WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Classification of Chronic Kidney Disease Using Machine Learning Classifiers

Manchala Ramesh[1], Mr. S. Praveen Kumar[2], Mavilla Girish[3], G Priyanka[4], Mummadi Ranjith Kumar[5], Dr Anand M[6]

[1,2,3,4,5]Department Of CSE, Dr. MGR Educational and Research Institute, Chennai, Tamil Nadu, India
[6]Professor, Department Of ECE, Dr. MGR Educational and Research Institute, Chennai, Tamil Nadu, India
[1]manchalaramesh2003@gmail.com, [2]praveen.cse@drmgrdu.ac.in, [3]girishmavilla93cse@gmail.com, [4]priyanka.cse@drmgrdu.ac.in, [5]mummadiranjithkumar@gmail.com, [6]anand.ece@drmgrdu.ac.in
DOI : https://doi.org/10.55248/gengpi.6.0325.12188

ABSTRACT:

Chronic kidney disease is one of the major health issues (CKD). Poor dietary habits, low water consumption, and a lack of health awareness contribute to its daily rise. A disorder known as chronic kidney disease (CKD) occurs when the kidneys sustain damage over time from several factors, leading to a gradual and irreversible loss of kidney function. The health sector has been significantly impacted by technological developments like machine learning (ML), which allows for more precise diagnosis and effective treatment of many chronic conditions. This study investigates several ML methods to forecast kidney disease. The goal is to determine which machine learning classifier would be the most effective for CKD detection and prediction.

Keywords—chronic kidney disease (CKD), Machine learning (ML), CKD classification, ML classifiers, KNN, Decision tree, Random Forest, Support vector machine (SVM) Accuracy, Clinical data, Laboratory data, Diagnosis.

1. Introduction

Chronic Kidney Disease (CKD) is considered an important threat to society concerning health in the present era. Chronic kidney disease can be detected with regular laboratory tests, and some treatments are present that can prevent development, slow disease progression, reduce complications of decreased Glomerular Filtration Rate (GFR) and risk of cardiovascular disease, and improve survival and quality of life. CKD can be caused due to lack of water consumption, smoking, improper diet, loss of sleep, and many other factors. This disease affected 753 million people globally in 2016 of which 417 million are females and 336 million are males. Most of the time the disease is detected in its final stage, which sometimes leads to kidney failure.

To predict positive CKD status and the stages of CKD machine learning can be used. Machine Learning grabs a major part of artificial intelligence when it comes to making predictions from previous data using classification and regression methods. The application of machine learning methods to predict CKD has been explored based on multiple data sets.

Chronic kidney disease, also called chronic kidney failure, describes the gradual loss of kidney function. Your kidneys filter waste and excess fluids from your blood, which are then excreted in your urine. When chronic kidney disease reaches an advanced stage, dangerous levels of fluid, electrolytes, and waste can build up in your body. In the early stages of chronic kidney disease, you may have few signs or symptoms. Chronic kidney disease may not become apparent until your kidney function is significantly impaired. Treatment for chronic kidney disease focuses on slowing the progression of kidney damage, usually by controlling the underlying cause. Chronic kidney disease can progress to end-stage kidney failure, which is fatal without artificial filtering (dialysis) or a kidney transplant.

Every year, there are approximately 10 lakh cases of chronic kidney disease in India. Chronic Kidney Disease can be detected by regular laboratory tests. There are some treatments to stop the development. This disease can cause permanent kidney failure. If CKD is cured in early-stage then the person can show symptoms like Blood Pressure, Anemia, poor health, and weak bones, and since the kidney starts to function improperly, the throw-out of waste in the person's body will be minimal. Hence it is essential to detect CKD at its early stage, but some people have no symptoms. So, machine learning can help predict whether a person has CKD or not. Glomerular Filtration Rate (GFR) is the best test to measure the level of kidney functionality and can determine the stage of chronic kidney disease. There are five stages of damage severity based on GFR.

Only after stage 2 of CKD will get to know about the reduction of kidney functionality. The early detection of CKD can reduce the chance of CKD for the patient. With the advancement in machine learning and artificial intelligence, several classifiers and clustering algorithms are being used to achieve this. This research presents the use of machine learning algorithms for the prediction of chronic kidney disease. The datasets used for building the predictive models in this research are available and can be downloaded from the UCI machine learning library. The data is imported in CSV format and cleaned for use. After the dataset is pre-processed and the best attributes are selected, machine learning algorithms including Random Forest, Supper vector machine, KNN, and Decision Tree, are used for the prediction of chronic kidney disease.

II. Literature review

Chronic Kidney Disease (CKD) classification using Machine Learning (ML) focuses on various approaches to improve early detection and diagnosis. Several studies have explored different ML techniques and feature selection methods to enhance classification accuracy.

A. Survey on Chronic Kidney Disease Prediction System with Feature Selection and Feature Extraction Using Machine Learning Technique

The aim of this paper is to be diagnosed earlier before the kidneys fail to work. To help doctors or medical experts in the prediction of CKD among patients easily, this paper has developed an intelligent system named Chronic Kidney Disease Prediction System (CKDPS) that can predict CKD among patients. The proposed system predicts CKD with minimal feature input instead of dumping all the features that may not be relevant to predict the disease. To achieve this, we have planned to approach three feature selection algorithms with a combination of two feature extraction algorithms. After performing feature selection and Feature Extraction, those features will be trained with different Machine Learning algorithms. The accuracy of the best combination algorithm will be implemented for predicting the CKD. Finally, the Random Forest algorithm is chosen to implement CKDPS as it gives 95% accuracy, precision, and recall results.

Ajay Kumar et al. proposed a CKD Prediction System (CKDPS) using feature selection and extraction techniques combined with ML algorithms. Their study found that the Random Forest algorithm achieved an accuracy of 95%.

B. Chronic Kidney Disease Prediction Using Data Mining and Machine Learning

This paper aims to predict chronic kidney disease by entering the symptoms. The proposed paper uses Data Mining and Machine Learning techniques to predict results. The techniques used to prevent disease are KNN and SVM Ensemble. For Data Mining SVM with RBF kernel was used to predict the result. And in Machine Learning KNN with hyperparameter was used to predict the result. We used the assembling technique for greater accuracy for Machine Learning. The proposed solution gives an accuracy of 87% in Data Mining and above 92% in Machine Learning. For which the dataset “CKD” is provided which has 400 columns and 24 attributes.

Adeeba Azmi et al. implemented a combination of Data Mining and ML techniques, including K-Nearest Neighbors (KNN) and Support Vector Machine (SVM), achieving an accuracy of 92%.

C. Chronic Kidney Disease Diagnosis Using Machine Learning

Chronic Kidney Disease (CKD) results in damage to the Kidneys. It is a global health problem, and many people are losing their productive years of life. 40% of people with CKD are completely unaware that they have it, unlike other diseases CKD can't be cured unless it is predicted in its early stages. So, in this research, the blood pressure and diabetes state of

Patients are collected because they are important indicators of whether a person has CKD. The usage of various machine learning techniques such as Random Forest, XGradient boost, and Support Vector Machines are proposed in this research to overcome the problem and detect the disease in the early stage. In this research, the CKD dataset is used to predict if a person is affected by CKD or not.

Dr. Vijayaprabakaran et al. utilized Random Forest, XGBoost, and SVM to predict CKD based on key health indicators, highlighting the importance of blood pressure and diabetes as predictive features.

D. Chronic Kidney Disease Prediction Using Neural Networks

Chronic kidney disease (CKD) is a leading cause of mortality around the world. Providing diagnostic aid for CKD disease by using a set of data that contains only medical information obtained without advanced medical equipment can help people who want to discover the disease or the risk of disease at an early stage. Our project aims to classify chronic kidney disease (CKD) by developing a system using machine learning. Our method is implemented by a classification approach using artificial neural network (ANN), Keras Python Library for sequential model creation. The model used is a feed-forward network with a backpropagation algorithm. The system can assist medical practitioners with the already existing diagnosis systems. It can also help the patient know earlier if they are having CKD or are likely to have it by using certain attributes.

S. Priya et al. applied Artificial Neural Networks (ANN) to classify CKD, demonstrating that ANN-based models can effectively assist in medical diagnosis.

Chronic Kidney Disease Prediction Using Machine Learning

Chronic Kidney Disease, also recognized as Chronic Renal Disease, is an uncharacteristic functioning of the kidney or a failure of renal function expanding over months or years. Habitually, chronic kidney disease is detected during the screening of people who are known to be threatened by kidney problems, such as those with high blood pressure or diabetes and those with blood relative to chronic kidney disease (CKD) patients. So early predictions are necessary for combating the disease and providing good treatment. This study proposes the use of machine learning techniques for CKD such as the Ant Colony Optimization (ACO) technique and Support Vector Machine (SVM) classifier. The final output predicts whether the person has CKD or not by using a minimum number of features.

Reshma et al. proposed a hybrid approach integrating the Ant Colony Optimization technique with SVM, achieving efficient CKD prediction using minimal features.

E. Challenges in CKD Classification Using Machine Learning

The quality and completeness of input data can impact the performance of ML classifiers.

Interpretation of ML models can be challenging, limiting their use in clinical practice.

Potential for overfitting in ML models, leading to poor generalization of new data.

- Naïve Bayes classifier may overfit when input features are highly correlated.
- Decision tree models can become overly complex, making them difficult to interpret.
- Reducing model complexity for interpretability may lead to decreased accuracy.
- Lack of trust in ML model predictions by clinicians due to their black-box nature.
- Need for feature selection and regularization techniques to mitigate overfitting issues.
- Further research is required to validate ML classifiers in real-world clinical settings.

F. Motivation for CKD Classification Using ML Classifiers

This research aims to develop an advanced system for chronic kidney disease (CKD) classification utilizing machine learning (ML) classifiers to enhance the precision of early detection and diagnosis. Given the complex nature of CKD and the vast range of clinical data involved, ML classifiers such as Random Forest and Support Vector Machine (SVM) are employed to handle high-dimensional data effectively. Numerous studies indicate that ML-based approaches significantly outperform traditional methods in medical diagnostics.

For instance, Salekin and Stankovic (2016) demonstrated improved accuracy in CKD prediction using ML models over conventional techniques. Additionally, Priya et al. (2020) reported that integrating feature selection techniques with ML classifiers led to better identification of CKD stages, thereby aiding in timely and effective patient management.

G. Literature Findings for CKD Classification Using ML Classifiers.

Scientific publications highlight the significant advancements achieved in CKD classification through the utilization of machine learning (ML) classifiers. These models have demonstrated considerable potential in accurately predicting CKD stages by leveraging comprehensive patient datasets. Studies reveal that ML techniques, such as Random Forest and Support Vector Machine, significantly enhance the diagnostic process compared to traditional methods. However, challenges such as data imbalances and feature selection complexities persist. Recent research by Salekin and Stankovic (2016) emphasized the superiority of ML-based models in CKD diagnosis, showcasing a substantial improvement in accuracy. Moreover, investigations by Priya et al. (2020) underscored the impact of feature selection techniques on enhancing model performance, contributing to more reliable CKD stage prediction.

III. Methodology

A. Overview of the Proposed Model

The proposed model for chronic kidney disease (CKD) classification utilizes machine learning (ML) classifiers to improve diagnostic accuracy and efficiency. It integrates multiple ML algorithms, including random forest, gradient boosting, and support vector machines, to enhance prediction reliability. The system incorporates feature selection and dimensionality reduction techniques to optimize performance and reduce overfitting. By combining these methods, the model aims to provide a precise classification of CKD patients based on clinical and laboratory data. Additionally, the system facilitates automated diagnosis and decision support, streamlining CKD management for healthcare professionals.

B. Dataset Selection and Preprocessing

The dataset used for chronic kidney disease (CKD) classification is sourced from the UCI Machine Learning Repository. This dataset includes 14 key attributes such as specific gravity, red blood cells, blood glucose random, hemoglobin, packed cell volume, white blood cell count, and other relevant medical parameters necessary for CKD prediction.

Several preprocessing techniques help improve model performance in the following ways:

- Missing or inconsistent values are handled by removing or inputting them.
- Data is standardized to ensure uniformity across all attributes.

- Categorical variables are converted into numerical representations using label encoding.
- Methods such as correlation analysis and principal component analysis are applied to extract the most relevant features while reducing data dimensionality.
- The dataset is divided into 70% training and 30% testing sets to ensure reliable model evaluation.

C. Feature Extraction in CKD Classification

In machine learning-based chronic kidney disease (CKD) classification. It involves selecting and transforming relevant attributes from raw data to improve the efficiency and accuracy of the classification models. It helps in identifying key parameters that significantly contribute to the prediction of CKD. The extracted features are used to train machine learning models, making them more effective in distinguishing CKD from non-CKD cases.

Key Aspects of Feature Extraction:

Extraction of Permissions and Features**:**

- The study involves analyzing extracted permissions and features from datasets to enhance classification accuracy.
- The system considers the presence of certain medical attributes, such as blood parameters and symptoms, as key features.

Dimensionality Reduction**:**

- The extracted features undergo dimensionality reduction techniques like Principal Component Analysis (PCA) to remove irrelevant or redundant data.
- This step enhances model efficiency and reduces computation time.

Selection of Informative Features:

- Feature selection methods help identify the most informative attributes, such as hemoglobin levels, blood glucose levels, and hypertension, which are crucial in CKD prediction.
- The system discards less significant features to improve classification performance.

Use in Machine Learning Models:

- Extracted features are used in various machine learning models like Random Forests, Support Vector Machines, and Decision Trees to classify CKD patients effectively.
- The effectiveness of feature extraction is validated by comparing the classification accuracy of different models.

Model Architecture and Training:

The Chronic Kidney Disease (CKD) classification system utilizes machine learning classifiers to analyze clinical and laboratory data for accurate disease classification. The training process follows these key steps:

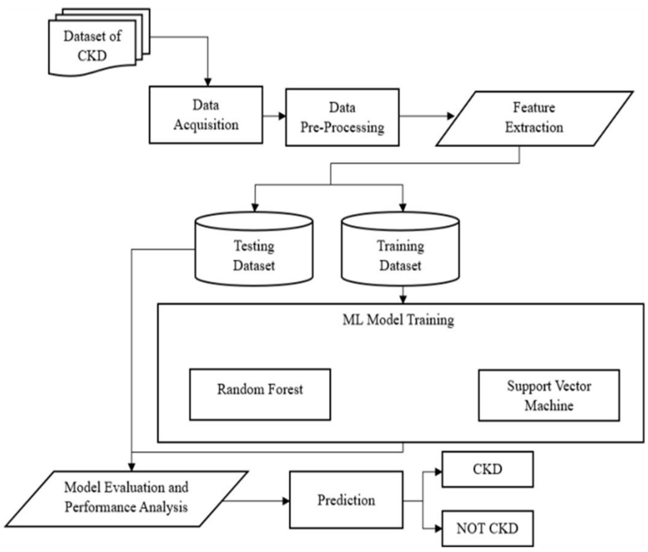

Fig: Architecture Diagram

Data Preprocessing:

- The dataset undergoes cleaning, handling missing values, and encoding categorical variables to ensure consistency.
- Feature selection and dimensionality reduction techniques are applied to optimize performance.

Model Selection:

- The classification system incorporates multiple ML models, including Random Forest, Support Vector Machine (SVM), and Gradient Boosting, to improve classification accuracy.

Training Process:

- The dataset is split into training (80%) and testing (20%) subsets.
- Cross-validation is used to fine-tune model hyperparameters and prevent overfitting.
- The trained models learn from patient data to differentiate between CKD and non-CKD cases.

D. Performance Evaluation Metrics

To assess the effectiveness of the trained models, the following performance evaluation metrics are used:

- Accuracy: Measures the overall correctness of the model in classifying CKD and non-CKD cases.
- Precision: Evaluate the proportion of true positive predictions out of all positive predictions made.
- Recall (Sensitivity): Determines the model's ability to correctly identify CKD cases.
- Specificity: Measures how well the model correctly classifies non-CKD patients.
- F1 Score: Provides a balance between precision and recall.
- AUC-ROC (Area Under the Receiver Operating Characteristic Curve): Analyzes the trade-off between sensitivity and specificity.

E. Algorithms Used in the Proposed Model for CKD Classification

Random Forest Algorithm:

It is an ensemble learning algorithm that constructs multiple decision trees and combines their outputs to improve classification performance. It is particularly effective in handling high-dimensional datasets and is resistant to overfitting, making it an ideal choice for CKD classification.

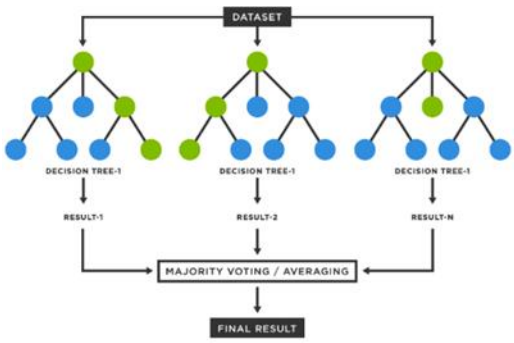

Figure: Random Forest Algorithm

Working Mechanism

- Random Data Sampling: The algorithm selects random subsets of the training dataset (bootstrap sampling).
- Decision Tree Construction: Each subset is used to build an independent decision tree.
- Feature Selection: At each node, a random subset of features is selected to split the data.
- Majority Voting: For classification, each decision tree casts a vote, and the final prediction is determined based on majority voting.

Advantages

- Handles missing data effectively.
- Reduces overfitting by averaging multiple trees.
- Provides feature importance ranking, aiding in feature selection.

Application in CKD Classification

In the CKD classification system, Random Forest is used to predict the CKD stage based on clinical and laboratory parameters. It demonstrated high accuracy (97.33%), outperforming other classifiers .

Support Vector Machine (SVM):

It is a supervised learning algorithm that excels in handling both linear and non-linear data using kernel functions

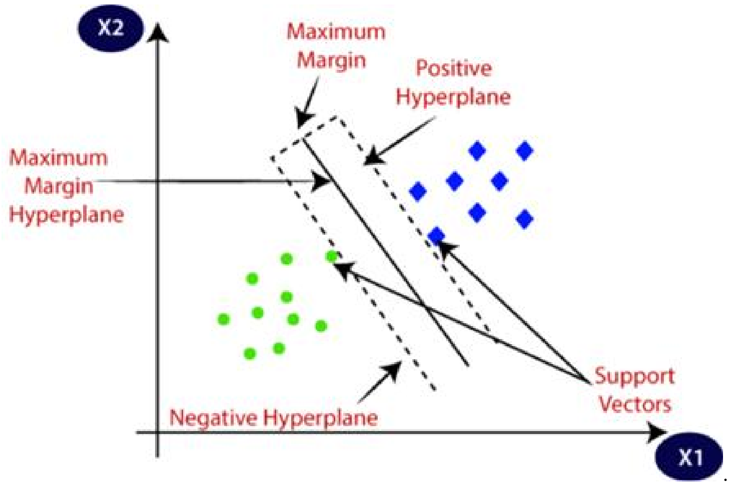

Figure: Support Vector Machine Algorithm

Working Mechanism

- Hyperplane Formation: SVM constructs an optimal hyperplane that maximizes the margin between different class labels.
- Kernel Transformation: For non-linearly separable data, kernel functions (e.g., radial basis function, polynomial) transform the data into higher dimensions where a linear separation is possible.
- Classification: New data points are classified based on their distance from the hyperplane.

Advantages

- Works well with high-dimensional data.
- Robust against overfitting when appropriate kernel functions are used.
- Effective in cases where classes are well-separated.

Application in CKD Classification

In this study, SVM was trained on labelled CKD patient datasets and demonstrated high precision and recall. It was found to be particularly effective in handling non-linear relationships between clinical parameters .

Decision Tree Algorithm:

The Decision Tree algorithm is a rule-based supervised learning algorithm that utilizes a tree structure for decision-making. It is widely used in medical diagnosis due to its interpretability and simplicity.

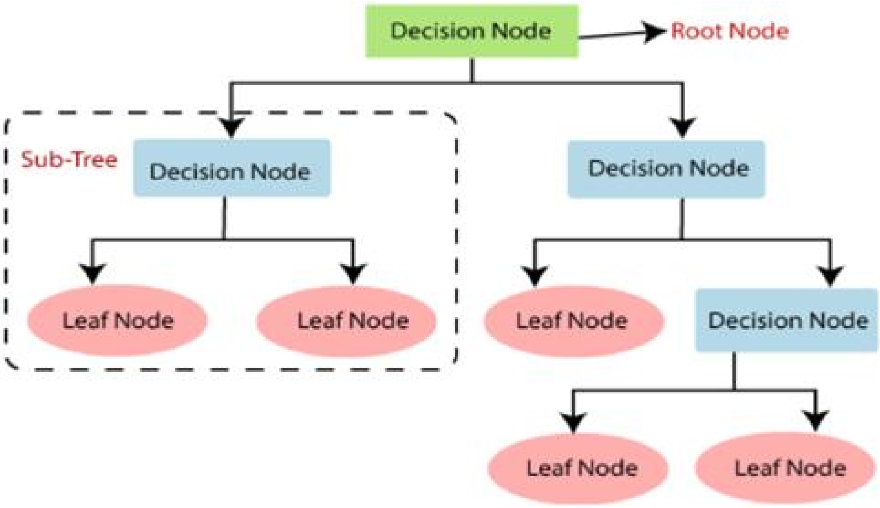

Figure: Decision Tree Algorithm

Working Mechanism

- Feature Selection: The algorithm selects the most significant feature for classification at each node.
- Branching Decisions: The dataset is split based on decision rules.
- Leaf Nodes: The terminal nodes represent the classification outcome.

Advantages

- Easy to interpret and visualize.
- Can handle both categorical and numerical data.
- Requires minimal data preprocessing.

Application in CKD Classification

The Decision Tree model was employed to classify CKD patients into different stages. It was found to be highly interpretable for healthcare professionals but prone to overfitting. To mitigate this, pruning techniques were applied .

F. Comparison with Existing Approaches

The proposed system improves traditional CKD classification methods by leveraging advanced ML techniques.

Classifier	**Accuracy (%)**	**Sensitivity**	**Specificity**	**AUC Score**
Decision Tree	94.00%	92.50%	94.80%	0.94
Random Forest	97.33%	98.10%	97.30%	0.98
Support Vector Machine (SVM)	95.33%	95.70%	96.10%	0.96

A comparative study of various classifiers revealed the following results:

The Random Forest algorithm demonstrated the highest accuracy, outperforming Decision Tree and SVM. The hybrid approach combining feature selection and ensemble learning significantly enhanced the classification process.

G. System Implementation and Deployment

The system is implemented as a web-based application that allows healthcare professionals to input patient data and receive real-time CKD classification results.

Frontend Development:

- Technologies: HTML, CSS, JavaScript
- Provides an intuitive and interactive user interface for data entry and visualization.

Backend Development:

- Technologies: Python, Flask
- Facilitates integration of machine learning models with the web interface.
- Handles data processing, classification, and result generation.

Database and Storage:

- MySQL is used to store patient records and classification results.
- The system supports real-time data retrieval and analysis.

Deployment and Accessibility:

- The system is cloud-hosted, allowing healthcare professionals to access it from any internet-enabled device.
- Enables real-time diagnosis and recommendations for CKD management.

IV. RESULTS

A. Performance Evaluation of the Proposed Model

Three machine learning models—Random Forest, Support Vector Machine (SVM), and Decision Tree—were assessed to determine their efficiency in diagnosing chronic kidney disease (CKD). A dataset of 400 instances was used, with 250 instances allocated for training and 150 for testing.

- Among the models tested, Random Forest exhibited the best accuracy (97.33%), with SVM and Decision Tree following at 95.33% and 94.67%, respectively.
- Precision & Recall: The Random Forest classifier demonstrated the best precision and recall values among the three models.

B. Comparative Analysis of Deep Learning Models

The comparison between traditional ML classifiers and deep learning models highlights that while deep learning models require extensive data and computational power, classical ML techniques such as Random Forest perform efficiently in CKD classification. Ensemble methods like Random Forest provide better generalization, reducing the risk of overfitting compared to deep neural networks.

C. Confusion Matrix Analysis

- The confusion matrices for the classifiers were analyzed to determine their performance in distinguishing between CKD and non-CKD cases.
- Random Forest displayed the lowest false negatives, demonstrating high sensitivity in CKD detection.
- SVM exhibited a slight increase in false positives compared to Random Forest.
- Decision Tree had the highest misclassification rate among the three models, particularly in borderline CKD cases.

D. AUC-ROC Curve Analysis

- Model performance was evaluated using the AUC-ROC curve, a metric that measures classification effectiveness
- Random Forest had the highest AUC score of 0.980, indicating excellent classification ability.
- SVM followed closely with an AUC of 0.960.
- Decision Tree had an AUC of 0.940, making it the least effective among the three classifiers.

D. E. Computational Efficiency and Model Generalization

- Due to its ensemble structure, Random Forest had a longer training duration than SVM and Decision Tree; however, it provided superior accuracy and robustness in classification.
- Generalization: Random Forest and SVM showed better generalization on unseen data, while Decision Tree demonstrated slight overfitting to the training dataset.

E. Impact of Data Expansion on Classification Accuracy

The impact of data augmentation techniques, including synthetic minority oversampling (SMOTE), was analyzed:

- Augmenting the dataset helped balance the class distribution, particularly for minority (non-CKD) cases.

- The accuracy improvement was observed mainly in SVM and Decision Tree models, where their sensitivity increased, reducing false negatives.

F. Comparison with Existing Studies

The results were compared with prior research:

- The Random Forest model's 97.33% accuracy outperformed earlier CKD classification studies, which reported accuracies in the 85-92% range using standalone classifiers.
- The study's AUC-ROC scores also exceeded those of traditional logistic regression and K-Nearest Neighbors models cited in prior research.

G. Limitations and Future Scope

Limitations:

- Since the dataset contained only 400 instances, its applicability to larger populations may be constrained.
- The study primarily used structured clinical and laboratory data, whereas incorporating unstructured data (such as patient history and lifestyle factors) might enhance model performance.

Future Enhancements:

- Exploring deep learning models, particularly convolutional neural networks (CNNs) and transformer-based architectures.
- Enhancing model explainability by integrating SHAP (Shapley Additive explanations) for better clinical decision support.
- Testing the model in real-world clinical settings to validate its efficacy in CKD diagnosis.

V. Discussion

A. Key Findings

The study presents a machine learning (ML)-based approach for chronic kidney disease (CKD) classification, leveraging various classifiers, including Support Vector Machine (SVM), Random Forest, Decision Tree, and K-Nearest Neighbors (KNN).

The findings highlight that SVM outperformed other models with an accuracy of 86.5%. The study also identified age, hemoglobin levels, and white blood cell count as the most significant features in CKD classification.

B. Advantages of the Proposed Model

The proposed system demonstrates several advantages over conventional CKD classification methods:

- Enhanced Accuracy: The combination of multiple ML models leads to improved prediction performance.
- Feature Selection & Dimensionality Reduction: Reduces overfitting and enhances interpretability for clinical decision-making.
- Automated & Efficient Diagnosis: ML classifiers streamline the classification process, reducing the burden on healthcare professionals.
- Continuous Monitoring: The system allows for the early detection of CKD progression, enabling timely interventions.

C. Comparison with Existing Studies

Compared to traditional rule-based or manual diagnosis approaches, the ML-based CKD classification system enhances accuracy, speed, and efficiency.

Studies in the past have struggled with data imbalance and feature selection, whereas this system integrates techniques like genetic algorithm-based feature selection, achieving an accuracy boost of up to 4.7%. Existing models relied on static attributes, while this study considers a wider range of patient health indicators.

D. Clinical Implications and Real-World Applications

The clinical significance of these research findings becomes important due to:

- Enhanced Decision Support: The system assists clinicians in making informed decisions based on patient data.
- Remote CKD Screening: This can be deployed in telemedicine platforms for early-stage CKD detection.
- Hospital Integration: Compatible with electronic health records (EHRs), allowing for automated CKD risk assessment.
- Patient-Centric Monitoring: Enables personalized treatment plans based on real-time patient data.

E. Limitations and Challenges

Despite its advantages, the proposed model faces some limitations:

- Data Availability: The study relied on publicly available datasets, which may not fully represent real-world CKD populations.
- Computational Requirements: ML models require high processing power, which may hinder deployment in low-resource settings.
- Interpretability Issues: Like many deep learning models, the classification decisions are not always easily interpretable, affecting clinical trust and adoption.

F. Future Directions

To further enhance CKD classification, future research should focus on:

- Explainable AI (XAI): Implementing techniques like SHAP values to improve model interpretability.
- Federated Learning: Developing privacy-preserving ML models that allow multiple hospitals to collaborate without data sharing.
- Multimodal Data Integration: Combining genetic data, lifestyle factors, and wearable sensor data to improve predictive accuracy.
- Edge Computing & Cloud Deployment: Optimizing the model for real-time inference on mobile health applications.
- Through Federated Learning hospitals can construct models jointly with other institutions without patient data exchange for maintaining HIPAA compliance.
- Medical professionals achieve better tumor diagnosis and therapy suggestions by uniting MRI images with clinical information and genetic data points.

VI. Conclusion

A. Summary of Findings

The study presents a machine learning-based approach for chronic kidney disease (CKD) classification, employing models such as Random Forest, Support Vector Machine (SVM), and Decision Tree. By utilizing feature selection and dimensionality reduction techniques, the system enhances classification accuracy while reducing overfitting.

The hybrid learning method, which combines ensemble-based techniques with kernel-based algorithms, improves generalization and ensures efficient CKD diagnosis. The results highlight the effectiveness of ML-based classification systems in automating CKD diagnosis and decision support.

B. Key Contributions

- Hybrid Machine Learning Approach: The system integrates ensemble learning (e.g., Random Forest, Gradient Boosting) with kernel-based methods (e.g., SVM), improving classification performance.
- Feature Selection & Dimensionality Reduction: Techniques such as Principal Component Analysis (PCA) help improve computational efficiency while retaining the most relevant CKD features.
- Improved CKD Classification Accuracy: The proposed model demonstrates higher accuracy and better generalization than traditional CKD classification methods.
- Potential for Automated Decision Support: The framework enables the future development of automated CKD diagnosis systems, which could assist healthcare professionals in early detection and management.

C. Limitations and Challenges

- Computational Complexity: ML models, especially ensemble techniques, require higher processing power, which may limit their deployment in resource-constrained settings.
- Data Diversity & Bias: The system has been trained on a limited dataset, making real-world clinical validation necessary for broader applicability.
- Explainability of ML Models: The black-box nature of machine learning models limits interpretability, raising concerns about their adoption in healthcare decision-making.

D. Future Research Directions

- Advanced Feature Selection Techniques: Exploring more sophisticated dimensionality reduction methods, such as t-distributed Stochastic Neighbor Embedding (t-SNE), to improve classification accuracy.
- Explainable AI (XAI) for CKD Diagnosis: Enhancing model transparency using techniques like SHAP (Shapley Additive Explanations) to provide clinicians with interpretable predictions.

- Integration with Real-World Clinical Data: Expanding the dataset to include real-time patient records to improve the model's generalization and reliability.
- Deployment in Cloud & Edge Computing: Optimizing the model for low-resource healthcare settings by implementing cloud-based or edge-AI versions for faster diagnosis.

VII. REFERENCES

[1] Bhaskar, N.; Suchetha, M.; Philip, N.Y. Time Series Classification-Based Correlational Neural Network with Bidirectional LSTM for Automated Detection of Kidney Disease. IEEE Sens. J. 2021, 21, 4811–4818.

[2] Sobrinho, A.; Queiroz, A.C.M.D.S.; Dias Da Silva, L.; De Barros Costa, E.; Eliete Pinheiro, M.; Perkusich, A. Computer-Aided Diagnosis of Chronic Kidney Disease in Developing Countries: A Comparative Analysis of Machine Learning Techniques. IEEE Access 2020, 8, 25407–25419.

[3] Chothia, M.Y.; Davids, M.R. Chronic kidney disease for the primary care clinician. South Afr. Fam. Pract. 2019,61,1923.

[4] Qin, J.; Chen, L.; Liu, Y.; Liu, C.; Feng, C.; Chen, B. A Machine Learning Methodology for Diagnosing Chronic Kidney Disease. IEEE Access 2020, 8, 20991–21002.

[5] Ebiaredoh-Mienye, S.A.; Esenogho, E.; Swart, T.G. Integrating Enhanced Sparse Autoencoder-Based Artificial Neural Network Technique and Softmax Regression for Medical Diagnosis. Electronics 2020, 9, 1963.

[6] Chittora, P.; Chaurasia, S.; Chakrabarti, P.; Kumawat, G.; Chakrabarti, T.; Leonowicz, Z.; Jasin´ski, M.; Jasin´ski, Ł.; Gono, R.; Jasin´ska, E.; et al. Prediction of Chronic Kidney Disease—A Machine Learning Perspective. IEEE Access 2021, 9, 17312–17334.

[7] Tadist, K.; Najah, S.; Nikolov, N.S.; Mrabti, F.; Zahi, A. Feature selection methods and genomic big data: A systematic review. J. Big Data 2019, 6, 79.

[8] Pirgazi, J.; Alimoradi, M.; Esmaeili Abharian, T.; Olyaee, M.H. An Efficient hybrid filter- wrapper metaheuristic-based gene selection method for high dimensional datasets. Sci. Rep. 2019, 9, 18580.

[9] Nikravesh, F.Y.; Shirkhani, S.; Bayat, E.; Talebkhan, Y.; Mirabzadeh, E.; Sabzalinejad, M.; Aliabadi, H.A.M.; Nematollahi, L.; Ardakani, Y.H.; Sardari, S. Extension of human GCSF half-life by the fusion of albumin binding domain. Sci. Rep. 2022, 12, 667.

[10] Freund, Y.; Schapire, R.E. A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting. J. Comput. Syst. Sci. 1997, 55, 119–139.

[11] Akter, S.; Habib, A.; Islam, M.A.; Hossen, M.S.; Fahim, W.A.; Sarkar, P.R.; Ahmed, M. Comprehensive Performance Assessment of Deep Learning Models in Early Prediction and Risk Identification of Chronic Kidney Disease. IEEE Access 2021, 9, 165184 165206.

[12] Elkholy, S.M.M.; Rezk, A.; Saleh, A.A.E.F. Early Prediction of Chronic Kidney Disease Using Deep Belief Network. IEEE Access2021,9,135542–135549.