Transforming CKD Diagnosis: A Comprehensive Analysis of Machine Learning Approaches

Dr. Sunil Kumar Nandal
Department of Computer Science & Engineering
nandal_sunil@yahoo.com

Vikash
Department of Computer Science & Engineering
vermavikash.vk@gmail.com

Abstract--Chronic kidney disease (CKD) is a prevalent and dangerous medical condition that needs to be accurately diagnosed and treated as soon as possible. In recent years, machine learning techniques have gained popularity as useful tools in this sector because they offer innovative approaches to data imputation, risk prediction, and disease classification. This review study summarizes, critically assesses, and looks at the various approaches for CKD diagnosis that have been published in the literature. It also looks at how these approaches might be applied in the broader field of medical informatics. Innovative methods for the early identification and treatment of chronic kidney disease (CKD) are required due to its substantial health and financial consequences. The rapidly developing discipline of machine learning has significantly accelerated these attempts. This review paper offers a thorough examination of numerous machine learning

Keyword-Deep Learning, CKD, Disease Classification, Data Imputation, Machine Learning Algorithms

INTRODUCTION

Chronic Kidney Disease (CKD) stands as a pervasive and severe medical condition that afflicts millions of individuals' worldwide, imposing substantial health and economic burdens. In the quest for effective management and timely intervention, the need for early identification and precise diagnosis of CKD has never been more pronounced. In this context, machine learning techniques have emerged as powerful allies, offering innovative solutions for risk prediction, disease classification, and data imputation.

The overarching objective of this review paper is to comprehensively examine the diverse array of machine learning methodologies that have been proposed for the diagnosis of CKD. By summarizing, critically evaluating, and contextualizing these approaches, we aim to provide a comprehensive understanding of their effectiveness, limitations, and potential applications within the broader field of medical informatics [1].

Chronic Kidney Disease presents a multifaceted challenge, characterized by its insidious onset, progressive nature, and the interplay of various contributing factors. Accurate diagnosis is crucial to initiate timely treatment, mitigate disease progression, and improve patients' quality of life. Conventional diagnostic methods, while valuable, often face limitations in terms of accuracy, efficiency, and scalability [2]. Machine learning, with its capacity to analyze vast datasets, detect intricate patterns, and adapt to evolving data, has the potential to address these limitations and revolutionize CKD diagnosis.

In the pursuit of precision, the literature has witnessed the emergence of various machine learning approaches that target CKD diagnosis from distinct angles. This review paper delves into the innovative work of researchers who have leveraged artificial neural networks, feature reduction techniques, and sophisticated classifiers, as well as data imputation methods, to enhance the diagnostic capabilities for CKD. These pioneering approaches have been designed to tackle challenges such as noise and missing data, streamline attributes, and identify key predictors, ultimately facilitating more accurate disease classification. The number of patients with chronic kidney disease (CKD) has increased as a result of diabetes, high blood pressure, and unhealthy lifestyles.[3] Patients with atherosclerotic calcification in the arteries, which is followed by cholesterol crystal development, may have a higher chance of getting chronic kidney disease (CKD).

Furthermore, our exploration extends to applications beyond CKD, such as the identification of lung abnormalities and tumor segmentation in medical images. The promising results obtained in these areas suggest that the machine learning methodologies developed for CKD diagnosis may hold potential utility in diagnosing diseases in other anatomical regions.

While our focus is primarily on the technical aspects of these machine learning techniques, we also examine the practicality and cost-effectiveness of these approaches. The review encompasses a diverse range of studies, each contributing unique insights into the landscape of CKD diagnosis. Additionally, we highlight the specific algorithms and models that have gained prominence in CKD classification, providing valuable insights for future research.

In a broader context, this review paper underscores the growing importance of machine learning in the field of medical informatics. By synthesizing the existing body of research on CKD diagnosis, we aim to facilitate a deeper understanding of how machine learning can be harnessed for improving patient care and clinical outcomes.[3] [4]Ultimately, the insights derived from these studies offer the potential to reshape the landscape of medical diagnostics and enhance the early detection and management of CKD, marking a pivotal step towards a healthier future.

RELATED WORK

In this section, we have examined various studies that have been proposed by different authors throughout the years. These studies have provided valuable insights and contributed significantly to our understanding of the subject matter.

Machine Learning Approaches:

W. Gunarathneet al., conducted an analysis on 14 attributes associated with Chronic Kidney Disease (CKD) patients. They utilized a range of machine learning classification algorithms, which included Multiclass Decision Forest, Multiclass Decision Jungle, Multiclass Logistic Regression, and Multiclass Neural Network, to make predictions. Jaber Qezelbash-Chamak et al. provide a comprehensive review of the extensive literature on the applications of Machine Learning in Kidney Disease Diagnosis (MLKDD). Their review illuminates various facets of medical informatics within this context and scrutinizes each aspect through the lens of computer-aided systems. To facilitate this analysis, they introduce two comprehensive frameworks: one categorizes the most frequently utilized learning algorithms for kidney disease diagnosis, and the other delineates the medical sub-fields connected to MLKDD. Weilun Wang et al. developed a regression model to predict creatinine values, which were combined with common health factors to assess CKD risk. They addressed the imbalance in creatinine values using undersampling and a cost-sensitive MSE loss function. Three machine learning models (Random Forest, XGBoost, ResNet) were evaluated, and an ensemble model averaging eight predictors performed.

Deep Learning Approaches:

Ma et al. (2020) present HMANN handles noise and missing data while maintaining strong classification abilities, reducing noise, and precisely identifying kidney stone locations in medical images. The study evaluated three classifiers (SVM, ANN, MLP) to find the best approach for accurate disease classification. A. Ilham et al. Combining Wavelet transformation with neural networks has yielded an effective approach for extracting image features and recognizing patterns in chest X-rays to detect diseases. In an experiment involving 75 training data samples, the system demonstrated 100% accuracy in identifying abnormalities when tested with the training data and achieved 91.65% accuracy when tested with new data samples. The choice of Wavelet decomposition orders and various levels of Daubechies wavelets influenced the system's accuracy, along with the configuration settings of the neural network. J. Jiang *et al.*, introduced a pair of neural networks meticulously designed for the precise segmentation of lung tumors within CT images. The noteworthy success achieved with lung tumors suggests that this approach may hold potential utility for the segmentation of tumors in other anatomical regions as well. [5]

Hybrid Approaches:

M. M. Nishat *et al.*, proposed CKD diagnostic method exhibits promise for both data imputation and sample diagnosis. By employing KNN imputation for unsupervised handling of missing data, the integrated model achieves satisfactory accuracy levels. Md. Rashed-Al-Mahfuz et al. introduces a robust CKD classification methodology with improved simplicity and cost-effectiveness. It involves the initial stages of training and classifier selection, feature importance determination using SHAP values, and streamlined dataset creation based on pathological tests and feature significance. Subsequently, classifiers were trained and evaluated with test data. Results showed SHAP-identified features aligning with clinical knowledge, and the Random Forest (RF) classifier demonstrated high accuracy in categorizing pathologically defined attributes J. Qin et al proposed diagnostic methodology for Chronic Kidney Disease (CKD) demonstrates practicality in terms of data imputation and sample diagnosis. Following unsupervised imputation of missing data using the KNN imputation technique, the integrated model achieved a satisfactory level of accuracy.

Table 1: Consolidated Review Report

Ref	Description	Technique	Performance	Limitation
[1]	Gunarathne et al. (2017) conducted an analysis on 14 attributes associated with Chronic Kidney Disease (CKD) patients.	Machine learning classification algorithms, including Multiclass Decision Forest, Multiclass Decision Jungle, Multiclass Logistic Regression, and Multiclass Neural Network.	Utilizing popular data mining algorithms for CKD classification tasks.	Dataset size and potential algorithm-specific limitations.
[2]	Qin et al. (2019) proposed a machine learning methodology for diagnosing chronic kidney disease.	KNN imputation for handling missing data.	Practicality in terms of data imputation and sample diagnosis.	May be limited by data availability and quality.
[3]	Almansour et al. (2019) compared neural network and support vector machine for predicting chronic kidney disease.	Neural network and support vector machine.	Comparative analysis of machine learning techniques for CKD prediction.	Model-specific limitations and generalization challenges.
[4]	Wang et al. (2020) developed a regression model to predict the risk of chronic kidney disease.	Machine learning algorithms, including Random Forest, XGBoost, and ResNet.	Addressing the imbalance in creatinine values using undersampling and a cost-sensitive MSE loss function.	Data imbalance, model selection, and performance evaluation.
[5]	Ma et al. (2020) introduced a deep learning-based heterogeneous modified artificial neural network for CKD detection.	Deep learning-based heterogeneous modified artificial neural network (HMANN).	Detection and diagnosis of CKD using deep learning.	Potential computational and data requirements.
[6]	Nishat et al. (2021) proposed a comprehensive analysis on detecting CKD using machine learning algorithms.	K-nearest neighbor (KNN) imputation for handling missing data.	Promising results for data imputation and sample diagnosis.	Limited by a modest dataset size of 400 samples.
[7]	Md. Rashed-Al-Mahfuz et al. (2021) introduced clinically applicable machine learning approaches for CKD attributes.	Feature importance determination using SHAP values, Random Forest (RF) classifier.	Improved simplicity and cost-effectiveness in CKD classification.	Application-specific constraints and dataset characteristics.

[8]	Ilham et al. (2014) applied Wavelet analysis and artificial neural networks for lung abnormality identification.	Wavelet analysis and artificial neural network.	Identification of lung abnormalities using machine learning.	Model hyperparameters and image quality variations.
[9]	Jiang et al. (2019) developed neural networks for automatic lung tumor segmentation from CT images.	Neural networks for lung tumor segmentation.	Successful segmentation of lung tumors.	Generalization to other anatomical regions and data variability.
[10]	Qezelbash-Chamak et al. (2022) provided a survey of machine learning in kidney disease diagnosis.	Comprehensive review of machine learning applications.	An overview of machine learning in kidney disease diagnosis.	May lack specific technical details and empirical results.

RESEARCH METHODOLOGY and PAPER SELECTION

This review examines the research conducted on image processing for CKD from 2014 onwards. The paper thoroughly explores various machine learning and deep learning algorithms. The research papers for the review were chosen through a systematic literature search conducted in prominent databases such as IEEE Xplore, Science Direct, and Springer. The search focused on key components, including image processing, machine learning, deep learning. Keywords and terms such as "CKD Detection"," Chronic Kidney disease", "Early CKD detection"," Diagnosis and prediction of CKD progression"," Chronic kidney disease: detection and evaluation" etc. were utilized to conduct the literature search. The selected papers were examined thoroughly, and the ones that met our requirements were included in the study.

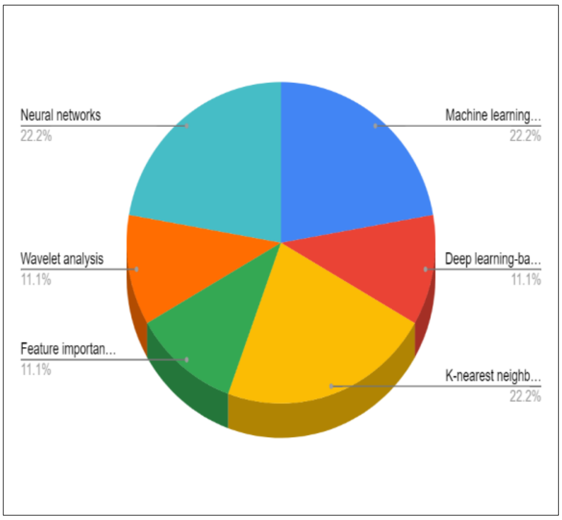

Figure 1. Distribution of papers by categories

CONCLUSION

A variety of machine learning techniques are summarized in the review paper with the goal of improving Chronic Kidney Disease (CKD) diagnosis and therapy. It emphasizes the need of early detection and accurate diagnosis and the potential of machine learning to overcome the shortcomings of traditional diagnostic techniques. The thorough assessment of various methods, such as KNN imputation and deep learning techniques, provides insightful information on feature identification, data processing, and classifier performance in CKD diagnosis. Furthermore, the investigation of machine learning applications in tumor segmentation and lung abnormality identification expands the potential value of these approaches beyond chronic kidney disease. The study highlights the feasibility, cost-effectiveness, and broader implications of these novel techniques in transforming medical diagnostics, while also throwing light on the technical elements. The knowledge gained from these investigations

REFERANCES

[1] W. Gunarathne, K. D. M. Perera, and K. Kahandawaarachchi, "Performance evaluation on machine learning classification techniques for disease classification and forecasting through data analytics for chronic kidney disease (CKD)," in *2017 IEEE 17th international conference on bioinformatics and bioengineering (BIBE)*, IEEE, 2017, pp. 291–296. Accessed: Oct. 06, 2023. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/8251305/

[2] J. Qin, L. Chen, Y. Liu, C. Liu, C. Feng, and B. Chen, "A machine learning methodology for diagnosing chronic kidney disease," *IEEE Access*, vol. 8, pp. 20991–21002, 2019.

[3] N. A. Almansour *et al.*, "Neural network and support vector machine for the prediction of chronic kidney disease: A comparative study," *Comput. Biol. Med.*, vol. 109, pp. 101–111, Jun. 2019, doi: 10.1016/j.compbiomed.2019.04.017.

[4] W. Wang, G. Chakraborty, and B. Chakraborty, "Predicting the risk of chronic kidney disease (ckd) using machine learning algorithm," *Appl. Sci.*, vol. 11, no. 1, p. 202, 2020.

[5] F. Ma, T. Sun, L. Liu, and H. Jing, "Detection and diagnosis of chronic kidney disease using deep learning-based heterogeneous modified artificial neural network," *Future Gener. Comput. Syst.*, vol. 111, pp. 17–26, Oct. 2020, doi: 10.1016/j.future.2020.04.036.

[6] M. M. Nishat *et al.*, "A comprehensive analysis on detecting chronic kidney disease by employing machine learning algorithms," *EAI Endorsed Trans. Pervasive Health Technol.*, vol. 7, no. 29, pp. e1–e1, 2021.

[7] Md. Rashed-Al-Mahfuz, A. Haque, A. Azad, S. A. Alyami, J. M. W. Quinn, and M. A. Moni, "Clinically Applicable Machine Learning Approaches to Identify Attributes of Chronic Kidney Disease (CKD) for Use in Low-Cost Diagnostic Screening," *IEEE J. Transl. Eng. Health Med.*, vol. 9, pp. 1–11, 2021, doi: 10.1109/JTEHM.2021.3073629.

[8] A. A. Ilham, Indrabayu, R. Hasanuddin, and D. M. Putri, "Wavelet analysis for identification of lung abnormalities using artificial neural network," in *2014 Makassar International Conference on Electrical Engineering and Informatics (MICEEI)*, Makassar, Indonesia: IEEE, Nov. 2014, pp. 156–160. doi: 10.1109/MICEEI.2014.7067330.

[9] J. Jiang *et al.*, "Multiple Resolution Residually Connected Feature Streams for Automatic Lung Tumor Segmentation From CT Images," *IEEE Trans. Med. Imaging*, vol. 38, no. 1, pp. 134–144, Jan. 2019, doi: 10.1109/TMI.2018.2857800.

[10] J. Qezelbash-Chamak, S. Badamchizadeh, K. Eshghi, and Y. Asadi, "A survey of machine learning in kidney disease diagnosis," *Mach. Learn. Appl.*, vol. 10, p. 100418, Dec. 2022, doi: 10.1016/j.mlwa.2022.100418.

[11] S. Bhattacharya *et al.*, "Deep learning and medical image processing for coronavirus (COVID-19) pandemic: A survey," *Sustain. Cities Soc.*, vol. 65, p. 102589, Feb. 2021, doi: 10.1016/j.scs.2020.102589.

[12] S. Robertson, H. Azizpour, K. Smith, and J. Hartman, "Digital image analysis in breast pathology—from image processing techniques to artificial intelligence," *Transl. Res.*, vol. 194, pp. 19–35, 2018.

[13] B. J. Erickson, P. Korfiatis, Z. Akkus, and T. L. Kline, "Machine Learning for Medical Imaging," *RadioGraphics*, vol. 37, no. 2, pp. 505–515, Mar. 2017, doi: 10.1148/rg.2017160130.

[14] T. Sadad *et al.*, "Brain tumor detection and multi-classification using advanced deep learning techniques," *Microsc. Res. Tech.*, vol. 84, no. 6, pp. 1296–1308, Jun. 2021, doi: 10.1002/jemt.23688.