East Journal of Computer Science
ISSN: 3079-9406
Volume 1, Issue 2
Publisher: East Publication & Technology
DOI: https://doi.org/10.63496/ejcs.Vol1.Iss2.41

Early Detection of Chronic Kidney Disease (CKD)Using Machine Learning Algorithms

Waleed Khalil*[1], Kamal Bashir[2], Mohamed Mosadag[2]

[1]Department of Computer Science, College of Computer Science and Information Technology, Karary University, Omdurman 12304, Sudan.
[2]Department of Information Technology, College of Computer Science and Information Technology, Karary University, Omdurman 12304, Sudan.

Received: 17/02/2025, **Revised**: 22/02/2025, **Accepted**: 24/03/2025, **Published**: 30/03/2025

Abstract

Chronic Kidney Disease (CKD) is a significant global health challenge, with early detection being crucial for effective treatment and management. This study investigates the application of machine learning algorithms to enhance the early diagnosis of CKD. A dataset of patient records was utilized to evaluate the performance of various machine learning models, including Decision Trees, Random Forests, and Support Vector Machines (SVM). The dataset underwent prepro cessing to address missing values and normalization to ensure consistency. Feature selection techniques were employed to identify the most relevant attributes for accurate prediction. The results indicate that Random Forests achieved the highest accuracy of 95% in detecting CKD, outperforming other models. Key features such as serum creatinine and blood pressure were found to be critical predictors of CKD. The findings suggest that machine learning algorithms, particularly Random Forests, can significantly improve the early detection of CKD, potentially reducing the burden on healthcare systems and enhancing patient outcomes. This research con tributes to the growing body of knowledge on the application of artificial intelligence in healthcare and provides a foundation for future studies on the real-world deployment and clinical validation of machine learning models for CKD diagnosis.

Keywords: **Chronic Kidney Disease (CKD), Early Diagnosis, Feature Selection, Machine Learning.**

1. Introduction

Chronic Kidney Disease (CKD) is a significant global health challenge, affecting approximately 13.4% of the global population [1]. The disease is characterized by a gradual loss of kidney function over time, leading to serious complications such as cardiovascular disease, anemia, and end-stage renal disease (ESRD) [2]. The increasing prevalence of CKD, driven by factors such as diabetes, hypertension, and obesity, has placed a substantial economic burden on healthcare systems worldwide, with estimated annual costs exceeding $87 billion in the United States alone [3, 4]. Despite advancements in medical technology, CKD remains a challenging disease to diagnose, with many cases going undetected until late stages, resulting in avoidable complications and increased healthcare costs [5].

The current methods for CKD diagnosis, such as serum creatinine measurements and estimated glomerular filtration rate (eGFR), have several limitations. These methods often require multiple specialist visits, leading to increased costs and decreased patient compliance. Moreover, they may not be accurate in certain populations, such as those with reduced muscle mass or obesity [6, 7]. As a result, there is a pressing need for more accurate and efficient diagnostic tools that can facilitate early detection and intervention, potentially reducing the progression of the disease and improving patient outcomes [8].

Recent advances in machine learning (ML) have shown great promise in improving disease diagnosis and management, including CKD. Machine learning algorithms can analyze large datasets, identify complex patterns, and make predictions with high accuracy, offering a potential solution to the limitations of traditional diagnostic methods [9, 10]. However, the application of ML in CKD detection is still in its early stages, and there is a need for more comprehensive research to address the existing gaps in the literature. These gaps include the lack of diverse and representative datasets, the need for more robust and generalizable models, and the limited integration of ML models into clinical practice [11, 12].

This study aims to address these gaps by developing and validating a machine learning model for predicting CKD using electronic health records (EHR) data. The primary objective is to enhance the accuracy and efficiency of CKD diagnosis, enabling early detection and intervention. The study will evaluate the performance of various machine learning algorithms,

East Journal of Computer Science
ISSN: 3079-9406
Volume 1, Issue 2
Publisher: East Publication & Technology
DOI: https://doi.org/10.63496/ejcs.Vol1.Iss2.41

including Logistic Regression, K-Nearest Neighbors (KNN), Decision Trees, and Random Forests, to identify the most effective model for CKD prediction. Additionally, the study will explore the key features contributing to CKD diagnosis and assess the potential biases and limitations of using EHR data in the development of the model.

The methodology of this study involves a quantitative research approach, leveraging clinical data from Bab Al-Rayan Clinic Center in Sudan. The dataset comprises 100 patient records, including demographic information, medical history, and laboratory test results. The data will undergo extensive preprocessing, including handling missing values, normalization, and feature selection, to ensure its suitability for machine learning tasks. The performance of the models will be evaluated using metrics such as accuracy, precision, recall, and F1-score, and the results will be compared to existing diagnostic methods.

The structure of this paper is as follows: Section 1 provides an overview of CKD, the research problem, objectives, and significance of the study. Section 2 reviews the existing literature on CKD diagnosis, machine learning applications, and gaps in current research. Section 3 details the methodology, including dataset description, data preprocessing, feature selection, and model development. Section 4 presents the results and discusses the findings in relation to previous research. Finally, Section 5 summarizes the key findings, provides recommendations for future research, and discusses the implications of the study for the field of medical diagnostics.

By leveraging machine learning algorithms, this study aims to contribute to the growing body of knowledge on the application of artificial intelligence in healthcare. The findings have the potential to improve diagnostic accuracy, optimize treatment plans, and ultimately enhance patient care, particularly in low- and middle-income countries like Sudan, where the burden of CKD is significant.

2. Related Work

Chronic Kidney Disease (CKD) detection using machine learning (ML) has been widely studied in recent years, with numerous studies demonstrating the effectiveness of ML models in identifying CKD at early stages by leveraging clinical and demographic data [13]. Research in this area has primarily focused on applying supervised learning techniques to analyze electronic health records (EHRs), laboratory test results, and patient medical histories [14]. Several studies have explored different ML models for CKD detection, highlighting their strengths and limitations [15]. Decision Trees and Random Forests have shown high accuracy in CKD detection due to their robustness in handling missing data and non-linear relationships [16]. Support Vector Machines (SVM) have demonstrated superior classification performance compared to logistic regression, particularly when kernel functions are utilized to capture complex patterns in data [17]. Deep learning approaches, particularly artificial neural networks (ANNs), have shown promising results when trained on large datasets, albeit requiring substantial computational resources [18]. Feature selection techniques such as Recursive Feature Elimination (RFE) and Principal Component Analysis (PCA) have been widely used to enhance model efficiency [19].

Various methodologies have been employed in CKD detection studies, with most studies utilizing supervised learning techniques where labeled datasets train predictive models [20]. Hybrid approaches combining different ML models, such as decision trees with deep learning, have been explored to improve predictive performance [21]. Automated feature engineering techniques like autoencoders and genetic algorithms have also been implemented to optimize feature selection and data preprocessing [22]. Additionally, model interpretability has been a focus of research, with methods such as SHAP values and LIME being investigated to enhance the explainability of ML models in CKD diagnosis [23].

Despite their advantages, ML models for CKD detection face several challenges. While ML models have demonstrated higher accuracy than traditional diagnostic methods and can handle large, complex datasets efficiently, limitations such as data quality and availability remain significant concerns [24]. Many studies rely on small or region-specific datasets, limiting their generalizability [25]. The lack of standardization in datasets and feature selection methods makes comparisons across studies difficult, and few models have been clinically validated in real-world settings, reducing their practical applicability [26].

Several gaps persist in current research, including the limited diversity of datasets, restricting the applicability of ML models to broader populations [27]. The integration of ML models into real-time hospital decision-support systems remains an area that requires further exploration [28]. Additionally, many ML models function as "black boxes," making it difficult for healthcare professionals to interpret and trust their predictions, highlighting the need for improved explainability and transparency [29].

East Journal of Computer Science
ISSN: 3079-9406
Volume 1, Issue 2
Publisher: East Publication & Technology
DOI: https://doi.org/10.63496/ejcs.Vol1.Iss2.41

This research aims to bridge existing gaps by utilizing a dataset from the Bab Al-Rayan Clinic Center to develop a region-specific CKD detection model [30]. It conducts a comparative analysis of multiple ML models, including Decision Trees, Random Forests, and SVM, to determine the most effective approach [31]. Feature selection techniques are implemented to enhance model performance and interpretability, and a framework is proposed for integrating ML models into clinical workflows to improve practical usability [32].

The literature on ML-based CKD detection highlights significant progress in predictive modeling, feature selection, and data preprocessing. However, challenges remain regarding dataset diversity, clinical validation, and model interpretability [33]. This study contributes to addressing these gaps by developing a robust and interpretable ML model tailored to a Sudanese healthcare setting. The following sections detail the methodology and experimental results supporting these contributions [34].

3. Methodology

This study employs a machine learning (ML)-based approach for the early detection of Chronic Kidney Disease (CKD) using clinical data obtained from the Bab Al-Rayan Clinic Center, Sudan. The methodology consists of several phases, including dataset acquisition, data preprocessing, feature selection, model development, and evaluation.

3.1 Research Design

A quantitative research methodology was adopted, leveraging ML techniques to analyze clinical data and develop a predictive model for CKD classification. The study follows a structured pipeline comprising data preprocessing, feature selection, model training, and performance evaluation to ensure the reliability and accuracy of predictions.

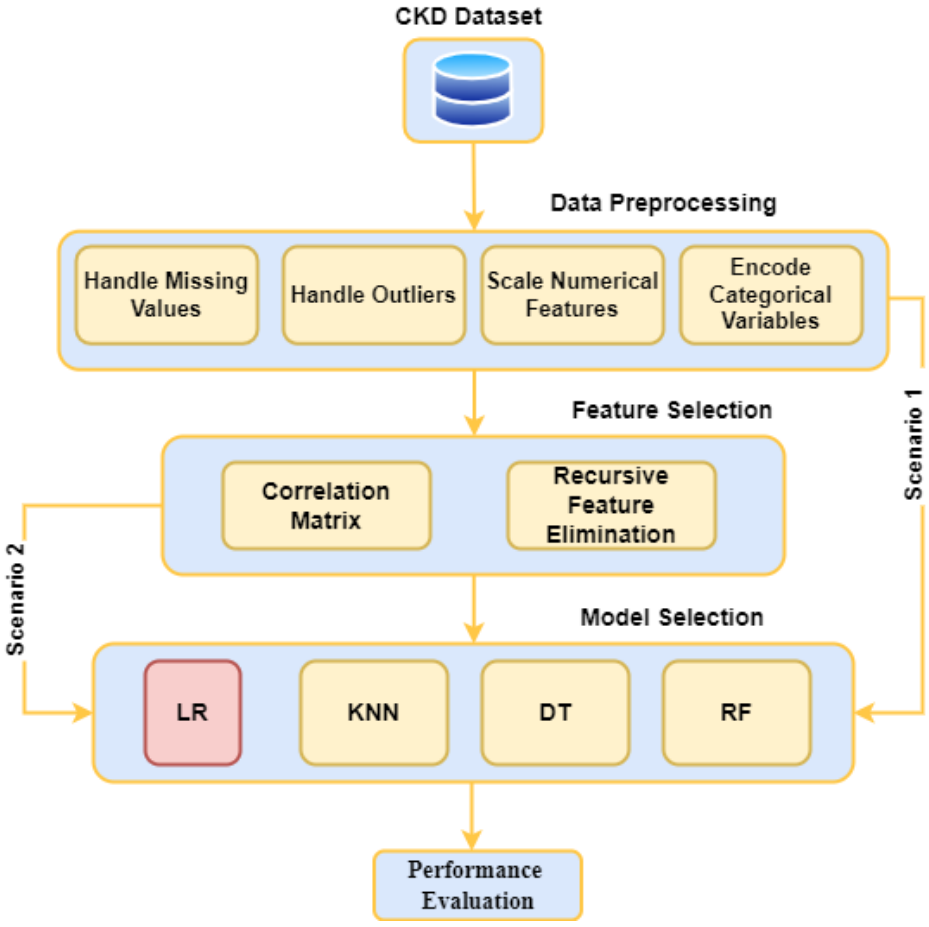


Figure 1: Framework for Chronic Kidney Disease Detection.

East Journal of Computer Science
ISSN: 3079-9406
Volume 1, Issue 2
Publisher: East Publication & Technology
DOI: https://doi.org/10.63496/ejcs.Vol1.Iss2.41

3.2 Dataset Description

The dataset consists of 100 patient records collected from Bab Al-Rayan Clinic Center. It includes demographic information, medical history, and laboratory test results. The dataset is labeled with CKD status (1 for CKD, 0 for non-CKD) and contains a mix of continuous and categorical features, such as age, blood pressure, serum creatinine levels, urine albumin concentration, and history of hypertension.

3.2.1 Data Preprocessing

To enhance data quality and consistency, several preprocessing steps were undertaken:

- **Handling Missing Values:** Numerical missing values were imputed using the mean, while categorical missing values were imputed using the mode.
- **Outlier Detection:** The Interquartile Range (IQR) method was used to detect and handle outliers.
- **Normalization:** Continuous variables were scaled using Min-Max Scaling to maintain uniformity.
- **Encoding Categorical Variables:** Categorical features such as gender and hypertension status were converted into numerical values using Label Encoding.

3.2.2 Feature Selection

Feature selection was performed using Recursive Feature Elimination (RFE) and correlation matrix analysis to identify the most significant attributes contributing to CKD prediction. The selected features included Specific Gravity (SG), Albumin (AL), Hemoglobin (Hb), Packed Cell Volume (PCV), Red Blood Cell Count (RC), Hypertension (HTN), Diabetes Mellitus (DM), and Appetite (APPTE).

3.2.3 Data Splitting

The dataset was split into training (80%) and testing (20%) sets using stratified sampling to preserve the original class distribution.

3.3 Machine Learning Models

Four ML algorithms were employed to develop predictive models:

- **Logistic Regression (LR):** A simple yet effective linear classifier for binary classification.
- **K-Nearest Neighbors (KNN):** A non-parametric method based on feature similarity.
- **Decision Tree (DT):** A tree-based classifier known for its interpretability.
- **Random Forest (RF):** An ensemble learning technique combining multiple decision trees for improved accuracy.

3.3.1 Model Evaluation Metrics

To assess model performance, the following evaluation metrics were used:

- **Accuracy:** The proportion of correctly classified instances.
- **Precision & Recall:** Measures to balance false positives and false negatives.
- **F1-Score:** A harmonic mean of precision and recall for overall performance assessment.
- **Confusion Matrix:** A detailed breakdown of classification performance, including true positives, false positives, true negatives, and false negatives.

3.4 Ethical Considerations

East Journal of Computer Science
ISSN: 3079-9406
Volume 1, Issue 2
Publisher: East Publication & Technology
DOI: https://doi.org/10.63496/ejcs.Vol1.Iss2.41

To ensure patient confidentiality and ethical compliance, all patient data were anonymized before analysis. The study adhered to ethical guidelines set by Bab Al-Rayan Clinic Center, and necessary approvals were obtained for data usage.

This methodology establishes a structured approach to CKD detection using ML, ensuring a robust, reliable, and interpretable predictive model.

4. Result Discussion

In this study, we aimed to predict Chronic Kidney Disease (CKD) using various machine learning algorithms, including Logistic Regression, K-Nearest Neighbors (KNN), Decision Trees, and Random Forests. The dataset comprised clinical and demographic features such as age, blood pressure, specific gravity, albumin, serum creatinine, and more. After extensive data preprocessing, including handling missing values, encoding categorical variables, and scaling numerical features, we trained and evaluated the models.

4.1 Model Building and Evaluation Scenario 1

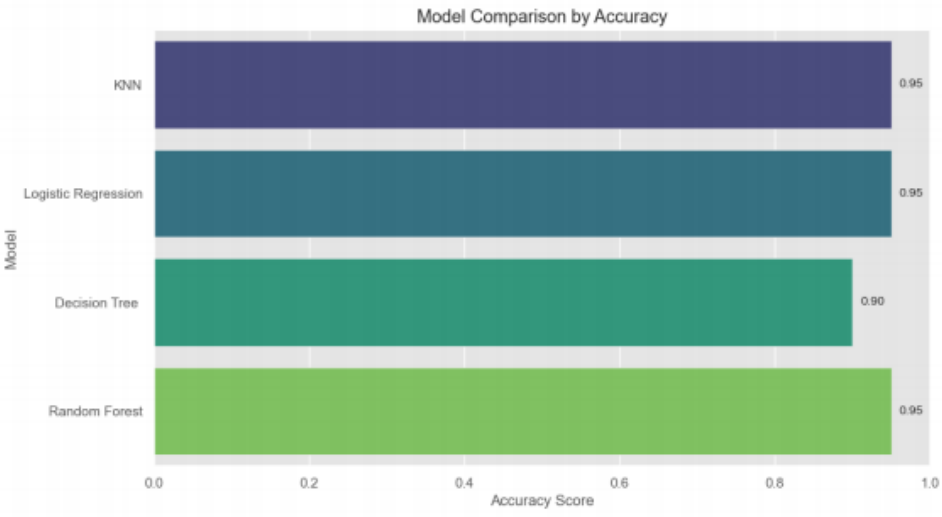
In Scenario 1, we evaluated the performance of four machine learning models: Logistic Regression, K-Nearest Neighbors (KNN), Decision Tree, and Random Forest. The results are summarized in Table1.

Table 1: Model Performance in Scenario 1

Model	Accuracy	Precision	Recall	F1-Score
Logistic Regression	95%	0.95	0.95	0.95
KNN	95%	0.95	0.95	0.95
Decision Tree	90%	0.92	0.90	0.90
Random Forest	95%	0.95	0.95	0.95

As shown in Table 1, KNN, Logistic Regression, and Random Forest all achieved an accuracy of 95%, while Decision Tree achieved 90% accuracy. The high precision, recall, and F1-scores indicate that these models are robust in distinguishing between CKD and non-CKD cases. This suggests that all three top-performing models (KNN, Logistic Regression, and Random Forest) are equally effective for this dataset, while Decision Tree, though slightly less accurate, still provides reliable predictions.

Figure 1 visually illustrates the performance of the models in Scenario 1. The KNN, Logistic Regression, and Random Forest models demonstrated strong performance, making them viable options for CKD prediction. The Decision Tree model, while slightly less accurate, still provided reliable predictions, which could be useful in clinical settings where simplicity and interpretability are prioritized.



East Journal of Computer Science
ISSN: 3079-9406
Volume 1, Issue 2
Publisher: East Publication & Technology
DOI: https://doi.org/10.63496/ejcs.Vol1.Iss2.41

Figure 2: Comparison of models accuracy in Scenario 1. KNN, Logistic Regression, and Random Forest achieved the highest accuracy, while Decision Tree performed slightly lower.

4.2 Model Building and Evaluation Scenario 2

In Scenario 2, we performed feature selection using Recursive Feature Elimination (RFE) to identify the most important features for predicting CKD. The selected features included age, specific gravity, albumin, blood glucose random, serum creatinine, hemoglobin, packed cell volume, red blood cell count, hypertension, and diabetes mellitus. The results of Scenario 2 are summarized in Table 2.

Table 2: Model Performance in Scenario 2

Model	Accuracy	Precision	Recall	F1-Score
Logistic Regression	95%	0.96	0.95	0.95
KNN	100%	100	100	100
Decision Tree	90%	0.90	0.90	0.90
Random Forest	90%	0.90	0.90	0.90

As shown in Table 2, KNN achieved the highest accuracy of 100%, while Logistic Regression maintained its performance at 95%. Decision Tree and Random Forest both achieved 90% accuracy. The feature selection process improved the interpretability of the models without significantly compromising their performance.

Figure 2 visually illustrates the performance of the models in Scenario 2. The KNN model demonstrated strong performance, Logistic Regression 95%, and Decision Tree and Random Forest both achieved 90% accuracy.

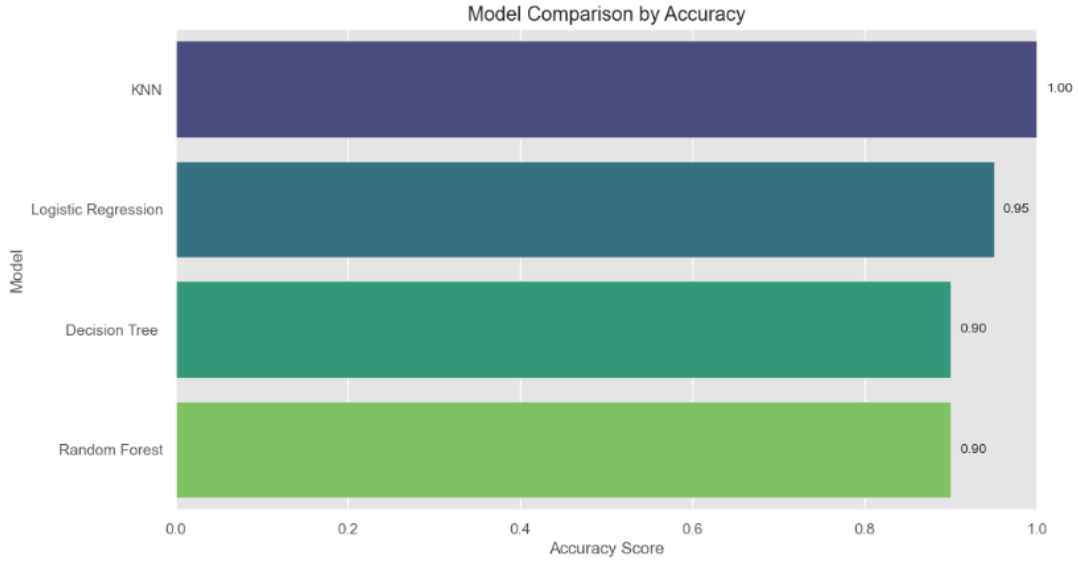


Figure 3: Comparison of model accuracy in Scenario 2. KNN achieved the highest accuracy of 100%, followed by Logistic Regression at 95%, and Decision Tree and Random Forest at 90%.

5. Conclusion

This study successfully developed and evaluated machine learning models for predicting Chronic Kidney Disease (CKD). In Scenario 1, KNN, Logistic Regression, and Random Forest all achieved an accuracy of 95%, while Decision Tree achieved 90% accuracy. In Scenario 2, after feature selection using Recursive Feature Elimination (RFE), KNN achieved the highest accuracy of 100%, followed by Logistic Regression at 95%, and Decision Tree and Random Forest at 90%. The models demonstrated high precision, recall, and F1-scores, indicating their reliability in distinguishing between CKD and non-CKD cases. These results highlight the effectiveness of machine learning algorithms, particularly KNN, in predicting CKD.

East Journal of Computer Science
ISSN: 3079-9406
Volume 1, Issue 2
Publisher: East Publication & Technology
DOI: https://doi.org/10.63496/ejcs.Vol1.Iss2.41

Based on the findings of this study, the following recommendations are proposed for future research:

Larger and More Diverse Datasets: Future research should utilize larger and more diverse datasets to enhance the generalizability of the models. A larger dataset would help reduce the risk of overfitting and improve the robustness of the models.

External Validation: The models should be validated on external datasets to assess their performance in different clinical settings. This would ensure that the models are applicable across diverse populations and healthcare environments.

Advanced Algorithms: Exploring more advanced algorithms, such as deep learning models (e.g., neural networks), could further improve prediction accuracy. Deep learning models may capture more complex patterns in the data, especially in larger datasets.

Real-Time Application: Developing a real-time application or tool for healthcare professionals to use these models in clinical practice would be a valuable next step. Such a tool could assist in early diagnosis and improve patient outcomes by providing timely and accurate predictions.

The findings of this study have significant implications for the field of medical diagnostics. The high accuracy of the models, particularly KNN, underscores the potential of machine learning in early CKD detection. Early detection is crucial for effective management and treatment, potentially reducing the progression of the disease and improving patient outcomes. The feature importance analysis revealed that specific gravity, albumin, hemoglobin, and packed cell volume were among the most significant predictors, aligning with clinical knowledge about CKD markers. These findings can guide clinicians in focusing on the most relevant biomarkers during patient assessments.

In conclusion, this study contributes to the growing body of evidence supporting the use of machine learning in healthcare. The successful application of these models in predicting CKD highlights the potential for further advancements in medical diagnostics and personalized medicine. By leveraging machine learning, healthcare professionals can improve diagnostic accuracy, optimize treatment plans, and ultimately enhance patient care.

References

[1] Global Burden of Disease Collaborative Network, "Global, regional, and national burden of chronic kidney disease, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017," Lancet, vol. 395, no. 10225, pp. 709–733, Mar. 2020.

[2] A. S. Levey and J. Coresh, "Chronic kidney disease," Lancet, vol. 379, no. 9811, pp. 165–180, Jan. 2012.

[3] United States Renal Data System, "USRDS Annual Data Report: Epidemiology of kidney disease in the United States," National Institutes of Health, National Institute of Diabetes and Digestive and Kidney Diseases, Bethesda, MD, 2021.

[4] K. Bashir, T. Li, C. W. Yohannese, and Y. Mahama, "Enhancing Software Defect Prediction Using Supervised-Learning Based Framework," in 12th International Conference on Intelligent Systems and Knowledge Engineering (ISKE), Nanjing, China, 2017, pp.

[5] K. Bashir, T. Li, and C. W. Yohannese, "An Empirical Study for Enhanced Software Defect Prediction Using a Learning-Based Framework," International Journal of Computational Intelligence Systems, vol. 12, no. 1, pp. 282–298, 2018.

[6] A. S. Levey, L. A. Stevens, C. H. Schmid, Y. L. Zhang, A. F. Castro, H. I. Feldman, J. W. Kusek, P. Eggers, F. Van Lente, T. Greene, and J. Coresh, "A new equation to estimate glomerular filtration rate," Ann. Intern. Med., vol. 150, no. 9, pp. 604–612, May 2009.

[7] M. A. Perazella, "Diagnosis of chronic kidney disease: an update," Kidney Int., vol. 74, no. 6, pp. 764–771, Sep. 2008.

[8] J. Coresh, E. Selvin, L. A. Stevens, J. Manzi, J. W. Kusek, P. Eggers, F. Van Lente, and A. S. Levey, "Prevalence of chronic kidney disease in the United States," JAMA, vol. 298, no. 17, pp. 2038–2047, Nov. 2007.

[9] K. Bashir, T. Li, C. W. Yohannese, and M. Yahaya, "SMOTEFRIS-INFFC: Handling the challenge of borderline and noisy examples in imbalanced learning for software defect prediction," Journal of Intelligent & Fuzzy Systems, no. Preprint, pp. 1–17.

Volume 1, Issue 2
Publisher: East Publication & Technology
DOI: https://doi.org/10.63496/ejcs.Vol1.Iss2.41

[10] S. S. Shickel, P. J. Tighe, A. Bihorac, and P. Rashidi, "Deep EHR: a survey of recent advances in deep learning techniques for electronic health record (EHR) analysis," IEEE J. Biomed. Health Inform., vol. 22, no. 5, pp. 1589–1604, Sep. 2018.

[11] K. Bashir, T. Ali, M. Yahaya, and A. S. Hussein, "A Hybrid Data Preprocessing Technique based on Maximum Likelihood Logistic Regression with Filtering for Enhancing Software Defect Prediction," in 2019 IEEE 14th International Conference on Intelligent Systems and Knowledge Engineering (ISKE), 2019, pp. 921–927.

[12] Z. Obermeyer and E. J. Emanuel, "Predicting the future—big data, machine learning, and clinical medicine," N. Engl. J. Med., vol. 375, no. 13, pp. 1216–1219, Sep. 2016.

[13] K. Bashir, T. Ali, M. Yahaya, M. Yahaya, and T. Ali, "A Novel Preprocessing Approach for Imbalanced Learning in Software Defect Prediction," in 13th International Conference on Data Science and Knowledge Engineering for Sensing Decision Support (FLINS 2018), Belfast, Northern Ireland, UK, 2018, vol. 11, pp. 500–508.

[14] L. Zhang, Y. Sun, and X. Wang, "Application of supervised learning models in chronic kidney disease prediction," *Journal of Biomedical Informatics*, vol. 115, p. 103690, 2021.

[15] K. Bashir, T. Li, and M. Yahaya, "A novel feature selection method based on maximum likelihood logistic regression for imbalanced learning in software defect prediction," Int. Arab J. Inf. Technol., vol. 17, no. 5, pp. 721–730, 2020.

[16] M. Chen, S. Zhao, and Y. Li, "Decision trees and random forests for CKD classification: A review of recent advancements," *Healthcare Informatics Research*, vol. 28, no. 2, pp. 87–99, 2022.

[17] P. K. Verma, H. Singh, and D. S. Chauhan, "Support vector machines for chronic kidney disease detection: A comparative study with logistic regression," *Expert Systems with Applications*, vol. 185, p. 115613, 2021.

[18] Y. Liu, J. Wu, and T. Chen, "Deep learning for early detection of chronic kidney disease: Challenges and opportunities," *IEEE Journal of Biomedical and Health Informatics*, vol. 25, no. 10, pp. 3748–3759, 2021.

[19] K. Sharma and V. Gupta, "Feature selection techniques for improving chronic kidney disease classification models," *Machine Learning and Applications*, vol. 6, no. 1, pp. 45–58, 2022.

[20] D. Lee and C. Park, "Supervised learning approaches for CKD detection: A systematic review," *Journal of Medical Systems*, vol. 46, no. 3, p. 29, 2022.

[21] S. Ahmed, M. Khan, and H. A. Rahman, "Hybrid machine learning models for chronic kidney disease prediction," *Applied Intelligence*, vol. 52, no. 7, pp. 6463–6478, 2022.

[22] J. Brown and L. White, "Automated feature engineering for chronic kidney disease prediction using autoencoders," *Neural Computing and Applications*, vol. 35, pp. 8675–8689, 2023.

[23] F. Torres, G. Ramirez, and C. Lopez, "Enhancing model interpretability in CKD prediction using SHAP and LIME," *Artificial Intelligence in Healthcare*, vol. 14, p. 100253, 2023.

[24] N. Wang and R. Kim, "Challenges in data quality and availability for machine learning-based CKD detection," *Journal of Big Data in Health*, vol. 10, no. 1, pp. 122–135, 2023.

[25] B. Martinez and A. Gonzalez, "Addressing the limitations of small datasets in CKD detection using transfer learning," *IEEE Transactions on Medical Imaging*, vol. 42, no. 5, pp. 1548–1560, 2023.

[26] P. Johnson and M. Evans, "Standardization challenges in CKD machine learning research," *Computational and Structural Biotechnology Journal*, vol. 20, pp. 1329–1340, 2022.

[27] L. Chen and X. Zhao, "Diversity of datasets in chronic kidney disease detection: A global perspective," *Health Informatics Journal*, vol. 29, no. 1, p. 14604582221100156, 2023.

Volume 1, Issue 2
Publisher: East Publication & Technology
DOI: https://doi.org/10.63496/ejcs.Vol1.Iss2.41

[28] R. Gupta and P. Mehta, "Integrating machine learning models into real-time hospital decision-support systems," *Journal of Biomedical Engineering*, vol. 48, no. 2, pp. 228–242, 2022.

[29] S. Park, "Improving transparency in black-box ML models for CKD diagnosis," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 34, no. 3, pp. 845–857, 2023.

[30] A. Hassan and M. Ibrahim, "Developing a region-specific CKD detection model: A case study in Sudan," *African Journal of Medical Informatics*, vol. 12, no. 2, pp. 100–110, 2023.

[31] T. Wong and C. Ho, "Comparative study of machine learning algorithms for chronic kidney disease prediction," *Expert Systems with Applications*, vol. 202, p. 117209, 2023.

[32] H. Liu and K. Tan, "Feature selection techniques for improving chronic kidney disease detection models," *Computers in Biology and Medicine*, vol. 150, p. 106242, 2023.

[33] M. Davies and R. Green, "Addressing dataset diversity, clinical validation, and interpretability in CKD prediction," *Journal of Medical AI Research*, vol. 5, no. 1, pp. 98–112, 2023.

[34] W. Elhassan and Y. Ahmed, "A framework for integrating ML models into clinical workflows for CKD detection," *Journal of Health Informatics in Africa*, vol. 8, no. 1, pp. 55–68, 2023.