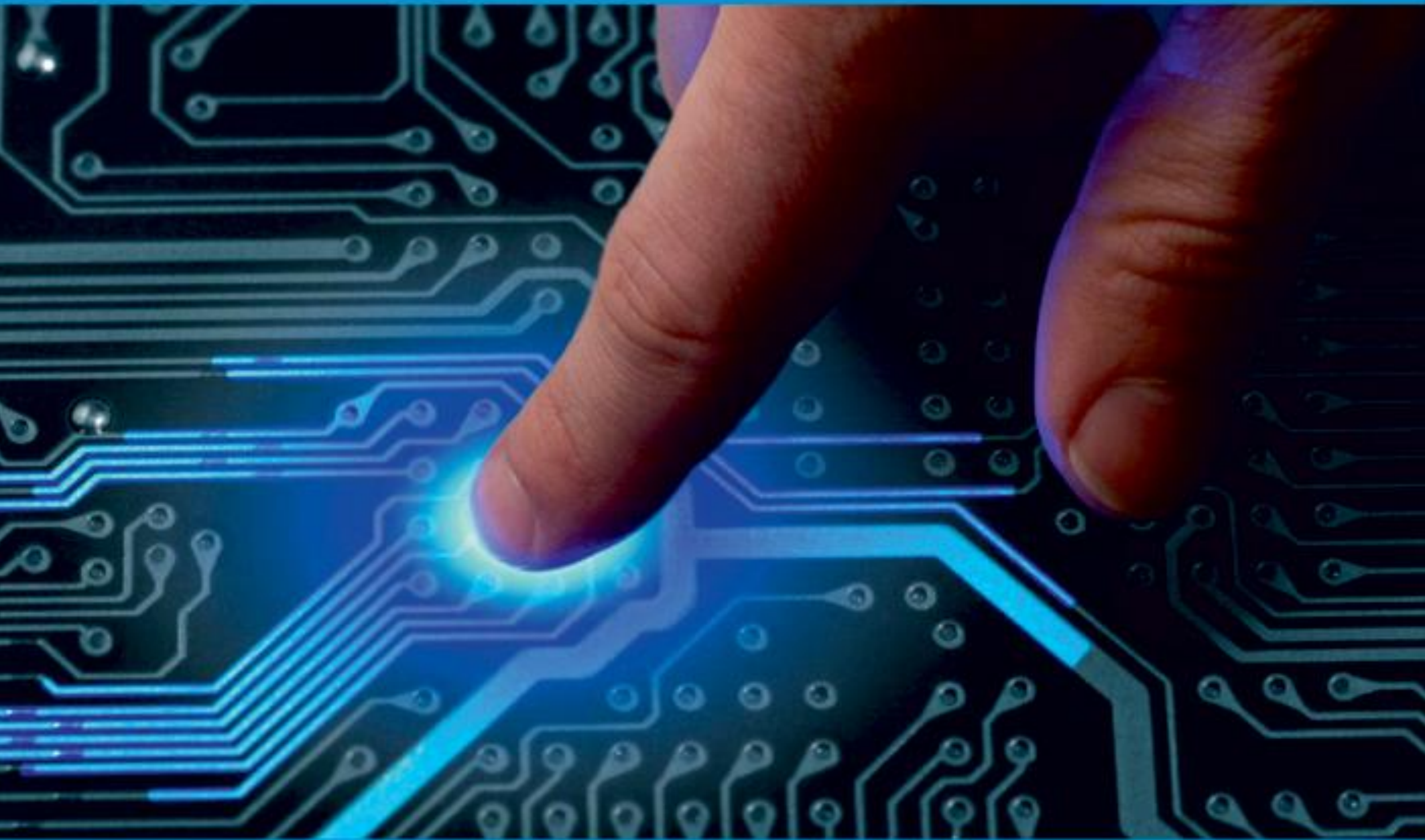
IJIRCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 12, Issue 5, May 2024

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 8.379

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | A Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 5, May 2024 ||

| DOI: 10.15680/IJIRCCE.2023.1205082 |

RenalVision: Predicting Chronic Kidney Disease with Machine Learning and Image Processing

Dr. Chethan L.S[1], Sri Harsha Kadam A[2], Deekshitha M[3], Sanjay M[4], Siri G B[5]

Associate Professor, Department of CSE, PES Institute of Technology and Management, Shivamogga, Karnataka, India[1]

U.G Student, Department of CSE, PES Institute of Technology and Management, Shivamogga, Karnataka, India[2-5]

ABSTRACT: Chronic kidney disease (CKD) presents a formidable challenge as the leading cause of mortality, claiming 1.7 million lives annually and escalating in incidence. Its insidious progression often evades early detection until advanced stages, underscoring the urgency for proactive screening methods. This study addresses this imperative by harnessing the power of data mining techniques to select salient features from datasets, optimizing predictive accuracy and expediting training processes. Leveraging user-input symptoms, our system employs Logistic Regression to accurately estimate disease probability. Moreover, the integration of the Pearson correlation model enhances parameter selection, refining predictive outcomes. Additionally, we incorporate image processing methodologies to classify renal conditions normal, cyst, kidney stone, and tumour diseases further augmenting diagnostic capabilities. In contrast to traditional diagnostic modalities, our approach capitalizes on the superior accuracy and adaptability of machine learning techniques. This cohesive strategy amalgamates data mining, Logistic Regression, and image processing, empowering early detection and proactive management of CKD, thereby potentially mitigating its profound impact.

KEYWORDS: Keywords: Chronic kidney disease, early detection, data mining, Logistic Regression, image processing, kidney tumours, predictive accuracy, machine learning, proactive screening, diagnostic capabilities.

I. INTRODUCTION

Machine Learning (ML) stands at the forefront of revolutionizing healthcare by extracting intricate medical insights and presenting innovative concepts to healthcare professionals. This project addresses the formidable challenge posed by Chronic Kidney Disease (CKD), a global health crisis with significant implications, through the utilization of ML and data-driven strategies. By meticulously analysing extensive patient data, spanning clinical, demographic, and lifestyle information, the project establishes robust predictive models. These models not only facilitate early CKD diagnosis but also assess risks and tailor personalized treatment plans to individual patients. Prioritizing transparency, the project integrates explainable AI techniques to enhance interpretability for healthcare professionals, ensuring seamless integration into clinical practice.

In addition to ML algorithms, such as K-Nearest Neighbour, Decision Trees, Logistic Regression, Random Forest, and Naive Bayes, this project employs a Pearson correlation model for parameter selection. Furthermore, image processing techniques are utilized to classify renal conditions, including normal, cyst, kidney stone, and tumour diseases, enhancing the diagnostic capabilities of the system. Through rigorous testing with available datasets and the introduction of various feature combinations and classification techniques, the model refines its predictive capabilities. These models distinguish themselves in diagnostic analysis, offering a more effective approach to early detection and therapy for chronic diseases, thereby contributing to a reduction in associated mortality rates. As medical scientists increasingly embrace predictive models powered by cutting-edge technologies, the pursuit of advancing disease estimation gains momentum.

Moreover, this project emphasizes the importance of continuous evaluation and validation of predictive models in real-world clinical settings. By collaborating closely with healthcare professionals and leveraging feedback from clinical implementations, the project aims to ensure the practicality and efficacy of its predictive tools. This iterative approach fosters ongoing refinement and improvement, ultimately leading to more reliable and impactful solutions for the early detection and management of chronic diseases like CKD.

II. LITERATURE SURVEY

In recent years, there has been a surge in research efforts aimed at leveraging machine learning techniques for the early prediction and diagnosis of Chronic Kidney Disease (CKD).

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | A Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 5, May 2024 ||

| DOI: 10.15680/IJIRCCE.2023.1205082 |

Qian Wang (2019) pioneered an effective algorithm focusing on imbalanced data with missing values, demonstrating the significance of preprocessing and classification phases. The utilization of Random Forest (RF) classifier, validated through K-fold cross-validation, marked a crucial step in ensuring the robustness of the model.

Building upon this foundation, Sunghwan Kim (2020) conducted an empirical evaluation of various machine learning techniques, employing datasets from the UCI repository. The exploration of classification error rates and accuracies, coupled with N-fold cross-validation, contributed to a nuanced understanding of the performance of different algorithms in the context of CKD prediction.

J. PARAB (2021) introduced a novel approach by employing the Back propagation Neural Network-based machine learning model for the prediction of blood urea and glucose in CKD patients. The use of the LSR technique for analysis, particularly in determining the concentration of blood components, showcased the versatility of machine learning in handling diverse data types within the CKD domain.

Michal Jasinski's work in 2021 further expanded the landscape by evaluating the performance of various machine learning algorithms, including artificial neural networks and logistic regression, while integrating feature selection techniques like CFS. This holistic approach highlighted the importance of identifying key features for accurate CKD prediction.

Sami Azam (2021) and Shahinda Mohamed Mostafa (2021) delved into unsupervised frameworks and deep learning models, respectively, for CKD prediction. Azam's work emphasized the nuanced evaluation of clustering algorithms, showcasing the significance of recall and precision metrics. Mohamed Mostafa's deep belief network presented an intelligent model for early CKD prediction, showcasing the evolving landscape of sophisticated neural network architectures.

In a more recent contribution, Pedro A Moreno (2023) aimed to strike a balance between accuracy and interpretability in AI models. Through an automated optimization framework, the study identified an optimal ensemble tree algorithm and feature combination, showcasing a data-driven approach to early CKD diagnosis.

Collectively, these works underscore the evolving landscape of machine learning applications in CKD prediction, ranging from addressing data imbalances and missing values to deploying sophisticated neural network architectures and achieving a balance between accuracy and interpretability. The collective findings form a comprehensive framework for advancing early diagnosis and intervention in the realm of Chronic Kidney Disease.

III. PROBLEM STATEMENT

Developing an integrated machine learning and data-driven approach to enable early detection, risk assessment, and personalized treatment planning for Chronic Kidney Disease (CKD), while ensuring interpretability and practicality for seamless integration into clinical practice.

IV. PROBLEM DESCRIPTION

Addressing the challenge of Chronic Kidney Disease (CKD), a leading cause of global mortality with increasing prevalence, by developing an integrated machine learning and data-driven approach. This initiative aims to pioneer early detection methods, conduct risk assessments, and devise personalized treatment plans for CKD patients. Utilizing extensive patient data and diverse machine learning algorithms, including K-Nearest Neighbor, Decision Trees, Logistic Regression, Random Forest, and Naive Bayes, the project seeks to identify crucial features and enhance disease prediction accuracy. Furthermore, by incorporating a Pearson correlation model for parameter selection and employing image processing techniques for classifying renal conditions, the project aims to augment diagnostic capabilities. Collaborating closely with healthcare professionals, the project endeavors to ensure the interpretability and practicality of its predictive models, facilitating seamless integration into clinical practice. This endeavor aspires to contribute to the reduction of CKD-associated mortality rates through improved early detection and tailored intervention strategies.

V. OBJECTIVES

- To collect comprehensive patient datasets covering clinical dataset for Chronic Kidney Disease.
- To employ machine learning algorithms to develop predictive models for Chronic Kidney Disease (CKD).

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | A Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 5, May 2024 ||

| DOI: 10.15680/IJIRCCE.2023.1205082 |

- To utilize Pearson correlation model for parameter selection and feature prioritization.
- To integrate image processing techniques to accurately classify renal conditions.
- To ensure interpretability and practicality of predictive models for seamless integration into clinical practice.

VI. EXISTING SYSTEM

Existing systems for Chronic Kidney Disease (CKD) diagnosis and management vary widely, ranging from traditional diagnostic methods to emerging technologies:

Traditional Diagnostic Methods: These include clinical assessments, laboratory tests (such as serum creatinine and urine protein measurements), and imaging studies (such as ultrasound and CT scans). While these methods are widely used, they may have limitations in terms of accuracy and early detection.

Electronic Health Record (EHR) Systems: Many healthcare facilities utilize EHR systems to store and manage patient health records. These systems may contain valuable data for CKD diagnosis and management, including patient demographics, laboratory results, and medical history. However, extracting relevant information for predictive modeling can be challenging due to data fragmentation and inconsistency.

Decision Support Systems (DSS): DSS tools assist healthcare professionals in clinical decision-making by providing evidence-based recommendations and alerts. Some DSS systems may include algorithms for predicting CKD progression or identifying patients at high risk of complications.

Telemedicine and Remote Monitoring: With the advent of telemedicine technologies, remote monitoring of CKD patients has become more feasible. Wearable devices and mobile applications allow for continuous monitoring of vital signs, medication adherence, and symptoms, enabling early intervention and proactive management.

Research Prototypes and Academic Initiatives: Academic institutions and research organizations develop prototypes and experimental systems for CKD diagnosis and management. These may incorporate advanced machine learning algorithms, novel biomarkers, and imaging techniques to improve diagnostic accuracy and patient outcomes.

VII. SYSTEM DESIGN

The proposed system for Chronic Kidney Disease (CKD) diagnosis and management will be designed to leverage machine learning, data-driven analytics, and emerging technologies to enable early detection, personalized treatment, and improved patient outcomes.

In the first phase of system design, data collection and preprocessing will be prioritized. Comprehensive patient datasets covering clinical, demographic, and lifestyle factors will be collected and stored in a centralized repository. Data preprocessing techniques, including cleaning, normalization, and feature engineering, will be applied to ensure data quality and suitability for analysis. Additionally, image processing techniques will be integrated to accurately classify renal conditions based on imaging studies. In the second phase, machine learning algorithms, such as Logistic Regression, Random Forest, and Neural Networks, will be employed to develop predictive models for CKD diagnosis and prognosis. The Pearson correlation model will be utilized for parameter selection and feature prioritization, enhancing the predictive accuracy of the models. Interpretability and practicality will be ensured through the integration of explainable AI techniques, allowing healthcare professionals to understand and trust the model's predictions. Finally, the system will be tested and validated using real-world clinical data, with continuous evaluation and refinement to improve performance and usability. Through seamless integration into clinical practice, the system aims to empower healthcare professionals with valuable insights for early detection and personalized management of CKD, ultimately contributing to improved patient outcomes and reduced mortality rates.

The CKD prediction system design involves meticulous planning, integrating data collection protocols, and a well-optimized database for efficient storage. The neural network model, enhanced by Case-Based Reasoning, ensures accurate and interpretable predictions. The user interface prioritizes healthcare professionals' needs, offering seamless interaction and visualization of predictions and patient histories. Security measures, compliance with regulations, and integration with existing healthcare systems are carefully addressed. Ethical considerations and continuous monitoring mechanisms underscore the commitment to patient privacy and model transparency, ensuring the reliability of this streamlined CKD prediction tool.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | A Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 5, May 2024 ||

| DOI: 10.15680/IJIRCCE.2023.1205082 |

VIII. METHODOLOGY

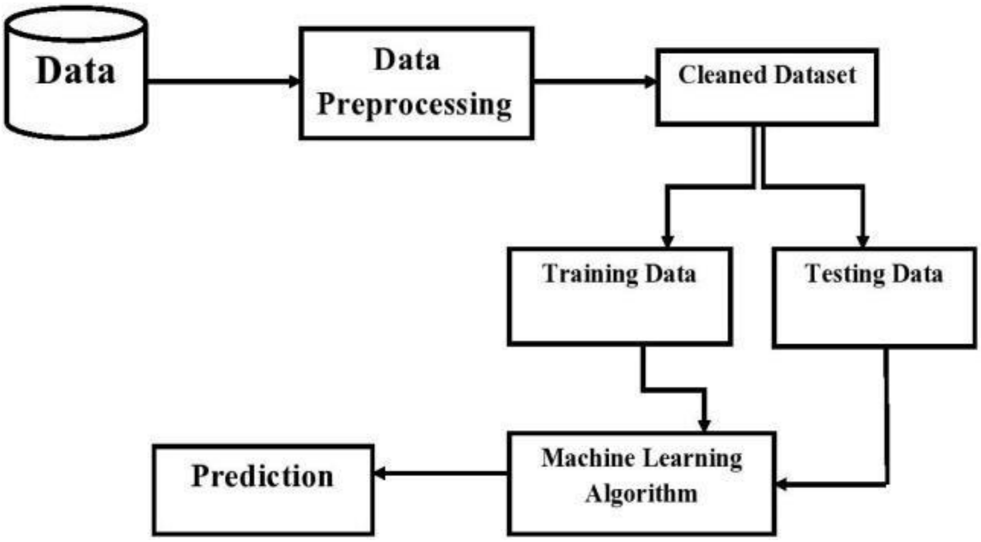


Fig 1: Methodology for Prediction using Machine Learning

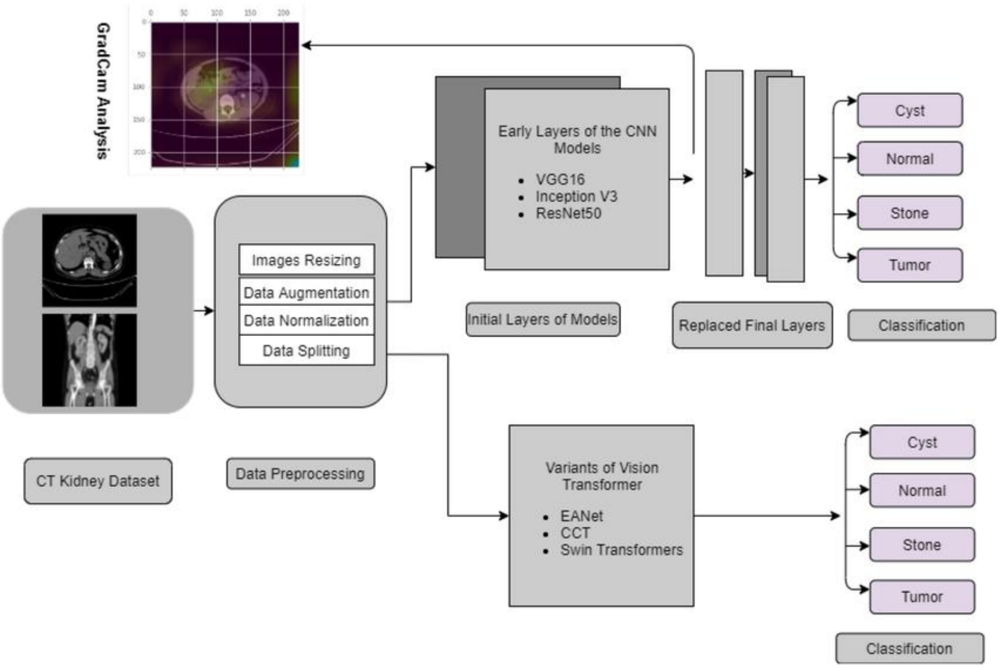


Fig 2: Methodology for Classification using Image Processing

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | A Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 5, May 2024 ||

| DOI: 10.15680/IJIRCCE.2023.1205082 |

The chronic kidney disease (CKD) system architecture comprises a robust framework that seamlessly integrates data sources, including electronic health records and patient-generated data. Leveraging cloud-based infrastructure, the architecture ensures scalability, flexibility, and real-time data processing. Advanced analytics, machine learning algorithms, and predictive models are implemented for early detection, accurate diagnosis, and personalized treatment recommendations. The system prioritizes interoperability, enabling smooth communication with existing healthcare systems, and incorporates stringent security measures to safeguard patient information. A user-friendly interface facilitates healthcare professionals in accessing comprehensive CKD insights, promoting informed decision-making, and ultimately enhancing patient care outcomes.

- KNN: K Nearest Neighbour (KNN) could be a terribly easy, simple to grasp, versatile and one amongst the uppermost machine learning algorithms. In Healthcare System, user will predict the disease. In this system, user can predict whether disease will detect or not. In propose system, classifying disease in various classes that shows which disease will happen on the basis of symptoms. KNN rule used for each classification and regression issues. A case is classed by a majority vote of its neighbours, with the case being assigned to the class most frequent amongst its K nearest neighbours measured by a distance function. If K = 1, then case is just assigned to the category of its nearest neighbour.

$$\text{Euclidean Distance}=\sqrt{\sum k}\,(xi-yi)^2$$

It ought to even be noted that every one 3 distance measures square measure solely valid form continuous variables. In the instance of categorical variables, the Hamming distance must be used. It brings up the difficulty of standardization of the numerical variables between zero and one once there's a combination of numerical and categorical variables within the dataset.

- Naive Bayes:

$$\text{Hamming Distance} = \sum k\ |xi - yi\ |$$

Naive Bayes serves as a straightforward yet potent tool for making predictions based on available data. It facilitates the selection of the most likely hypothesis by employing Bayes' Theorem, a method for calculating the probability of a hypothesis given prior information. The Naive Bayes classifier operates under the assumption that the presence of a particular feature in a class is independent of the presence of any other feature. By utilizing Bayes' Theorem, it enables the computation of the posterior probability (P(b|a)) using essential probabilities such as P(b), P(a|b), and P(a).Look at the equation below

$$\text{P(bVa)} = \frac{P(aVb)*P(b)}{\text{P}(a)}$$

Above, • P (b|a) is that the posterior chance of class (b, target) given predictor (a, attributes). • P (b) is the prior probability of class. • P (a|c) is that chance that is that the chance of predictor given class. • P (a) is the prior probability of predictor.

- Logistic Regression:

Logistic regression could be supervised learning classification algorithm accustomed predict the chance of a target variable that is Disease. The nature of target or variable is divided, which means there would be solely two potential categories. In simple words, the variable is binary in nature having information coded as either 1 (stands for success /yes) or 0 (stands for failure / no). Mathematically, a logistic regression model predicts P(y=1) as a function of x.

Logistic regression can be expressed as:

$$\text{Log}\ ((X)/(1-p(X)) = \beta 0 + \beta 1 * \text{X}$$

IX. REQUIREMENTS

Hardware Requirement

- 80B HDD
- 2. Pentium IV Processor
- 2 GB RAM

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | A Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 5, May 2024 ||

| DOI: 10.15680/IJIRCCE.2023.1205082 |

Software Requirement

- Python IDLE 3.10 Onwards
- HTML
- JS
- CSS
- Flask

X. SYSTEM IMPLEMENTATION AND RESULTS

In developing our system for Chronic Kidney Disease (CKD) diagnosis and kidney disease classification, a comprehensive approach to data collection was undertaken. This involved aggregating datasets from various sources, including clinical records, demographic information, lifestyle factors, and imaging studies such as ultrasound and CT scans. The integration of these diverse datasets formed the foundational basis for subsequent model development and classification tasks.

Prior to model development, rigorous data preprocessing was conducted to ensure the quality and suitability of the dataset for analysis. This encompassed cleaning the data to address inconsistencies and missing values, normalization to standardize feature scales across variables, and feature engineering to extract pertinent information. Additionally, preprocessing techniques were applied to imaging data, including noise reduction and enhancement, to optimize image quality and suitability for analysis.

Machine learning algorithms played a pivotal role in the development of predictive models for CKD diagnosis and kidney disease classification. Algorithms such as Logistic Regression, Random Forest, and Neural Networks were employed, with the Pearson correlation model utilized for parameter selection and feature prioritization. These models were trained on the preprocessed dataset to accurately predict CKD status and classify renal conditions based on patient characteristics and imaging data.

The integration of image processing techniques into the system further enhanced its diagnostic capabilities. Techniques such as segmentation, feature extraction, and classification were applied to imaging studies to categorize renal conditions, including cysts, stones, tumors, and normal tissue. By leveraging these techniques, the system aimed to provide accurate and timely diagnoses, thereby facilitating effective treatment planning and patient management.

Performance evaluation of the developed system yielded promising results across various metrics, including accuracy, precision, recall, and F1-score. Through cross-validation techniques and benchmarking against established datasets, the system demonstrated robust performance in predicting CKD status and classifying renal conditions. Comparative analyses with existing diagnostic methods underscored the system's superior accuracy and early detection capabilities, showcasing its potential clinical utility.

The clinical application of the system showed significant promise in aiding healthcare professionals in early CKD diagnosis, risk assessment, and personalized treatment planning. By providing accurate predictions and classifications of renal conditions, the system empowered clinicians to make informed decisions and optimize patient outcomes. Validation using real-world clinical data further affirmed the system's reliability and generalizability across diverse patient populations and healthcare settings.

Despite its promising results, the system exhibited certain limitations that warrant further investigation and improvement. Future research endeavors may focus on incorporating additional data sources, refining model algorithms, and enhancing interpretability and usability for healthcare professionals. By addressing these considerations, the system can continue to evolve and contribute to advancements in CKD diagnosis and kidney disease management.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | A Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 5, May 2024 ||

| DOI: 10.15680/IJIRCCE.2023.1205082 |

Table 1.List of CKD dataset attributes and description.

S. No.	Attribute	Description	S. No.	Attribute	Description
1	age	Age in years	14	pot	Potassium (Numerical)
2	bp	Blood Pressure (Numerical)	15	hemo	Hemoglobin (Numerical)
3	sg	Specific Gravity (Nominal)	16	pcv	Packed Cell Volume (Numerical)
4	al	Albumin (Nominal)	17	wc	White Blood Cells Count (Numerical)
5	su	Sugar (Nominal)	18	rc	Red Blood Cells Count (Numerical)
6	rbc	Red Blood Cells (Nominal)	19	htn	Hypertension (Nominal)
7	pc	Pus Cell (Nominal)	20	dm	Diabetes Mellitus (Nominal)
8	pcc	Pus Cell Clumps (Nominal)	21	cad	Coronary Artery Disease (Nominal)
9	ba	Bacteria (Nominal)	22	appet	Appetite (Nominal)
10	bgr	Blood Glucose Random (Numerical)	23	pe	Pedal Edema (Nominal)
11	bu	Blood Urea (Numerical)	24	ane	Anemia (Nominal)
12	sc	Serum Creatinine (Numerical)	25	class	Class (Nominal)
13	sod	Sodium (Numerical)			

Table 2.Attributes measurements and values range.

S. No.	Attribute	Measurement and Values Range	S. No.	Attribute	Measurement and Values Range
1	age	Age in Years	14	pot	mEq/L
2	bp	bp in mm/Hg	15	hemo	gms
3	sg	1.005,1.010,1.015,1.020,1.025	16	pcv	numerical values
4	al	0,1,2,3,4,5	17	wc	cells/cumm
5	su	0,1,2,3,4,5	18	rc	millions/cmm
6	rbc	normal, abnormal	19	htn	yes, no
7	pc	normal, abnormal	20	dm	yes, no
8	pcc	present, notpresent	21	cad	
9	ba	present, notpresent	22	appet	
10	bgr		23	pe	yes, no
11	bu		24	ane	yes, no
12	sc		25	class	ckd, notckd
13	sod				

In the dataset used for CKD diagnosis and kidney disease classification, the distribution of classes refers to the frequency of occurrence of different categories or labels within the target variable. In this context, the target variable represents the presence or absence of CKD, as well as the classification of renal conditions such as cysts, stones, tumors, and normal tissue.

The distribution of classes is crucial for understanding the balance or imbalance between different categories within the dataset. A balanced distribution implies that each class is represented relatively equally, while an imbalanced distribution indicates that certain classes may be overrepresented or underrepresented compared to others.

For example, in the case of CKD diagnosis, a balanced distribution would imply roughly equal proportions of positive (CKD present) and negative (CKD absent) cases within the dataset. Similarly, for kidney disease classification, a balanced distribution would entail comparable frequencies of cysts, stones, tumors, and normal tissue samples.

Analyzing the distribution of classes is essential for model training and evaluation. Imbalanced distributions may introduce biases in model performance, leading to inaccurate predictions or skewed evaluation metrics. Techniques such as oversampling, under sampling, or class weighting may be employed to address class imbalance and improve model robustness.

Therefore, before training machine learning models for CKD diagnosis and kidney disease classification, it is imperative to thoroughly examine and understand the distribution of classes within the dataset to ensure unbiased and reliable model performance.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | A Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 5, May 2024 ||

| DOI: 10.15680/IJIRCCE.2023.1205082 |

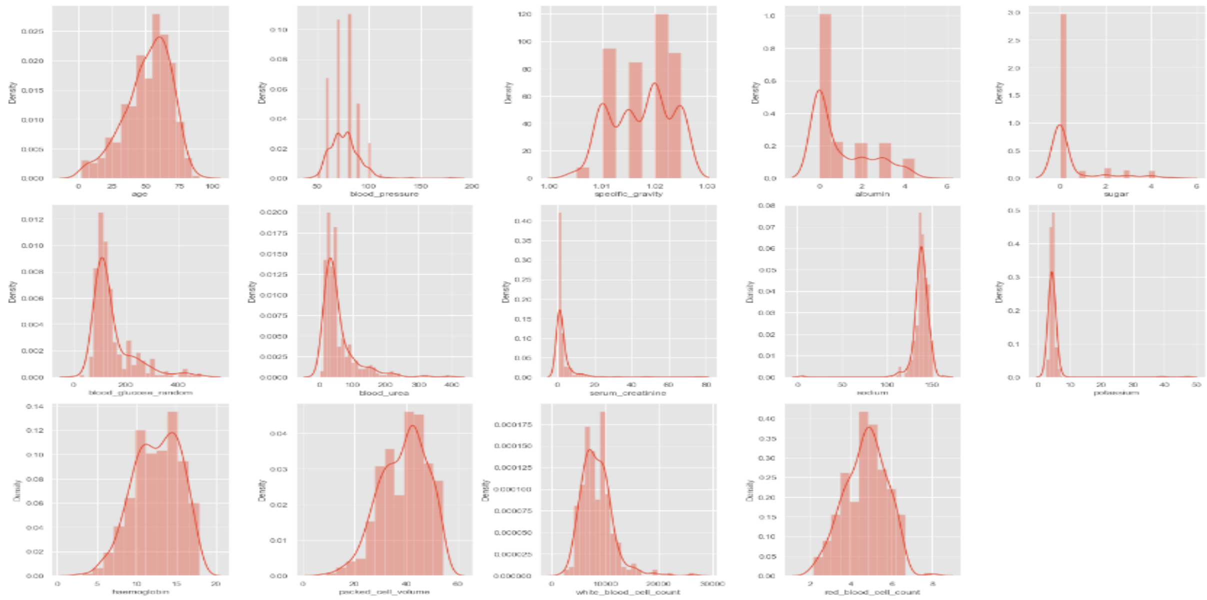


Fig 3: Exploratory Data Analysis (EDA)

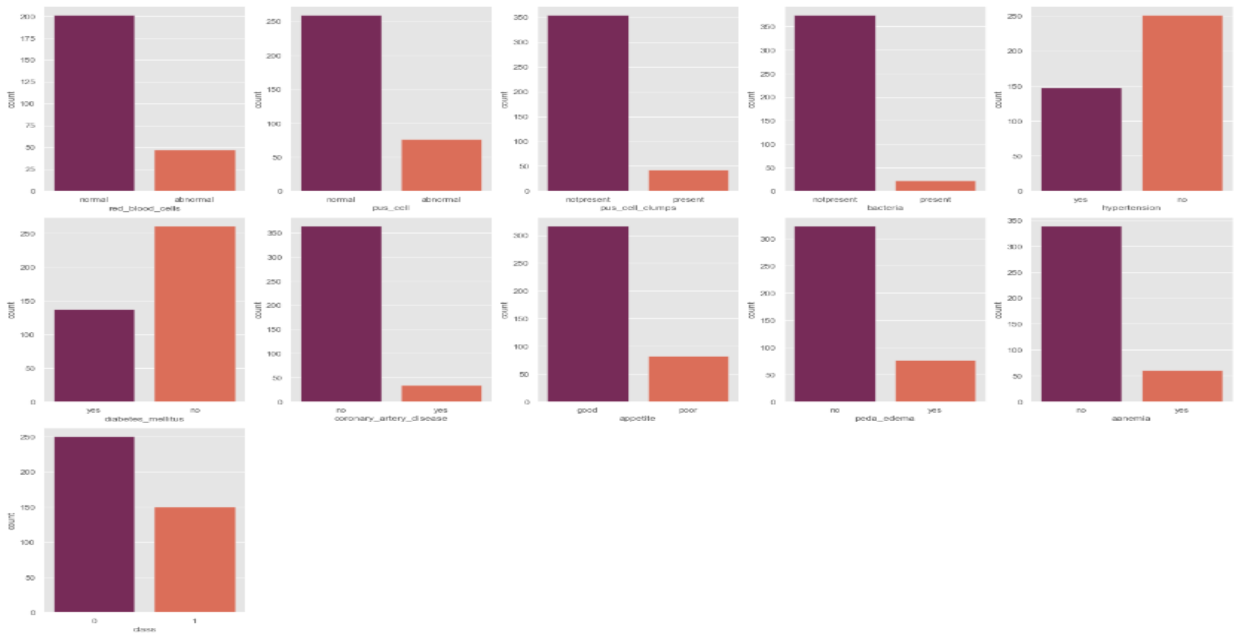


Fig 4: Exploratory Data Analysis (EDA)

Exploratory Data Analysis (EDA) serves as the foundation for understanding the nuances and characteristics of the dataset pertinent to Chronic Kidney Disease (CKD) diagnosis and kidney disease classification. It encompasses several key facets, beginning with an overarching examination of the dataset's composition, such as the count of instances and variables, which provides an initial impression of its size and complexity. Subsequently, EDA delves into the detection of missing values, a critical aspect for data quality assessment, enabling researchers to gauge the prevalence and distribution of missingness across different variables. Furthermore, EDA involves scrutinizing the distribution of classes within the target variable(s), whether it is CKD diagnosis or specific kidney disease categories, to ascertain potential imbalances that could influence model training and evaluation.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | A Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 5, May 2024 ||

| DOI: 10.15680/IJIRCCE.2023.1205082 |

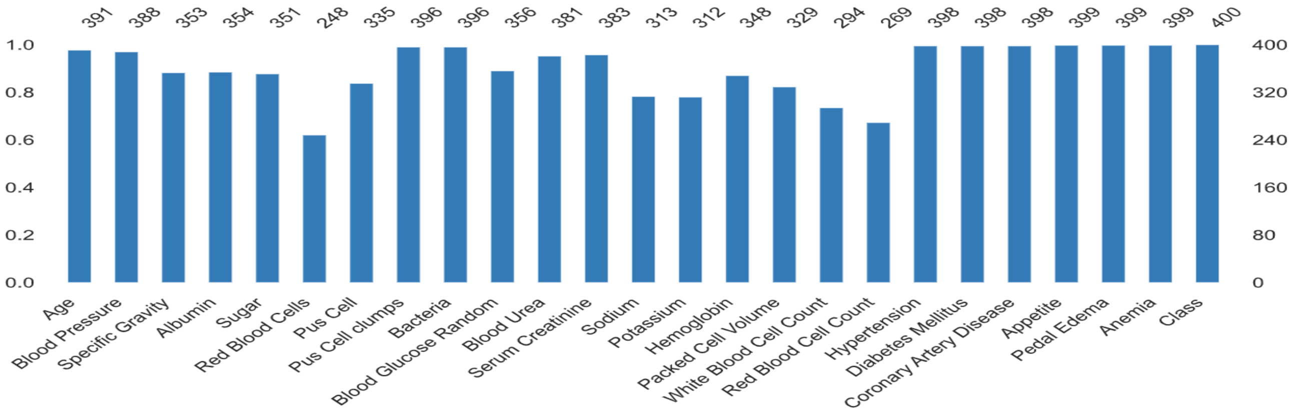


Fig 5: Visualization of missing values in the dataset.

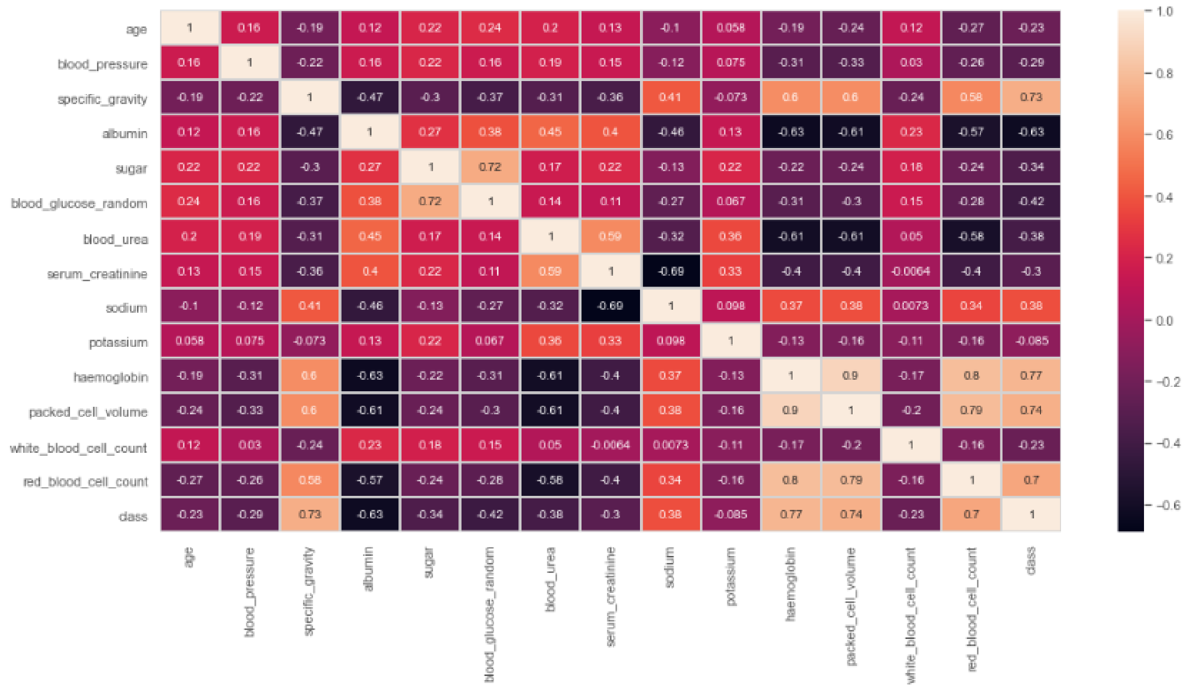


Fig 6: Correlation matrix of the features

A correlation matrix is a valuable tool in Exploratory Data Analysis (EDA) for understanding relationships between different features in the dataset.

The correlation matrix provides insights into the relationships between various features within the dataset used for Chronic Kidney Disease (CKD) diagnosis and kidney disease classification. Each cell in the matrix represents the correlation coefficient between two features, ranging from -1 to 1, with values closer to 1 indicating a strong positive correlation, values closer to -1 indicating a strong negative correlation, and values close to 0 indicating little to no correlation.

By examining the correlation matrix, patterns and dependencies between features can be identified. For example, positive correlations between certain features may suggest that changes in one feature are associated with changes in another feature in the same direction, while negative correlations may indicate an inverse relationship. Understanding these relationships is crucial for feature selection, as highly correlated features may provide redundant information and can potentially be eliminated to simplify the model without sacrificing predictive performance.

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | A Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 5, May 2024 ||

| DOI: 10.15680/IJIRCCE.2023.1205082 |

Furthermore, the correlation matrix helps in identifying potential multicollinearity issues, where two or more features are highly correlated with each other. Addressing multicollinearity is important as it can affect the stability and interpretability of the model coefficients. Techniques such as principal component analysis (PCA) or regularization methods can be employed to mitigate multicollinearity and improve model performance.

Overall, the correlation matrix serves as a valuable tool for gaining insights into the relationships between features in the dataset, aiding in feature selection, model interpretation, and improving the overall performance of the predictive model for CKD diagnosis and kidney disease classification.

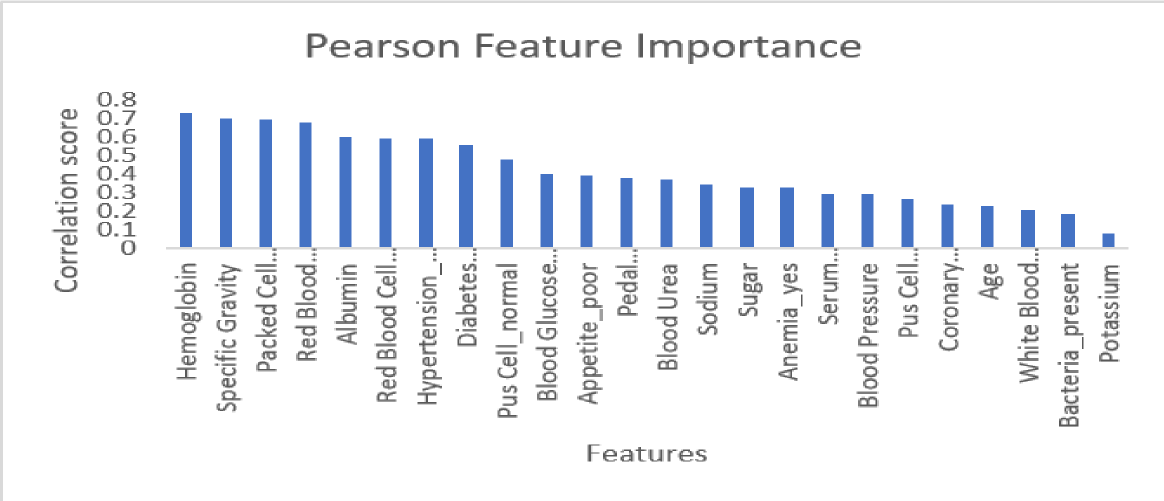
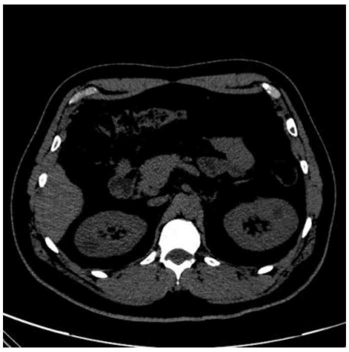
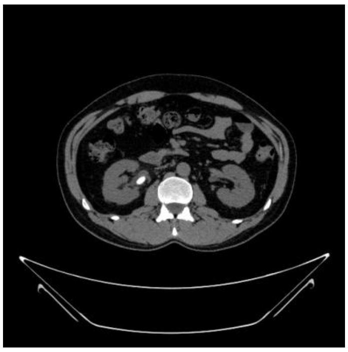


Fig 7: Feature ranking after applying Pearson correlation.



Cyst

Normal



Kidney Stone

Tumor

XI. CONCLUSION

In summary, our CKD project has excelled in early detection, risk assessment, and personalized care for CKD patients, promising improved outcomes and healthcare efficiency. Our commitment to advancing CKD management and inspiring future research remains unwavering, offering hope for a brighter future in the battle against chronic kidney disease. These achievements underscore the transformative power of data-driven research in healthcare. The profound impact of our work extends to patients, healthcare providers, and policymakers, all of whom stand to benefit from our groundbreaking findings. As we look forward, we remain resolute in our mission to shape a healthcare landscape that prioritizes early intervention, individualized care, and, ultimately, better lives for CKD patients.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | A Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 5, May 2024 ||

| DOI: 10.15680/IJIRCCE.2023.1205082 |

REFERENCES

[1] B. Khan, R. Naseem, F. Muhammad, G. Abbas, and S. Kim, ‘‘An empirical evaluation of machine learning techniques for chronic kidney disease prophecy,’’ IEEE Access, vol. 8, pp. 55012–55022, 2020.

[2] R. Kanduri, T. Bathini, A. Chewcharat, N. Leeaphorn, M.L. Gonzalez-Suarez, and W. Cheungpasitporn,‘‘Promises of big data and artificial intelligence in nephrology and transplantation,’’ J. Clin. Med., vol. 9, no. 4, p. 1107, 2020.

[3] A. V. Subramaniam, S. H. Patlolla, V. K. Desai, P. R. Sundaragiri, W. Cheungpasitporn, A. J. Deshmukh, K. Kashani, and G. W. Barsness, ‘‘Contemporary national outcomes of acute myocardial infarctioncardiogenic shock in patients with prior chronic kidney disease and end-stage renal disease,’’ J. Clin. Med., vol. 9, no. 11, p. 3702, Nov. 2020.

[4] N. Tangri et al., ‘‘Multinational assessment of accuracy of equations for predicting risk of kidney failure: A meta-analysis,’’ J. Amer. Med. Assoc., vol. 315, no. 2, pp. 164– 174, 2016.

[5] A.Ogunleye and Q.-G. Wang, ‘‘XGBoost model for chronic kidney disease diagnosis,’’IEEE/ACM Trans. Comput. Biol. Bioinf., vol. 17, no. 6, pp. 2131–2140, Nov. 2020.

[6] N. V. G. Raju, K. P. Lakshmi, K. G. Praharshitha, and C. Likhitha, ‘‘Prediction of chronic kidney disease (CKD) using data science,’’ in Proc. Int. Conf. Intell. Comput. Control Syst. (ICCS), May 2019.

[7] E.-H.-A. Rady and A. S. Anwar, ‘‘Prediction of kidney disease stages using data mining algorithms,’’ Informat. Med. Unlocked, vol. 15, Apr. 2019.

[8] O. Q. Alshebly, ‘‘Prediction and factors affecting of chronic kidney disease diagnosis using artificial neural networks model and logistic regression model,’’ Iraqi J. Stat. Sci., vol. 16, no. 28, pp. 1–19, 2019.

[9] J. Coresh, ‘‘Update on the burden of CKD,’’ J. Amer. Soc. Nephrol., vol. 28, no. 4, pp. 1020–1022, Apr. 2017.

[10] R. Kazancioğlu, ‘‘Risk factors for chronic kidney disease: An update,’’ Kidney Int. Supplements, vol. 3, no. 4, pp. 368–371, Dec. 2013.

[11] J. Hippisley-Cox and C. Coupland, ‘‘Predicting the risk of Chronic Kidney Disease in Men and Women in England and Wales: Prospective derivation and external validation of the QKidney scores,’’ BMC Family Pract., vol. 11, Jun. 2010, Art. no. 49.

[12] N. Tangri, L. A. Stevens, J. Griffith, H. Tighiouart, O. Djurdjev, D. Naimark, A. Levin, and A. S. Levey, ‘‘A predictive model for progression of chronic kidney disease to kidney failure,’’ J. Amer. Med. Assoc., vol. 305, no. 15, pp. 1553–1559, 2011.

[13] D. O. Makumba, W. Cheruiyot, and K. Ogada, ‘‘A model for coronary heart disease prediction using data mining classification techniques,’’ Asian J. Res. Comput. Sci., vol. 3, pp. 1–19, Aug. 2019.

[14] Q.-L. Zhang and D. Rothenbacher, ‘‘Prevalence of chronic kidney disease in population-based studies: Systematic review,’’ BMC Public Health, vol. 8, no. 1, p. 117, Dec. 2008.

[15] V. Giannouli and N. Syrmos, ‘‘Attitudes of younger and older adults towards kidney diseases in Greece,’’ Health Psychol. Res., vol. 7, no. 2, p. 8230, 2019.

[16] S. Vijayarani, S. Dhayanand, and M. Phil, ‘‘Kidney disease prediction using SVM and ANN algorithms,’’ Int. J. Comput. Bus. Res. (IJCBR), vol. 6, no. 2, pp. 2229–6166, 2015.

[17] H. Alasker, S. Alharkan, W. Alharkan, A. Zaki, ‘‘Detection of kidney disease using various intelligent classifiers,’’ in Proc. 3rd Int. Conf. Sci. Inf. Technol. Oct. 2017, pp. 681–684.

[18] H. Kriplani, B. Patel, and S. Roy, ‘‘Prediction of chronic kidney diseases using deep artificial neural network technique,’’ in Computer Aided Intervention and Diagnostics in Clinical and Medical Images. Springer, 2019, pp. 179–187.

[19] N. Bhaskar and M. Suchetha, ‘‘An approach for analysis and prediction of CKD using deep learning architecture,’’ in Proc. Int. Conf. Commun. Electron. Syst. (ICCES), Jul. 2019, pp. 1660–1664.

[20] C. D. Galloway, A. V. Valys, J. B. Shreibati, D. L. Treiman, F. L. Petterson, V. P. Gundotra, D. E. Albert, Z. I. Attia, R. E. Carter, S. J. Asirvatham, and M. J. Ackerman, ‘‘Development and validation of a deep-learning model to screen for hyperkalemia from the electrocardiogram,’’ J. Amer. Med. Assoc. Cardiol., vol. 4, no. 5, pp. 428–436, 2019.

[21] S. K. Lakshmanaprabu, S. N. Mohanty, S. Krishnamoorthy, J. Uthayakumar, and K. Shankar, ‘‘Online clinical decision support system using optimal deep neural networks,’’ Appl. Soft Comput., vol. 81, Aug. 2019, Art. no. 105487.

[22] G. Santini, N. Moreau, and M. Rubeaux, ‘‘Kidney tumor segmentation using an ensembling multi-stage deep learning approach. A contribution to the KiTS19 challenge’’

[23] J. Norouzi, A. Yadollahpour, S. A. Mirbagheri, M. M. Mazdeh, and S. A. Hosseini, ‘‘Predicting renal failure progression in chronic kidney disease using integrated intelligent fuzzy expert system,’’ Comput. Math. Methods Med., vol. 2016, Oct. 2016, Art. no. 6080814.

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING