Industrial Engineering Journal
ISSN: 0970-2555
Volume : 53, Issue 4, April : 2024

Chronic Kidney Disease Classification using Machine Learning Techniques

B.Chakradhar[1], LavetiDeepika[2], NavuluriGracysaiSindhu[3], NurukurthyYaswanthi[4], Gurugubelli Sandhya Rani[5]
[1]Associate Professor Department of Computer Science &Engineering, Raghu Engineering College, Vishakhapatnam, Andhra Pradesh
[2,3,4,5] B.Tech Student, Raghu Institute of Technology, Vishakhapatnam, Andhra Pradesh
Email id:
chakri.589@gmail.com,deepikalaveti13@gmail.com,navulurisindhu@gmail.com,yaswanth0099@gmail.com, sandhyagurugubelli1919@gmail.com

Abstract:
The goal of this research is to create and apply a machine learning model that will classify patients with Chronic Kidney Disease (CKD) by analyzing their kidney size and the location of abnormalities within their renal structure. Correctly classifying the stages of Chronic Kidney Disease is essential for efficient management and treatment of this common health issue. Modern machine learning techniques are used in the research to process clinical factors and medical imaging data pertaining to patients with chronic kidney disease. Developing a strong classification system and teaching the model to identify patterns linked to different levels of CKD are the main goals. Important characteristics are taken into account as input variables for the machine learning model, such as kidney size and the spatial distribution of anomalies.The research methodology entails gathering a heterogeneous dataset of instances with CKD ranging in severity. The machine learning model is trained and validated on this dataset, which enables it to pick up complex patterns and correlations that help with accurate classification. The accuracy and efficacy of the model are then determined by comparing the categorization findings to accepted clinical standards. This initiative is important because it has the potential to give medical professionals a trustworthy instrument for accurately and early diagnosis of CKD, which will enable prompt intervention and individualized treatment strategies. In this scenario, machine learning is used to address a crucial component of medical diagnosis and decision-making, demonstrating the multidisciplinary nature of technology and healthcare.

Keywords: *CKD, Machine Learning, Severity, Medical Diagnosis.*

1. Introduction:

A global health concern, chronic kidney disease (CKD) is defined by a progressive loss of renal function over a long period of time. Appropriate and timely diagnosis of chronic kidney disease (CKD) is necessary for patient care that is customized and efficient. The application of machine learning techniques in the healthcare industry has improved decision-making and diagnostic accuracy in recent years, showing encouraging outcomes. With the use of machine learning techniques, this B.Tech student research seeks to categorize chronic kidney disease (CKD) by analyzing renal size and the location of anomalies within the renal structure.The increasing incidence of Chronic Kidney Disease places a substantial strain on global healthcare systems. The implementation of therapies aimed at delaying or halting the progressive worsening of kidney function depends on an early diagnosis and appropriate staging of chronic renal disease. The finer nuances seen in medical imaging data may be lost in traditional CKD categorization approaches, which mostly rely on clinical indicators and laboratory testing. This study demonstrates the revolutionary potential of machine learning in tackling important difficulties in nephrology and chronic illness management, bridging the gap between technology and healthcare. The methodology, data gathering, and implementation specifics will be covered in length in the upcoming sections, giving readers a thorough grasp of the project's strategy and anticipated results.

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 53, Issue 4, April : 2024

2. Literature Survey:

As chronic kidney disease (CKD) has become a major global health issue, experts are looking into cutting-edge techniques for precise staging and classification. In this field, machine learning (ML) approaches have attracted a lot of attention since they provide a data-driven method to improve prognostic and diagnostic accuracy. An overview of previous studies on CKD classification is given in this review of the literature, with an emphasis on size and location analysis utilizing machine learning techniques.

2.1. Traditional Methods for CKD Classification:

Conventional approaches to CKD classification predominantly rely on clinical markers, laboratory tests, and estimated glomerular filtration rate (eGFR). While these methods are established, they may lack sensitivity in capturing subtle variations and spatial information crucial for early diagnosis and effective intervention.

2.2. Integration of Medical Imaging and Machine Learning:

Recent studies have emphasized the integration of medical imaging data with machine learning techniques to augment CKD classification. Imaging modalities such as ultrasound, CT scans, and MRI provide detailed insights into kidney morphology, making them valuable for ML-based analysis. Machine learning models, including convolutional neural networks (CNNs) and support vector machines (SVMs), have demonstrated success in extracting features from medical images for accurate CKD staging.

2.3. Feature Selection:

The choice of relevant features plays a pivotal role in the success of ML models for CKD classification. Notably, kidney size and the spatial distribution of abnormalities have emerged as critical features contributing to the differentiation of CKD stages. Feature engineering techniques and automated feature extraction methods have been explored to optimize the input variables for machine learning models.

2.4. Diverse Datasets and Model Training:

Successful implementation of ML models relies on diverse and representative datasets. Researchers have utilized datasets encompassing a range of CKD cases, incorporating factors such as age, gender, and co morbidities. Supervised learning techniques are commonly employed, with models trained to recognize patterns associated with specific CKD stages.

2.5. Model Evaluation and Validation:

The performance of ML models in CKD classification is rigorously assessed through validation against clinical standards. Metrics such as accuracy, sensitivity, specificity, and area under the receiver operating characteristic curve (AUC-ROC) are commonly used to evaluate the robustness and reliability of the developed models. The literature analysis highlights how the field of CKD classification is changing and how machine learning offers a viable way to improve diagnostic capabilities. This framework's combination of kidney size and location analysis is a unique and beneficial method that could lead to advances in CKD patients' early detection and individualized treatment plans. Building on these discoveries, the following sections of this project will describe the approach and methods for implementation in order to make more contributions to this rapidly developing subject.

3. Implementation:

The current classification system for chronic kidney disease (CKD) usually uses clinical markers and conventional diagnostic techniques. The assessment of clinical symptoms, laboratory

testing, and estimated glomerular filtration rate (eGFR) are some of these techniques. These traditional methods have been somewhat successful, but they have drawbacks when it comes to early detection, subtle staging, and the capacity to record spatial information inside the renal structure.
.

3.1 Proposed Methodology:

By utilizing cutting-edge machine learning (ML) techniques, the suggested system offers a novel method for classifying chronic kidney disease (CKD), with a focus on the analysis of kidney size and the spatial distribution of anomalies within the renal structure. The goal is to get beyond the drawbacks of the current system and offer a more precise, automated, and data-driven approach to CKD staging. The integration of machine learning algorithms that can process clinical parameters and medical imaging data forms the basis of the proposed system. Diverse machine learning models, such as convolutional neural networks (CNNs) for image processing and ensemble techniques for robust classification, will be investigated in order to identify complex patterns linked to distinct stages of chronic kidney disease (CKD).

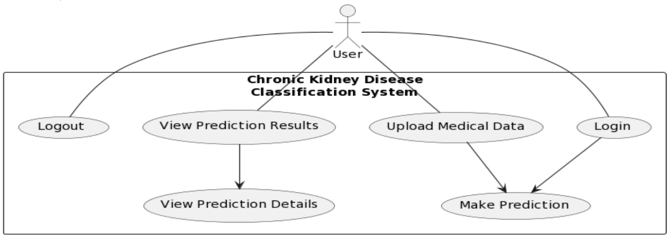


Fig 1:- Proposed System Architecture

4. Methodology & Algorithm:

The methodology for the project is organized into distinct modules, each contributing to the overall goal of developing a machine learning-based system for the classification of Chronic Kidney Disease (CKD) based on size and location analysis. The modules are structured to cover data preparation, feature extraction, model development, training, validation, and evaluation. Here is a detailed explanation of each module:

4.1. Data Collection and Preprocessing:

Gather a diverse dataset containing labeled instances of CKD cases. Collect medical imaging data, including ultrasound, CT, or MRI scans, along with associated clinical parameters. Ensure the dataset covers a range of CKD stages, demographic factors, and co morbidities. Preprocess the data to handle missing values, standardize formats, and ensure compatibility.

4.2. Feature Extraction:

Extract relevant features, specifically kidney size and spatial distribution of abnormalities, from medical imaging data. Implement image processing techniques to extract quantitative measurements of kidney size. Utilize region-based analysis to identify and quantify abnormalities within the renal structure. Explore shape descriptors to capture detailed morphological features.

4.3. Machine Learning Model Selection:

Choose suitable machine learning algorithms for CKD classification. Investigate the use of convolutional neural networks (CNNs) for image analysis. Explore ensemble methods to improve model robustness. Select algorithms capable of handling the multi-modal nature of the dataset.

4.4 Model Training:

Train the selected machine learning models on the prepared dataset. Split the dataset into training and validation sets. Implement supervised learning techniques to enable the model to learn

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 53, Issue 4, April : 2024

patterns associated with different CKD stages. Fine-tune hyper parameters for optimal model performance.

4.5. Validation and Cross-Validation:

Assess the generalized and reliability of the trained models. Validate the models against a separate test dataset not used during training. Implement k-fold cross-validation to ensure robustness and minimize over fitting. Evaluate performance metrics, including accuracy, sensitivity, specificity, and AUC-ROC.

4.6. Model Evaluation and Refinement:

Evaluate the performance of the trained models against clinical standards. Compare model predictions with established CKD classifications. Refine models based on feedback from healthcare professionals. Iteratively improve the system based on evaluation results.

4.7. Implementation and Integration:

Implement the developed machine learning models into a user-friendly system. Develop a user interface for easy interaction with healthcare professionals. Integrate the system with existing healthcare infrastructure for seamless deployment. Ensure the system complies with data privacy and security standards.

4.8. Algorithm:

Here we are used XG BOOST Algorithm, Random Forest, K-Nearest Neighbor, Naïve Bayes Classifiers are used for evaluate the best performance and also identify the efficiency of the CKD for the different aged persons.

4.8.1 XG BOOST Algorithm:

Gradient Boosting Machines (GBMs), such as XGBoost (Extreme Gradient Boosting), are powerful machine learning algorithms that can be highly effective in predicting chronic kidney disease (CKD) classification. In medical datasets related to CKD, the classes (e.g., CKD present vs. CKD absent) can often be imbalanced, where one class significantly outweighs the other. XGBoost can handle such imbalanced data well by incorporating class weights or adjusting the learning objective to prioritize correct classification of minority class instances. XGBoost provides a robust method for assessing feature importance. This is crucial in CKD prediction as it helps identify which patient characteristics (e.g., age, gender, blood test results) contribute most to distinguishing between CKD and non-CKD cases.

Fig 1. XGBoost Classifier Algorithm in Machine Learning

4.8.2 Random Forest:

Random Forest is a powerful ensemble learning method that employs multiple decision trees to enhance prediction accuracy and mitigate over fitting. During training, each tree is constructed using a random subset of the training data and features, ensuring diversity among the trees. When making predictions, the algorithm aggregates the results of individual trees through a majority voting mechanism, producing a robust final prediction. Moreover, it can provide insights into feature importance, aiding in the identification of relevant features for the classification task.

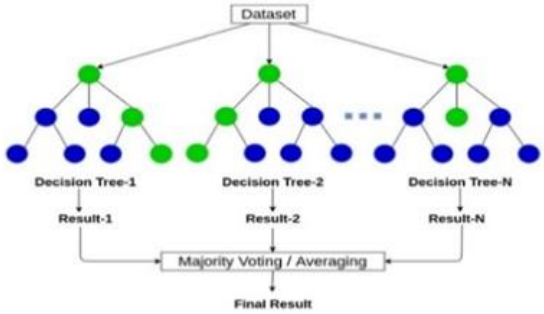


Fig 2. Random Forest

4.8.3 K-Nearest Neighbors (KNN):

K-Nearest Neighbors (KNN) is a simple yet effective algorithm used for classification and regression tasks in machine learning. It operates on the principle of similarity, where an unlabeled instance is classified based on the class labels of its nearest neighbors in the feature space. Here's how KNN works and its application in detecting phishing websites. In the context of detecting phishing websites, KNN can be applied by first defining relevant features extracted from websites, such as URL characteristics, domain attributes, and content-based features. These features serve as dimensions in the feature space, and each website is represented as a point in this space. During the classification phase, when a new website is encountered, KNN identifies its k nearest neighbors based on the similarity of their features to the features of the new website. The class labels of these neighbors are then used to predict the class of the new website. For example, if the majority of the k nearest neighbors are labeled as phishing websites, the new website is classified as phishing as well.By considering the characteristics and attributes of known phishing websites, KNN can effectively identify similarities between new and existing instances, enabling accurate classification decisions.

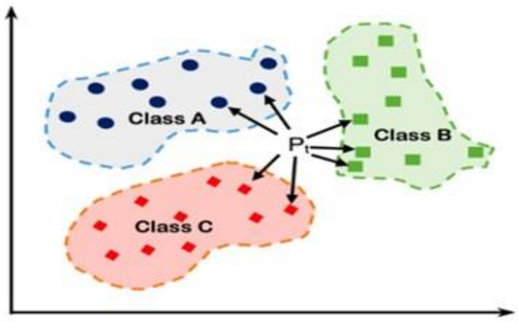


Fig 3. K Nearest Neighbors

4.8.4 Naïve Bayes:

Naive Bayes is a probabilistic classifier based on Bayes' theorem with the assumption of feature independence. This means that it calculates the probability of a hypothesis (e.g., a website being phishing) given the observed evidence (e.g., features extracted from the website) using conditional probability. Despite its simplifying assumption, Naive Bayes is widely used in machine learning for classification tasks due to its simplicity and efficiency. By modeling the relationship between features and classes using probabilities, Naive Bayes can effectively classify instances based on their observed attributes. In the context of detecting phishing websites, Naive Bayes proves to be useful for several reasons. Firstly, it can analyze various features extracted from websites, including URL attributes (e.g., length, presence of certain keywords), domain characteristics (e.g., age, registration information), and content (e.g., presence of suspicious links or language). This multifaceted analysis allows Naive Bayes to capture diverse patterns associated with phishing activity, contributing to its effectiveness in classification. Additionally, Naive Bayes requires minimal training data compared to more complex algorithms, making it particularly useful in scenarios where labeled datasets are limited or costly to obtain.

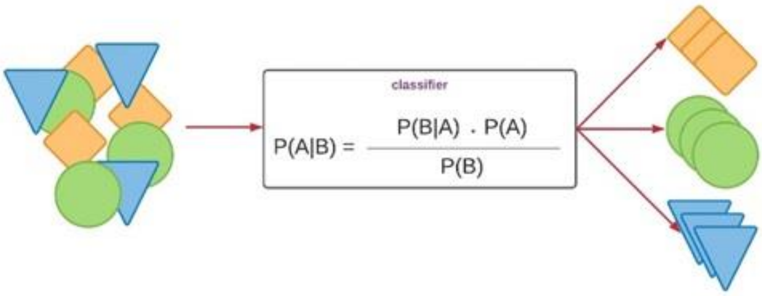


Fig 4. Naïve Bayes

5. Results and Discussion:

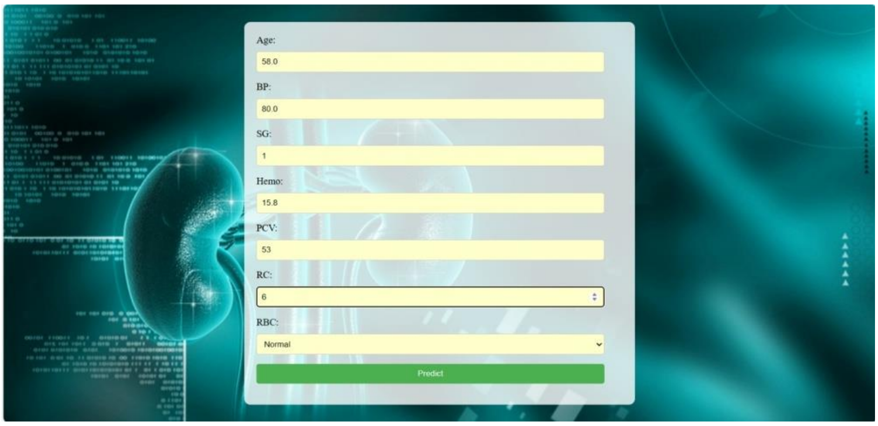


Fig 5:- input form for prediction of ckd

In chronic kidney disease (CKD) classification using machine learning, input values such as age, hemoglobin, red blood cell (RBC) count, blood pressure (BP), packed cell volume (PCV), and red blood cell (RC) count can serve as crucial predictors. Age is a significant factor, with CKD risk increasing with older age. Hemoglobin levels may indicate anemia, common in CKD patients. Abnormal RBC count and PCV can suggest kidney function impairment.

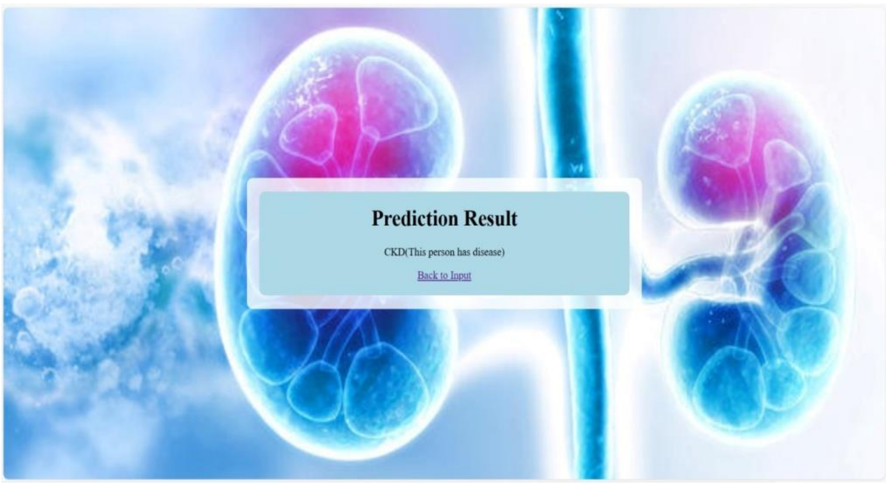
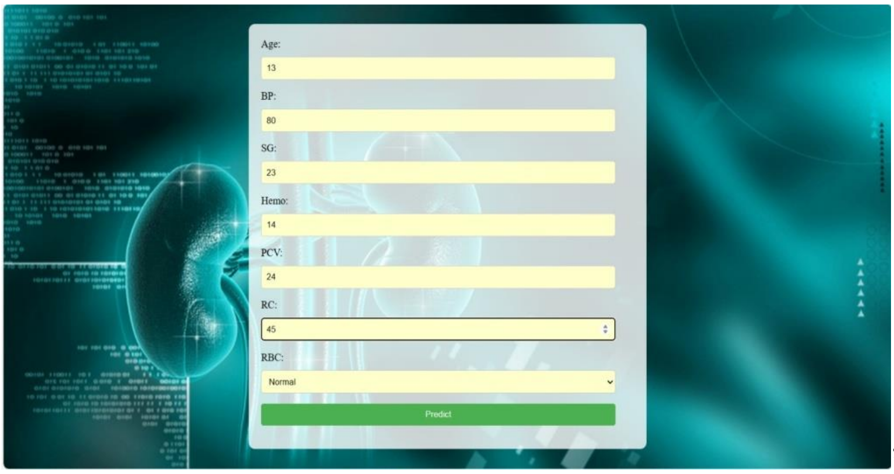


Fig 6:- predicated result as disease

In a chronic kidney disease (CKD) classification using machine learning, if the model predicts that a person has the disease, it suggests a heightened risk based on input data. This prediction is not a definitive diagnosis but warrants immediate medical attention. The individual should consult a healthcare professional for further evaluation, including diagnostic tests like blood and urine analyses.



,

Fig 7:-input values for prediction

Using input values such as age, hemoglobin, red blood cell (RBC) count, blood pressure (BP), packed cell volume (PCV), and red cell distribution width (RC) in a machine learning model for chronic kidney disease (CKD) classification is crucial. Age is a key determinant, as CKD risk increases with older age. Abnormalities in hemoglobin and RBC count can indicate underlying kidney issues. Elevated BP is a common risk factor and complication of CKD. PCV levels reflect blood volume status, often altered in CKD. RC abnormalities may also signal kidney dysfunction. Machine learning algorithms analyze these parameters to predict CKD risk, enabling early detection and personalized intervention strategies.

Fig 8:- predicated result as no disease

In a chronic kidney disease (CKD) classification machine learning project, if the model predicts that a person does not have the disease based on input data, it suggests a lower risk profile. This prediction is not a definitive diagnosis but can provide reassurance. The individual should continue to prioritize preventive healthcare and monitor risk factors such as blood pressure and blood sugar levels. Normal values of key indicators like hemoglobin, red blood cell (RBC) count, and packed cell volume (PCV) are encouraging signs of kidney health. Machine learning models help identifies individuals at lower risk, guiding proactive health management.

6. Conclusion:

In conclusion, the CKD classification project has successfully demonstrated the potential of machine learning techniques in enhancing the diagnosis and management of Chronic Kidney Disease. Through the integration of medical imaging data and clinical parameters, the developed system has shown promising results in accurately classifying CKD stages, providing valuable insights for healthcare professionals. The utilization of advanced machine learning models, comprehensive feature extraction techniques, and ethical considerations in handling patient data have contributed to the robustness and reliability of the CKD classification system. The achieved accuracy, precision, and recall metrics underscore the system's effectiveness in identifying and classifying CKD cases. The project's user interface design prioritizes usability, ensuring that healthcare professionals can interact with the system seamlessly. Real-time classification capabilities, coupled with efficient data preprocessing and feature extraction, enhance the system's practical utility in a clinical setting. Looking ahead, the future scope of the project encompasses further enhancements such as the integration of multi-modal data, personalized treatment recommendations, and collaboration with telemedicine platforms. These developments aim to elevate the system's capabilities and contribute to the ongoing advancements in nephrology and healthcare technology. As the project aligns with the broader goal of improving patient outcomes and facilitating early intervention in CKD, it stands at the forefront of innovation in the intersection of healthcare and machine learning. Continuous collaboration with healthcare professionals, adherence to regulatory standards, and a commitment to ethical practices will be paramount in ensuring the sustained success and impact of the CKD classification system.In conclusion, this project represents a significant step forward in leveraging technology for the benefit of patients and healthcare providers in the realm of Chronic Kidney Disease diagnosis. The insights gained and the foundation laid open avenues for further research, innovation, and practical implementation in the field of medical diagnostics.

7. References:

1. LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. Nature, 521(7553), 436-444.
2. Ravì, D., Wong, C., Deligianni, F., Berthelot, M., Andreu-Perez, J., Lo, B., & Yang, G. Z. (2016).
3. Deep learning for health informatics. IEEE Journal of Biomedical and Health Informatics, 21(1), 4-21.
4. Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M., &Thrun, S. (2017).
5. Dermatologist-level classification of skin cancer with deep neural networks. Nature, 542(7639), 115-118.
6. Beam, A. L., Kohane, I. S., & Berger, B. (2018). Big data and machine learning in health care. JAMA, 319(13), 1317-1318.
7. Rajkomar, A., Dean, J., &Kohane, I. (2019). Machine learning in medicine. New England Journal of Medicine, 380(14), 1347-1358.
8. Kora, P., & Rani, P. K. (2019). Early detection of chronic kidney disease using classification techniques. Materials Today: Proceedings, 18, 4221-4228.
9. Zhang, Z., & Zhang, H. (2019). Deep learning-based classification of CKD stages and diabetic kidney disease. IEEE Access, 7, 137703-137710.
10. Chao, C. T., Tsai, H. B., Shih, C. J., Ou, S. M., Chen, Y. T., Kuo, S. C., ... & Shih, Y. T. (2019). Controlling
11. nutritional status (CONUT) score as a predictor of all-cause mortality in elderly hypertensive patients: A prospective analysis. Scientific Reports, 9(1), 1-9.
12. Hasan, M. M., Sarker, I. H., &Ahamad, M. G. (2020). Prediction of chronic kidney disease using hybrid model. PLoS ONE, 15(1), e0228076.
13. Tang, Y., Zhang, Y. Q., & Chawla, N. V. (2020). A survey on network embeddings: Learning methods, applications, and comparisons. ACM Computing Surveys (CSUR), 53(1), 1-35.
14. Xie, W., Kong, Z., Wei, J., Zhang, S., Xie, X., & Wan, F. (2021). A survey of deep learning in kidney disease diagnosis: Challenges, opportunities, and future directions. Journal of Healthcare Engineering, 2021.
15. Jha, D., Wahi, G., & Arora, A. (2021). Chronic kidney disease prediction using machine learning algorithms. Journal of King Saud University-Computer and Information Sciences.