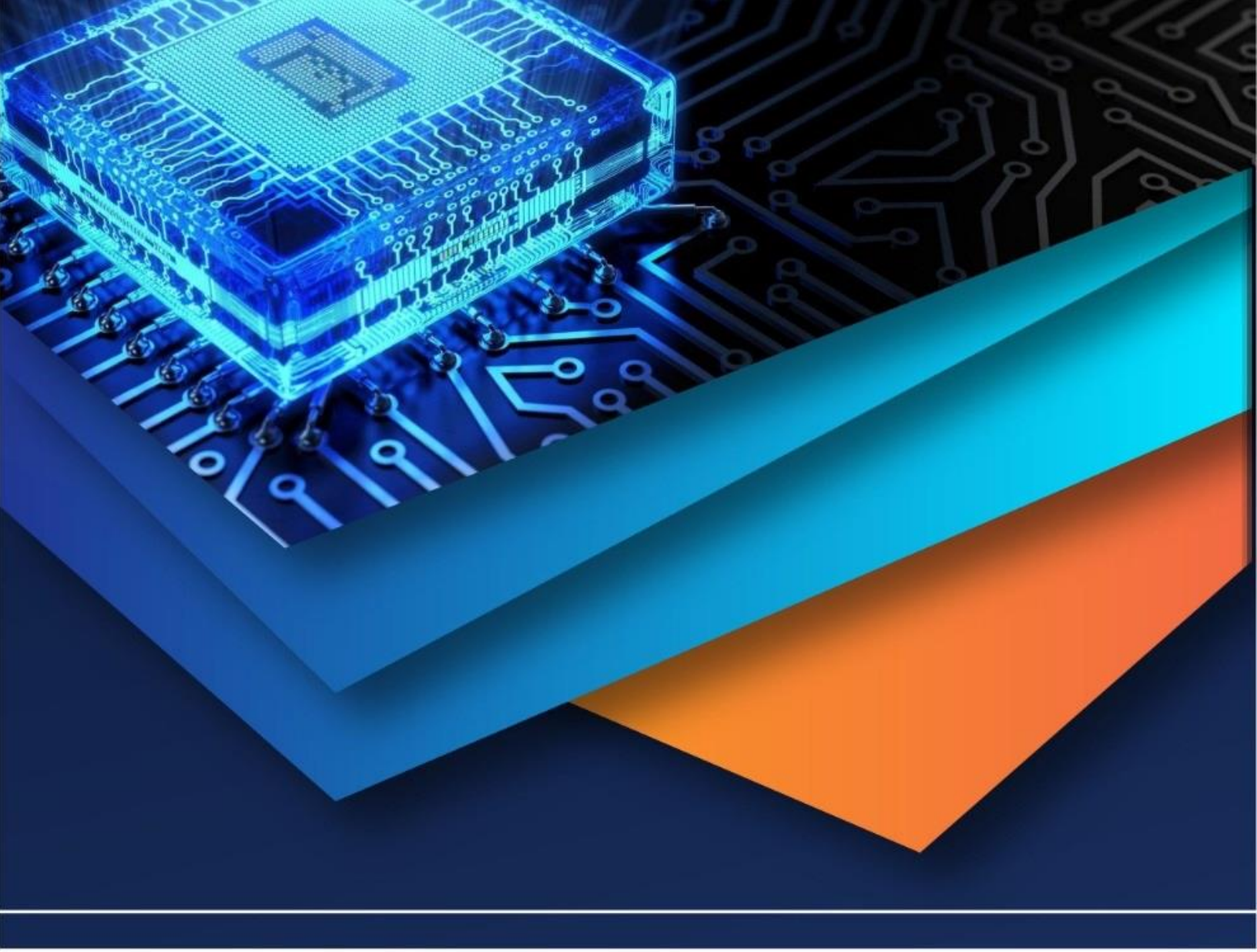
©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

Despite the high accuracy scores, the models' performance must be critically assessed for their practical applicability. It is essential to ensure that the models are robust, generalizable, and applicable in diverse clinical environments. The research advocates for further validation with external datasets and in clinical trials to establish the models' reliability and utility in diagnosing CKD.

In conclusion, this study contributes to the burgeoning field of machine learning in healthcare, offering promising models for CKD detection that could potentially improve patient outcomes through early diagnosis. Nonetheless, the journey from a research setting to clinical practice involves rigorous validation and ethical considerations, ensuring that the technology developed is both effective and equitable when deployed in patient care. Future work will focus on these aspects, striving to bridge the gap between data-driven insights and clinical applicability.

REFERENCES

[1] Qin, J., Chen, L., Liu, Y., Liu, C., Feng, C., & Chen, B. (2020). A Machine Learning Methodology for Diagnosing Chronic Kidney Disease. IEEE Access, 20991–21002.
[2] Devika R, Sai Vaishnavi Avilala, V. Subramaniyaswamy, "Comparative Study of Classifier for Chronic Kidney Disease.
[3] Rahman, M., Sorna, J. F., Rana, M., Islam, L., Tazim, M., & Alvi, S. T. (2022, January). A Predictive Analysis of Chronic Kidney Disease by Exploring Important Features. International Journal of Computing and Digital Systems, 167–176.
[4] Hosseinzadeh, M., Koohpayehzadeh, J., Bali, A. O., Asghari, P., Souri, A., Mazaherinezhad, A., Bohlouli, M., & Rawassizadeh, R. (2020, July). A diagnostic prediction model for chronic kidney disease in the internet of things platform. Multimedia Tools and Applications, 16933–16950.
[5] Alaiad, A., Najadat, H., Mohsen, B., & Balhaf, K. (2020,March). Classification and Association Rule Mining Technique for Predicting Chronic Kidney Disease. Journal of Information & Knowledge Management, 2040015.
[6] Akhter T., Islam M.A., Islam S. Artificial neural network based covid-19 suspected area identification. J Eng Adv. 2020;1:188–194. [Google Scholar]
[7] Aljaaf A.J., Al-Jumeily D., Haglan H.M., et al. 2018 IEEE Congress on Evolutionary Computation (CEC) IEEE; 2018. Early prediction of chronic kidney disease using machine learning supported by predictive analytics; pp. 1–9. [Google Scholar]
[8] Almasoud M., Ward T.E. Detection of chronic kidney disease using machine learning algorithms with least number of predictors. Int J Soft Comput Appl. 2019;10 [Google Scholar]
[9] Ani R., Sasi G., Sankar U.R., Deepa O. 2016 International Conference on Advances in Computing, Communications and Informatics (ICACCI) IEEE; 2016. Decision support system for diagnosis and prediction of chronic renal failure using random subspace classification; pp. 1287–1292. Arora, M., Sharma, E.A., 2016. Chronic kidney disease detection by analyzing medical datasets in weka. International Journal of Computer of machine learning algorithms. New advances in machine learning 3, 19–48. [Google Scholar]
[10] Arora M., Sharma E.A. Chronic kidney disease detection by analyzing medical datasets in weka. Int J Comput Mach Learn Algor New Adv Mach Learn. 2016;3:19–48. [Google Scholar]
[11] Banik S., Ghosh A. Prevalence of chronic kidney disease in Bangladesh: a systematic review and meta-analysis. Int Urol Nephrol. 2021;53:713–718. [PubMed] [Google Scholar]
[12] Charleonnan A., Fufaung T., Niyomwong T., Chokchueypattanakit W., Suwannawach S., Ninchawee N. 2016 Management and Innovation Technology International Conference (MITicon) IEEE; 2016. Predictive analytics for chronic kidney disease using machine learning techniques. pp. MIT–80. [Google Scholar]
[13] Chen Z., Zhang X., Zhang Z. Clinical risk assessment of patients with chronic kidney disease by using clinical data and multivariate models. Int Urol Nephrol. 2016;48:2069–2075. [PubMed] [Google Scholar]
[14] Chittora P., Chaurasia S., Chakrabarti P., et al. Prediction of chronic kidney disease-a machine learning perspective. IEEE Access. 2021;9:17312–17334. [Google Scholar]
[15] Cueto-Manzano A.M., Cortés-Sanabria L., Martínez-Ramírez H.R., Rojas-Campos E., Gómez-Navarro B., Castillero-Manzano M. Prevalence of chronic kidney disease in an adult population. Arch Med Res. 2014;45:507–513. [PubMed] [Google Scholar]
[16] Eroğlu K., Palabaş T. 2016 National Conference on Electrical, Electronics and Biomedical Engineering (ELECO) IEEE; 2016. The impact on the classification performance of the combined use of different classification methods and different ensemble algorithms in chronic kidney disease detection; pp. 512–516. [Google Scholar]
[17] Fatima M., Pasha M. Survey of machine learning algorithms for disease diagnostic. J Intel Learn Syst Appl. 2017;9(01):1. [Google Scholar]
[18] Gudeti B., Mishra S., Malik S., Fernandez T.F., Tyagi A.K., Kumari S. 2020 4th International Conference on Electronics, Communication and Aerospace Technology (ICECA) IEEE; 2020. A novel approach to predict chronic kidney disease using machine learning algorithms; pp. 1630–1635. [Google Scholar]
[19] Heung M., Chawla L.S. Predicting progression to chronic kidney disease after recovery from acute kidney injury. Curr Opin Nephrol Hypertens. 2012;21:628–634. [PubMed] [Google Scholar]
[20] Tekale S., Shingavi P., & Wandhekar, S. (2018, October). Prediction of Chronic Kidney Disease Using Machine Learning Algorithm. IJARCCE, 92–96.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)