SEMINAR REPORT

ON

"CHRONIC KIDNEY DISEASE DETECTION AND GUIDANCE"

SUBMITTED BY

Vedant Dindore

Anurag Ganvir

Varad Gondkar

Under the Guidance of

Dr. Prof. Mrs. D. P. Adhyapak

DEPARTMENT OF ELECTRONICS & TELECOMMUNICATION

(Program: ELECTRONICS AND COMPUTER ENGINEERING)

P.E.S'S MODERN COLLEGE OF ENGINEERING

PUNE – 411 005.

SAVITRIBAI PHULE PUNE UNIVERSITY

ACADEMIC YEAR: 2024-2025

CERTIFICATE

This is to certify that

Vedant Dindore	**Exam No: B1903105014**
Anurag Ganvir	**Exam No: B1903105017**
Varad Gondkar	**Exam No: B1903105018**

of B.E. (E&TC) have successfully completed the seminar titled '**Chronic Kidney Disease Detection And Guidance**' during the academic year 2024-25. This report is submitted as partial fulfillment of the requirement of degree in E&TC Engineering as prescribed by Savitribai Phule Pune University.

Dr. Prof. Mrs. R. S. Kamathe	**Dr. Prof. Mrs. D. P. Adhyapak**
Head Of Department	**Project Co-ordinator**

Place: Pune

Date: / /

ACKNOWLEDGEMENT

We would like to take this opportunity to express our heartfelt gratitude to our project guide, Dr. Prof. Mrs. D. P. Adhyapak. Her exceptional guidance and unwavering support throughout the duration of this project have been invaluable. Working under her mentorship has not only enriched our understanding but also made the research process an engaging and enlightening experience. We are deeply appreciative of her insightfulsuggestions, constructive feedback, and the numerous discussions that have greatly enhanced our work.

We also extend our sincere thanks to Dr. Prof. Mrs.R. S. Kamathe, the Head of the Department of Electronics & Computer Engineering, for her valuable guidance and support. Her encouragement and the facilities provided during our seminar preparations have played a crucial role in our project's development. The resources and insights she offered were instrumental in shaping our approach.

Additionally, we are profoundly grateful to Dr. Prof. Mrs. K. R. Joshi, Principal of PES's Modern College of Engineering. Her commitment to creating a conducive research environment has greatly inspired us. Her belief in our potential and her kind encouragement have motivated us to strive for excellence.

Lastly, we thank all those who have contributed, both directly and indirectly, to the successful completion of this project. Each contribution, no matter how small, has made a significant difference, and we are grateful fortheir involvement in this journey.

With Deep Reverence.

Vedant Dindore
Anurag Ganvir
Varad Gondkar

ABSTRACT

Chronic Kidney Disease (CKD) is a progressive condition that often remains undetected until advanced stages, leading to severe complications such as kidney failure. Traditional detection methods like blood and urine tests are inadequate for early diagnosis, making early intervention challenging. This project addresses the problem by utilizing neuralnetworks for early CKD prediction, leveraging their ability to capture complex, non-linear relationships in patient data.

The novelty of this approach lies in its automatic feature extraction and adaptability, allowing the model to improve over time and scale with new data. Key results demonstrate that neural networks can accurately predict CKD risk in earlier stages compared to traditional methods. Additionally, these models have been successfully applied in clinical decision support systems, personalized treatment plans, and telehealth platforms, enhancing early detection and patient care.

In conclusion, this project highlights the potential of AI, particularly neural networks, to revolutionize CKD detection by offering more accurate, scalable, and adaptive solutions, leading to better patient outcomes and healthcare efficiency.

LIST OF FIGURES

Table of Contents

1. Introduction

Chronic Kidney Disease (CKD) is a long-term medical condition characterized by the gradual loss of kidney function, where the kidneys fail to effectively filter waste products and excess fluids from the blood. This leads to the buildup of harmful substances in the body, eventually resulting in severe health complications such as kidney failure, cardiovascular disease, and anemia. CKD is a global public health concern, affecting millions of individuals, and its prevalence continues to rise due to risk factors such as diabetes and hypertension.

One of the major challenges with CKD is its "silent" nature in the early stages. Most individuals do not experience noticeable symptoms until the disease has advanced, by which time irreversible damage to the kidneys may have occurred. Common symptoms that manifest in later stages include fatigue, swelling in the legs and ankles, changes in urination, and difficulty concentrating. Early detection of CKD is crucial for slowing disease progression, managing symptoms, and preventing the need for costly and invasive treatments such as dialysis or kidney transplantation.

Traditional methods for CKD detection, such as blood tests to measure creatinine and the glomerular filtration rate (GFR), urine tests for protein levels, and kidney biopsies, are limited in their ability to identify the disease in its early stages. Blood and urine tests often fail to detect subtle changes in kidney function until the disease is well advanced, and kidney biopsies, while more accurate, are invasive and not commonly performed as routine screening tools.

Given these limitations, there is a growing need for more advanced and accurate methods of early CKD detection. In this context, machine learning (ML) has emerged as a promising solution, with neural networks in particular showing great potential. Neural networks excel in handling large datasets and detecting non-linear relationships between variables, making them well-suited for medical applications such as CKD prediction. These models can automatically learn from patient data, identify early indicators of kidney damage, and provide more precise risk assessments than traditional methods.

This report explores the integration of neural networks in CKD prediction, focusing on their ability to enhance early detection, improve patient outcomes, and reduce healthcare costs. By utilizing real-world clinical data and continuously adapting to new information, neural networks offer a scalable and flexible approach to CKD management. Additionally, the report examines practical applications of these models, including their use in clinical decision support systems, personalized treatment plans, telehealth platforms, and screening tools for pharmacies and insurance companies.

2. LITERATURE SURVEY

Imesh Udara Ekanayake et al. proposed a machine learning-based system for predicting Chronic Kidney Disease (CKD), focusing on early detection through clinical data. They utilized Random Forest and Extra Trees classifiers, achieving high accuracy with effective handling of missing values using the K-nearest neighbor imputer algorithm. Although the model is highly accurate, challenges exist in dealing with missing data and ensuring that the system generalizes well across different patient datasets [1]

Nikhila et al. presented a machine learning-based system for predicting Chronic Kidney Disease (CKD) usingensemble algorithms like AdaBoost, Random Forest, Gradient Boosting, and Bagging. The study achieved 100% accuracy with AdaBoost and Random Forest while evaluating the models using performance metrics such as accuracy, sensitivity, precision, and MCC. However, Bagging and Gradient Boosting showed slightly lower performance, with challenges in handling imbalanced datasets. The paper emphasizes the potential of ensemble learning in providing early diagnosis of CKD [2]

Pankaj Chittora et al. explores the application of machine learning models for predicting Chronic Kidney Disease (CKD). Using the UCI CKD dataset, the authors apply seven machine learning classifiers (e.g., artificial neural network, C5.0, logistic regression, linear support vector machine, and random tree) and evaluate their performance. They also experiment with feature selection methods, such as Correlation-based Feature Selection (CFS), Wrapper, and LASSO, to identify the most relevant features for CKD prediction. Additionally, Synthetic Minority Oversampling Technique (SMOTE) is used to balance the dataset, enhancing model accuracy. Among the models, linear support vector machine with L2 penalty achieves the highest accuracy of 98.86% when used with SMOTE and all features. However, a deep neural network, trained with the same dataset, reaches an even higher accuracy of 99.6%, demonstrating the effectiveness of deep learning for CKD prediction. The authors conclude that optimized feature selection combined with SMOTE improves model performance and provides a robust solution for early CKD detection, particularly with deep learning approaches. [3]

This study by *Asif Salekin* and *Stankovic* explores machine learning methods for early detection of Chronic Kidney Disease (CKD). Using a dataset of 400 individuals (250 with CKD), the authors examine 24 predictive attributes and evaluate three machine learning classifiers: k-nearest neighbors, random forest, and neural networks. The random forest classifier achieves the highest detection accuracy with an F1 score of 0.993 and a root mean square error (RMSE) of 0.1084, significantly outperforming traditional methods like the CKD-EPI equation for estimating glomerular filtration rate (GFR). The study also identifies a reduced subset of 12 key attributes (e.g., hemoglobin, albumin, specific gravity) that maintains high accuracy, optimizing cost and effectiveness. This approach proposes a more accessible, cost-effective machine learning-based CKD detection method, potentially beneficial for early clinical diagnosis and management. This paper supports the potential of machine learning models in enhancing CKD detection accuracy and highlights novel predictive attributes beyond traditional GFR estimation formulas. [4]

Rajeshwari and H.K. Yogish presented a machine learning-based system for predicting Chronic Kidney Disease (CKD) to assist doctors in early diagnosis. The study evaluated four machine learning techniques: Naïve Bayes, Random Forest, Decision Tree, and Support Vector Machine. The Random Forest model outperformed other classifiers with an accuracy of 98.75%, F1-score of 99%, precision of 99%, and recall of 99%. This approach highlights the effectiveness of using machine learning models, especially Random Forest, in improving the accuracy and efficiency of CKD detection. [5]

Chilakamarthi Prem Kashyap, Gollapudi Sai Dakar Reddy, and M. Balamurugan investigated the use of machine learning techniques for the timely diagnosis of Chronic Kidney Disease (CKD), a prevalent condition characterized by the progressive decline in kidney function. The study employed four machine learning algorithms: Support Vector Machine (SVM), K-Nearest Neighbor (KNN), Random Forest, and Decision Tree, using a dataset from the UCI repository. Data preprocessing techniques were applied to enhance the accuracy of CKD predictions. The research highlights the increasing incidence of CKD, drivenprimarily by diabetes and hypertension, and emphasizes the need for effective diagnostic methodologies in managing this critical public health issue. [6]

Zhang and Rothenbacher reviewed the global prevalence of Chronic Kidney Disease (CKD), noting its rise as a major health issue, now primarily caused by hypertension and diabetes. Using the K/DOQI guidelines, they compared GFR estimation methods (Cockcroft-Gault and MDRD equations) and found a CKD prevalence of 7.2% in people aged 30 and older, with higher rates in older populations. The study emphasized the need for early detection to prevent progression to end-stage renal disease (ESRD), which presents a significant financial burden, especially in developing countries. They also highlighted CKD's link to cardiovascular disease. [7]

Wesley de Souza (2019) highlights the concerning trends in chronic kidney disease (CKD) hospitalizations and mortality rates in Espírito Santo, Brazil, from 1996 to 2017. It found a stationary overall mortality rate but noted a significant increase in mortality among women since 2005, alongside a yearly rise in hospitalization rates. The findings underscore the need for targeted public health measures and improved management strategies to address the rising burden of CKD, particularly in vulnerable populations. [8]

Vijendra Singh, Vijayan Asari and Rajkumar Rajasekaran et al. proposes a deep neural network (DNN) for early detection of chronic kidney disease (CKD), designed to enhance diagnostic accuracy and efficiency over traditional methods. Leveraging the UCI CKD dataset, the study emphasizes data preprocessing steps, including mean and mode imputation for handling missing values, normalization, and categorical encoding. Feature selection is conducted via Recursive Feature Elimination (RFE), identifying key indicators such as hemoglobin, serum creatinine, specific gravity, and hypertension as critical in predicting CKD. The DNN model, which includes five dense and dropout layers optimized using the Adam optimizer, achieves a 100% accuracy rate, significantly outperforming existing classifiers like K-Nearest Neighbor (KNN), Support Vector Machine (SVM), and Decision Tree. This model's success underscores its potential as a reliable tool for early CKD diagnosis, offering nephrologists an advanced approach to enhance patient outcomes. Limitations include reliance on limited attributes, which may reduce generalizability, and the need for further validation with larger datasets.[9]

Hamida Ilyas et al. explore the application of decision tree algorithms, J48 and Random Forest, for diagnosing Chronic Kidney Disease (CKD) stages, focusing on early detection to improve patient outcomes. Using the CKD-EPI equation to calculate Glomerular Filtration Rate (GFR)—a key indicator of kidney health—the researchers classify CKD stages on a dataset of 400 records from the UCI repository, which includes attributes like age, serum creatinine, sex, and race. Data preprocessing methods, such as normalization and imputation, enhance model accuracy. The algorithms were tested using WEKA, with J48 achieving 85.5% accuracy in 0.03 seconds, significantly outperforming Random Forest's 78.25% in 0.28 seconds. J48's ability to handle both categorical and numerical data proved advantageous, delivering higher sensitivity, specificity, and efficiency across all CKD stages. This study positions the J48 algorithm as a robust tool for automated CKD diagnosis, contributing valuable insights into machine learning's role in medical decision support systems. The study is limited by a small dataset, restricted features, lack of external validation, and evaluation of only two algorithms without exploring other potential models.[10]

Jiongming qin et al. comprehensive machine learning methodology for chronic kidney disease (CKD) detection, focusing on multiple classification algorithms to improve diagnostic accuracy. The methodology includes the use of decision trees, support vector machines (SVM), neural networks, and ensemble methods to determine the most effective approach for CKD prediction. Key steps in the methodology involve data preprocessing (such as handling missing values and feature scaling), feature selection to identify the most relevant medical indicators (e.g., serum creatinine levels and glomerular filtration rate), and model tuning for optimizing prediction accuracy. Each model's performance is evaluated using metrics like accuracy, sensitivity, specificity, and the area under the curve (AUC) to assess its suitability for clinical application. For the literature survey, this methodology demonstrates a structured approach to using machine learning for medical diagnosis, showcasing how diverse algorithms can support early detection of CKD. However, the study could improve by addressing gaps such as enhancing model explainability, validating results in real-world clinical settings, and exploring novel biomarkers to strengthen predictive power.[11]

*Namyong Park, Eunjeong Kang, Minsu Park, Hajeong Lee, Hee-Gyung Kang, Hyung-Jin Yoon, and U. Kang*The paper on "Predicting Acute Kidney Injury in Cancer Patients Using Heterogeneous and Irregular Data" explores the development of a machine learning model designed to predict acute kidney injury (AKI) in cancer patients. By utilizing irregular and heterogeneous clinical data, including serum creatinine (SCr) measurements, the model achieved strong performance metrics, with a precision of 0.7892 and an F-measure of 0.7576 in predicting AKI within 14 days. This approach shows potential for improving clinical decision-making and preventive care for cancer patients at risk of AKI. [12]

Dibaba Adeba Debal and Tilahun Melak Sitote focuses on predicting Chronic Kidney Disease (CKD) using advanced machine learning techniques. The study employs models such as Random Forest (RF), Support Vector Machine (SVM), and Decision Tree (DT) to classify CKD stages. Through feature selection methods like Recursive Feature Elimination with Cross-Validation (RFECV), the key predictive attributes were identified. Among the classifiers, Random Forest demonstrated the best performance in predicting both binary and multiclass CKD stages, highlighting its potential to enhance early detection and intervention for CKD patients. [13]

Abdulhamit Subasi, Emina Alickovic, and Jasmin Kevric in their study, Diagnosis of Chronic Kidney Disease by Using Random Forest, investigate chronic kidney disease (CKD) detection using various machine learning classifiers, particularly focusing on the effectiveness of the Random Forest (RF) algorithm. Given CKD's global impact, the authors emphasize the importance of developing accurate, automated diagnostic tools. They evaluate multiple machine learning algorithms, including Artificial Neural Networks (ANN), Support Vector Machines (SVM), k-Nearest Neighbors (k-NN), C4.5 Decision Trees, and RF, using a dataset from the UCI Machine Learning Repository. Among these, RF emerges as the most effective, delivering robust performance in CKD classification due to its ability to handle data variability and complexity. The study highlights how machine learning models, especially RF, can play a critical role in translating CKD symptoms into diagnostic categories, offering potential for use in healthcare settings. The study is constrained by its reliance on a single dataset, which limits its generalizability to broader, more diverse populations. Additionally, the research does not explore hybrid or ensemble methods that might further enhance predictive accuracy.[14]

Abdulhamit Subasi, Emina Alickovic, and Jasmin Kevric in Chronic Kidney Disease Prediction and Recommendation of Suitable Diet Plan Using ML present a methodology that integrates machine learning (ML) techniques for CKD prediction along with diet recommendations tailored to patient needs. Their approach focuses on using supervised learning algorithms, particularly Random Forest (RF), Support Vector Machine (SVM), and Artificial Neural Networks (ANN), to analyze patient data and classify CKD stages. The study utilizes data from the UCI CKD dataset and performs pre-processing steps, including data cleaning, feature selection, and model training, to enhance prediction accuracy. Additionally, it proposes a dietary recommendation system based on patient-specific data, leveraging classification outcomes to suggest diet plans that may help manage CKD progression. This paper highlights the importance of integrating predictive analytics with personalized health recommendations for comprehensive CKD management. The primary limitations include the lack of validation in diverse patient populations and real-world clinical settings, which limits the study's generalizability.[15]

Asst. Prof. Mrs. D. Navya Narayana Kumari, T. Praveen Satya, B. Manikanta, A. Phani Chandana, and Y. L. S. Aditya, in their work Personalized Diet Recommendation System Using Machine Learning, propose a machine learning-based model to deliver tailored diet recommendations focusing on individual health needs. This model incorporates user-specific factors such as age, gender, activity level, and weight goals to generate diet plans that align with nutritional guidelines. Leveraging the Nearest Neighbors algorithm with cosine similarity, the system compares the nutritional profiles of recipes, optimizing meal suggestions based on user preferences and nutritional requirements. The model’s design integrates multiple components: data acquisition from Food.com, feature selection for essential dietary information, and a user-friendly interface developed in Streamlit to provide interactive and personalized meal plans. The recommendation system estimates daily nutritional intake, presenting it through visual tools like pie charts for clarity, and offers users ingredient details and step-by-step meal preparation instructions. The study’s limitations include the use of a single dataset, which may restrict the generalizability of the recommendations, and limited adaptability for users with specific medical conditions requiring customized dietary considerations.[16]

Anonnya Banerjee, Ala Noor, Nasrin Siddiqua et al, In Food Recommendation using Machine Learning for CKD Patients, the authors present a machine learning-based dietary recommendation system tailored for chronic kidney disease (CKD) patients. This system integrates patient-specific data such as age, CKD stage, and dietary restrictions to recommend suitable foods that align with CKD management requirements. The recommendation model uses supervised learning algorithms, with a focus on algorithms like Decision Trees and Random Forests, which effectively classify dietary items based on nutritional profiles that are beneficial or harmful to CKD patients. The dataset used includes CKD-relevant nutritional details, and the model emphasizes the importance of avoiding high potassium and phosphorus foods, which are detrimental to CKD patients. The system also incorporates a user interface that allows patients to input preferences, enhancing personalization in dietary recommendations. The study's primary limitation is its reliance on a limited dataset, which may affect its generalizability across diverse CKD patient populations. Additionally, the model could benefit from real-world clinical validation to ensure effectiveness and practical applicability in healthcare settings[17]

3. SPECIFICATIONS OF THE PROJECT

- **Data Collection:** Gather data from various sources, including:
  - Demographic Features: Age, gender, race, ethnicity, socioeconomic status.
  - Clinical Features: Blood pressure, blood sugar, cholesterol levels, proteinuria, hematuria.

- **Data Preprocessing:** Clean and preprocess the data to ensure quality:
  - Handle missing values.
  - Normalize or standardize numerical features.
  - Encode categorical variables.

- **Feature Selection:** Identify relevant features for model training:
  - Utilize both existing features and potential new features (lifestyle and genetic).
  - Consider techniques for automatic feature extraction using neural networks.

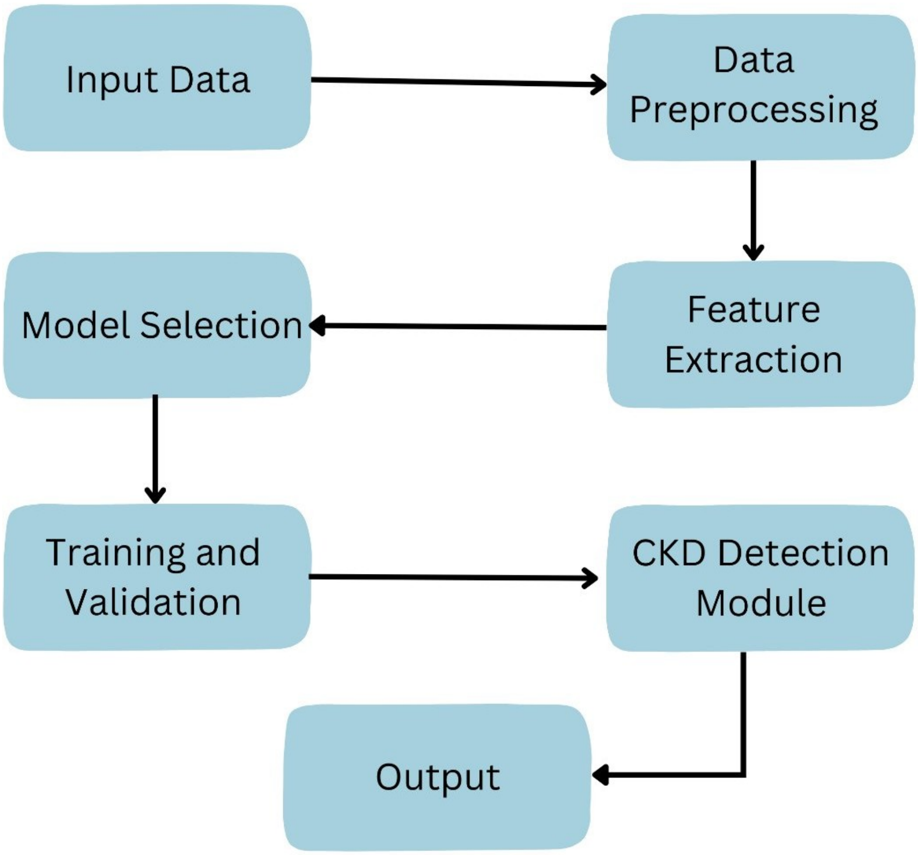
- **Model Development:** Choose and implement machine learning algorithms:
  - Focus on neural networks to capture complex, non-linear relationships.
  - Train the model using the preprocessed dataset.

- **Model Evaluation:** Assess the model's performance using metrics such as:
  - Accuracy
  - Precision
  - Recall
  - F1-score
  - ROC-AUC curve analysis

- **Model Optimization:**
  - Fine-tune hyperparameters to improve performance.
  - Use techniques like cross-validation to ensure robustness.
  - Fine-tune hyperparameters to improve performance.

- **Deployment:** Implement the trained model into a clinical setting:
    - Provide a user-friendly interface for healthcare providers.
    - Ensure compatibility with existing health records systems.

4. BLOCK DIAGRAM AND DISCRIPTION



1. Data Collection Block: This block is responsible for gathering data from various sources, including:

- Clinical Data: Information from Electronic Health Records (EHRs), such as patient demographics,medical history, and laboratory results.
- Lifestyle Data: Data on patient habits, including smoking history, dietary practices, and physicalactivity levels.
- Genetic Data: Family history of kidney disease and other relevant genetic information. This blockensures comprehensive data collection for accurate CKD prediction.

2. Data Preprocessing Block: This block focuses on preparing the collected data for analysis:

- Handling Missing Values: Utilizes algorithms (like K-Nearest Neighbors) to fill in gaps in the dataset.
- Normalization/Standardization: Adjusts numerical data to a common scale without distorting differences in ranges.
- Encoding Categorical Variables: Converts categorical data into numerical format, making it suitablefor machine learning algorithms.

3. **Feature Selection Block:** This block is crucial for identifying the most relevant features that contribute toCKD prediction:

- Feature Extraction Techniques: Implements methods such as LASSO (Least Absolute Shrinkage andSelection Operator) to select significant variables.
- Dimensionality Reduction: Reduces the number of input variables, helping to enhance model performance and interpretability.

4. **Machine Learning Model Block:** This block houses the core predictive algorithms:

- Model Selection: Chooses appropriate machine learning techniques, primarily neural networks, toanalyze the preprocessed data.
- Training Process: Involves training the model on historical data to recognize patterns associated withCKD.
- Performance Metrics: Evaluates model performance through metrics such as accuracy, precision,recall, and F1-score.

5. **User Interface Block:** This block provides the interface through which healthcare professionals interactwith the system:

- Web Dashboard: A user-friendly platform where clinicians can input patient data, view predictions,and access recommendations.
- Visualization Tools: Graphical representations of risk assessments and trends to aid decision-making.

6. **Risk Assessment Block:** This block categorizes patients based on their CKD risk levels:

- Risk Stratification: Classifies patients into low, moderate, and high-risk groups based on the model'soutput.
- Automated Recommendations: Offers personalized suggestions for monitoring and lifestyle adjustments to mitigate risks.

7. **Alert System Block:** This block ensures timely notifications for healthcare providers:

- Real-time Alerts: Notifies clinicians of significant changes in patient risk status or when new clinicaldata is available.

- Feedback Mechanism: Provides immediate insights based on model predictions, facilitating promptinterventions.

8. Continuous Learning Block: This block emphasizes the adaptability of the system:

- Model Retraining: Allows for ongoing updates and improvements to the model using new patient data,ensuring its accuracy over time.
- Data Integration: Incorporates new findings and trends in CKD management into the system.

9. Integration with Healthcare Systems Block: This block focuses on the system's interoperability:

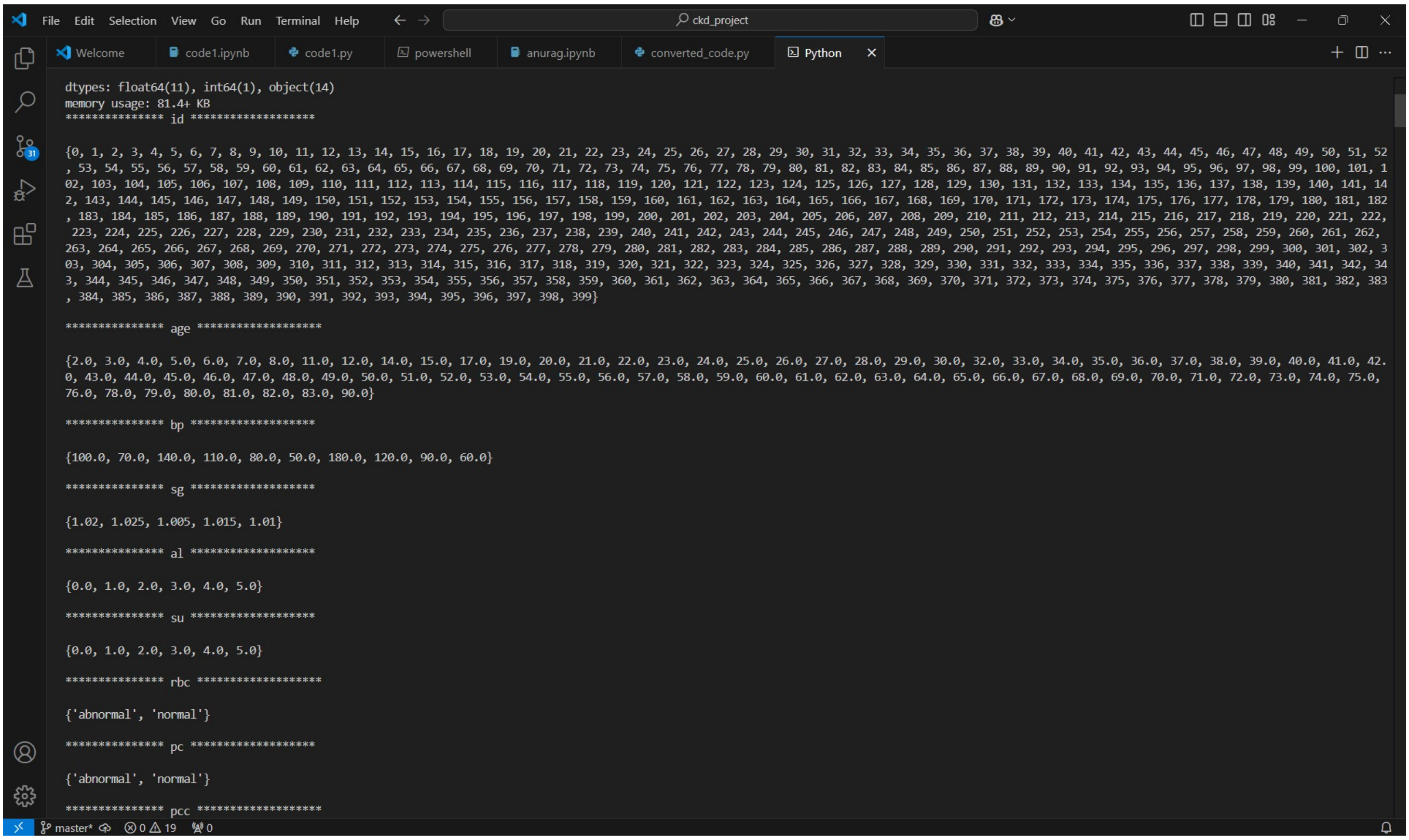
- Compatibility: Ensures the CKD detection system can seamlessly integrate with existing health recordsystems.
- Data Security: Implements protocols to protect patient data and maintain confidentiality.

Gathering the dataset:

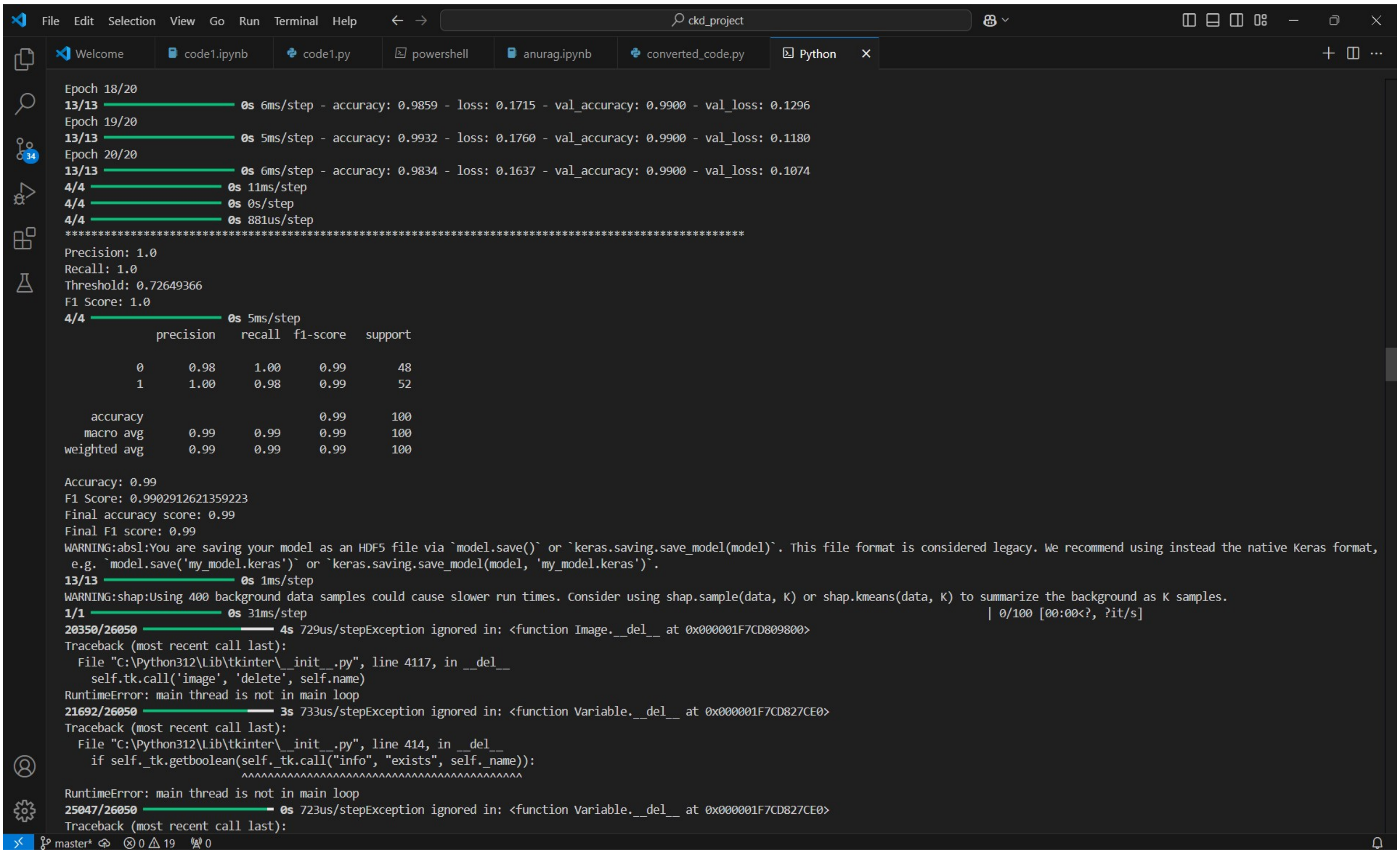
id	age	bp	sg	al	su	rbc	pc	pcc	ba	bgr	bu	sc	sod	pot	hemo	pcv	wc	rc	htn	dm	cad	appe
0	48	80	1.02	1	0		normal	notpresent	notpresent	121	36	1.2			15.4	44	7800	5.2	yes	yes	no	gooc
1	7	50	1.02	4	0		normal	notpresent	notpresent		18	0.8			11.3	38	6000		no	no	no	gooc
2	62	80	1.01	2	3	normal	normal	notpresent	notpresent	423	53	1.8			9.6	31	7500		no	yes	no	poor
3	48	70	1.005	4	0	normal	abnormal	present	notpresent	117	56	3.8	111	2.5	11.2	32	6700	3.9	yes	no	no	poor
4	51	80	1.01	2	0	normal	normal	notpresent	notpresent	106	26	1.4			11.6	35	7300	4.6	no	no	no	gooc
5	60	90	1.015	3	0			notpresent	notpresent	74	25	1.1	142	3.2	12.2	39	7800	4.4	yes	yes	no	gooc
6	68	70	1.01	0	0		normal	notpresent	notpresent	100	54	24	104	4	12.4	36			no	no	no	gooc
7	24		1.015	2	4	normal	abnormal	notpresent	notpresent	410	31	1.1			12.4	44	6900	5	no	yes	no	gooc
8	52	100	1.015	3	0	normal	abnormal	present	notpresent	138	60	1.9			10.8	33	9600	4	yes	yes	no	gooc
9	53	90	1.02	2	0	abnormal	abnormal	present	notpresent	70	107	7.2	114	3.7	9.5	29	12100	3.7	yes	yes	no	poor
10	50	60	1.01	2	4		abnormal	present	notpresent	490	55	4			9.4	28			yes	yes	no	gooc
11	63	70	1.01	3	0	abnormal	abnormal	present	notpresent	380	60	2.7	131	4.2	10.8	32	4500	3.8	yes	yes	no	poor
12	68	70	1.015	3	1		normal	present	notpresent	208	72	2.1	138	5.8	9.7	28	12200	3.4	yes	yes	yes	poor
13	68	70						notpresent	notpresent	98	86	4.6	135	3.4	9.8				yes	yes	yes	poor
14	68	80	1.01	3	2	normal	abnormal	present	present	157	90	4.1	130	6.4	5.6	16	11000	2.6	yes	yes	yes	poor
15	40	80	1.015	3	0		normal	notpresent	notpresent	76	162	9.6	141	4.9	7.6	24	3800	2.8	yes	no	no	gooc
16	47	70	1.015	2	0		normal	notpresent	notpresent	99	46	2.2	138	4.1	12.6				no	no	no	gooc
17	47	80						notpresent	notpresent	114	87	5.2	139	3.7	12.1				yes	no	no	poor
18	60	100	1.025	0	3		normal	notpresent	notpresent	263	27	1.3	135	4.3	12.7	37	11400	4.3	yes	yes	yes	gooc
19	62	60	1.015	1	0		abnormal	present	notpresent	100	31	1.6			10.3	30	5300	3.7	yes	no	yes	gooc
20	61	80	1.015	2	0	abnormal	abnormal	notpresent	notpresent	173	148	3.9	135	5.2	7.7	24	9200	3.2	yes	yes	yes	poor
21	60	90						notpresent	notpresent		180	76	4.5		10.9	32	6200	3.6	yes	yes	yes	gooc
22	48	80	1.025	4	0	normal	abnormal	notpresent	notpresent	95	163	7.7	136	3.8	9.8	32	6900	3.4	yes	no	no	gooc
23	21	70	1.01	0	0		normal	notpresent	notpresent										no	no	no	poor
24	42	100	1.015	4	0	normal	abnormal	notpresent	present		50	1.4	129	4	11.1	39	8300	4.6	yes	no	no	poor
25	61	60	1.025	0	0		normal	notpresent	notpresent	108	75	1.9	141	5.2	9.9	29	8400	3.7	yes	yes	no	gooc
26	75	80	1.015	0	0		normal	notpresent	notpresent	156	45	2.4	140	3.4	11.6	35	10300	4	yes	yes	no	poor
27	69	70	1.01	3	4	normal	abnormal	notpresent	notpresent	264	87	2.7	130	4	12.5	37	9600	4.1	yes	yes	yes	gooc
28	75	70		1	3			notpresent	notpresent	123	31	1.4							no	yes	no	gooc
29	68	70	1.005	1	0	abnormal	abnormal	present	notpresent		28	1.4			12.9	38			no	no	yes	gooc
30		70						notpresent	notpresent	93	155	7.3	132	4.9					yes	yes	no	gooc
31	73	90	1.015	3	0		abnormal	present	notpresent	107	33	1.5	141	4.6	10.1	30	7800	4	no	no	no	poor
32	61	90	1.01	1	1		normal	notpresent	notpresent	159	39	1.5	133	4.9	11.3	34	9600	4	yes	yes	no	poor
33	60	100	1.02	2	0	abnormal	abnormal	notpresent	notpresent	140	55	2.5			10.1	29			yes	no	no	poor
34	70	70	1.01	1	0	normal		present	present	171	153	5.2							no	yes	no	poor
35	65	90	1.02	2	1	abnormal	normal	notpresent	notpresent	270	39	2			12	36	9800	4.9	yes	yes	no	poor

Finding missing values from the dataset and normalizing them with other features.

5. SOFTWARE SYSTEM



```
dtypes: float64(11), int64(1), object(14)
memory usage: 81.4+ KB
*************** id *******************

{0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52
, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 101, 1
02, 103, 104, 105, 106, 107, 108, 109, 110, 111, 112, 113, 114, 115, 116, 117, 118, 119, 120, 121, 122, 123, 124, 125, 126, 127, 128, 129, 130, 131, 132, 133, 134, 135, 136, 137, 138, 139, 140, 141, 14
2, 143, 144, 145, 146, 147, 148, 149, 150, 151, 152, 153, 154, 155, 156, 157, 158, 159, 160, 161, 162, 163, 164, 165, 166, 167, 168, 169, 170, 171, 172, 173, 174, 175, 176, 177, 178, 179, 180, 181, 182
, 183, 184, 185, 186, 187, 188, 189, 190, 191, 192, 193, 194, 195, 196, 197, 198, 199, 200, 201, 202, 203, 204, 205, 206, 207, 208, 209, 210, 211, 212, 213, 214, 215, 216, 217, 218, 219, 220, 221, 222,
 223, 224, 225, 226, 227, 228, 229, 230, 231, 232, 233, 234, 235, 236, 237, 238, 239, 240, 241, 242, 243, 244, 245, 246, 247, 248, 249, 250, 251, 252, 253, 254, 255, 256, 257, 258, 259, 260, 261, 262,
263, 264, 265, 266, 267, 268, 269, 270, 271, 272, 273, 274, 275, 276, 277, 278, 279, 280, 281, 282, 283, 284, 285, 286, 287, 288, 289, 290, 291, 292, 293, 294, 295, 296, 297, 298, 299, 300, 301, 302, 3
03, 304, 305, 306, 307, 308, 309, 310, 311, 312, 313, 314, 315, 316, 317, 318, 319, 320, 321, 322, 323, 324, 325, 326, 327, 328, 329, 330, 331, 332, 333, 334, 335, 336, 337, 338, 339, 340, 341, 342, 34
3, 344, 345, 346, 347, 348, 349, 350, 351, 352, 353, 354, 355, 356, 357, 358, 359, 360, 361, 362, 363, 364, 365, 366, 367, 368, 369, 370, 371, 372, 373, 374, 375, 376, 377, 378, 379, 380, 381, 382, 383
, 384, 385, 386, 387, 388, 389, 390, 391, 392, 393, 394, 395, 396, 397, 398, 399}

*************** age *******************

{2.0, 3.0, 4.0, 5.0, 6.0, 7.0, 8.0, 11.0, 12.0, 14.0, 15.0, 17.0, 19.0, 20.0, 21.0, 22.0, 23.0, 24.0, 25.0, 26.0, 27.0, 28.0, 29.0, 30.0, 32.0, 33.0, 34.0, 35.0, 36.0, 37.0, 38.0, 39.0, 40.0, 41.0, 42.
0, 43.0, 44.0, 45.0, 46.0, 47.0, 48.0, 49.0, 50.0, 51.0, 52.0, 53.0, 54.0, 55.0, 56.0, 57.0, 58.0, 59.0, 60.0, 61.0, 62.0, 63.0, 64.0, 65.0, 66.0, 67.0, 68.0, 69.0, 70.0, 71.0, 72.0, 73.0, 74.0, 75.0,
76.0, 78.0, 79.0, 80.0, 81.0, 82.0, 83.0, 90.0}

*************** bp *******************

{100.0, 70.0, 140.0, 110.0, 80.0, 50.0, 180.0, 120.0, 90.0, 60.0}

*************** sg *******************

{1.02, 1.025, 1.005, 1.015, 1.01}

*************** al *******************

{0.0, 1.0, 2.0, 3.0, 4.0, 5.0}

*************** su *******************

{0.0, 1.0, 2.0, 3.0, 4.0, 5.0}

*************** rbc *******************

{'abnormal', 'normal'}

*************** pc *******************

{'abnormal', 'normal'}

*************** pcc *******************
```



```
Epoch 18/20
13/13 0s 6ms/step - accuracy: 0.9859 - loss: 0.1715 - val_accuracy: 0.9900 - val_loss: 0.1296
Epoch 19/20
13/13 0s 5ms/step - accuracy: 0.9932 - loss: 0.1760 - val_accuracy: 0.9900 - val_loss: 0.1180
Epoch 20/20
13/13 0s 6ms/step - accuracy: 0.9834 - loss: 0.1637 - val_accuracy: 0.9900 - val_loss: 0.1074
4/4 0s 11ms/step
4/4 0s 0s/step
4/4 0s 881us/step
*************************************************************************************************
Precision: 1.0
Recall: 1.0
Threshold: 0.72649366
F1 Score: 1.0
4/4 0s 5ms/step
              precision    recall  f1-score   support

           0       0.98      1.00      0.99        48
           1       1.00      0.98      0.99        52

    accuracy                           0.99       100
   macro avg       0.99      0.99      0.99       100
weighted avg       0.99      0.99      0.99       100

Accuracy: 0.99
F1 Score: 0.9902912621359223
Final accuracy score: 0.99
Final F1 score: 0.99
WARNING:absl:You are saving your model as an HDF5 file via `model.save()` or `keras.saving.save_model(model)`. This file format is considered legacy. We recommend using instead the native Keras format,
 e.g. `model.save('my_model.keras')` or `keras.saving.save_model(model, 'my_model.keras')`.
13/13 0s 1ms/step
WARNING:shap:Using 400 background data samples could cause slower run times. Consider using shap.sample(data, K) or shap.kmeans(data, K) to summarize the background as K samples.
1/1 0s 31ms/step                                                                        | 0/100 [00:00<?, ?it/s]
20350/26050 4s 729us/stepException ignored in: <function Image.__del__ at 0x000001F7CD809800>
Traceback (most recent call last):
  File "C:\Python312\Lib\tkinter\__init__.py", line 4117, in __del__
    self.tk.call('image', 'delete', self.name)
RuntimeError: main thread is not in main loop
21692/26050 3s 733us/stepException ignored in: <function Variable.__del__ at 0x000001F7CD827CE0>
Traceback (most recent call last):
  File "C:\Python312\Lib\tkinter\__init__.py", line 414, in __del__
    if self._tk.getboolean(self._tk.call("info", "exists", self._name)):
                           ^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^
RuntimeError: main thread is not in main loop
25047/26050 0s 723us/stepException ignored in: <function Variable.__del__ at 0x000001F7CD827CE0>
Traceback (most recent call last):
```

6. EXPECTED RESULTS

The Chronic Kidney Disease (CKD) Detection and Guidance System Using Neural Networks is expected to significantly enhance the accuracy of CKD diagnosis, particularly in its early stages, leading to a higher percentage of correct identifications compared to traditional methods. By facilitating timely medical interventions, the system aims to reduce the progression of the disease and its associated complications, ultimately improving patient outcomes. The project will categorize patients based on their risk levels, allowing healthcare providers to prioritize monitoring and management of high-risk individuals. Moreover, personalized treatment plans will be generated using individual patient data, including lifestyle and genetic factors, fostering better engagement and adherence to treatment protocols.

This early detection capability is projected to decrease the need for costly interventions, such as dialysis and kidney transplants, resulting in overall healthcare cost savings. The integration of diverse data sources will provide a comprehensive risk assessment for CKD, empowering healthcare professionals with a user-friendly interface to access critical patient information and recommendations efficiently. Additionally, the system's adaptability will allow it to continuously improve as new patient data becomes available, further enhancing its predictive capabilities. Lastly, the project is poised to contribute to CKD research by identifying trends and potential new risk factors, ultimately advancing the medical community's understanding of this critical public health issue.

7. APPLICATION

Applications :

Early Diagnosis :

Purpose: Detect CKD in its early stages to prevent the progression to advanced stages that require costly treatments like dialysis or transplantation.
Impact: Early detection enables timely interventions, such as medication, dietary changes, and lifestyle modifications, to slow or stop the progression of the disease.
Technology Use: The machine learning model processes clinical data (e.g., glomerular filtration rate, serum creatinine levels) to spot early indicators of kidney dysfunction, even before noticeable symptoms occur.

Risk Stratification :

Purpose: Identify and categorize patients into different risk levels (low, medium, high) based on factors like age, gender, lifestyle habits, and genetic predispositions.
Impact: High-risk patients can be prioritized for closer monitoring and frequent medical checkups, while lower-risk individuals can maintain routine follow-ups.
Technology Use: The system uses predictive analytics and machine learning algorithms to assess the combination of clinical, demographic, and lifestyle data, creating a comprehensive risk profile for each patient.

Predictive Analytics :

Purpose: Predict future disease progression by analyzing trends in a patient's health data, enabling the creation of personalized treatment plans.
Impact: With personalized predictions, healthcare providers can proactively manage potential complications, adjust treatments, and improve patient outcomes. For example, the model may predict when a patient's condition is likely to worsen, enabling adjustments to medication before complications arise.
Technology Use: Predictive models use historical and real-time data to forecast CKD progression and guide treatment strategies tailored to the individual's medical history and condition.

Decision Support Systems :

Purpose: Help healthcare providers make data-driven, informed decisions on patient treatment by offering insights based on the latest medical data and AI-driven predictions.
Impact: The DSS assists clinicians in diagnosing CKD earlier, recommending treatment options, and identifying patients who need immediate interventions. It minimizes diagnostic errors and improves the quality of healthcare.
Technology Use: The system combines clinical decision-making protocols with machine learning to provide actionable recommendations, improving accuracy and efficiency in medical practice.

Resource Allocation :

Purpose: Optimize the use of healthcare resources by identifying which patients need immediate care, which can wait, and how to efficiently distribute resources like medical staff, dialysis machines, and hospital beds.
Impact: Ensuring that patients with urgent needs are treated promptly increases operational efficiency in hospitals and clinics while reducing unnecessary resource usage for lower-risk patients.
Technology Use: The model processes large amounts of patient data to predict the severity of CKD and suggest how resources should be allocated, helping hospitals better manage their facilities, equipment, and healthcare professionals.

Research and Development :

Purpose: Facilitate ongoing research into CKD by analyzing large, real-world datasets to identify trends, risk factors, and new patterns that may not be obvious through traditional methods.
Impact: This contributes to medical research by discovering potential new markers for CKD, better understanding of how it progresses, and improving future treatments. Research efforts can also focus on developing new therapies or preventive measures for CKD based on insights from machine learning.
Technology Use: Machine learning models analyze large datasets, which can lead to breakthroughs in CKD research by identifying non-obvious correlations between clinical and lifestyle factors that contribute to kidney disease.

8. ADVANTAGES AND DISADVANTAGES

Advantages:

Improved Accuracy : Machine learning models enhance diagnostic accuracy by learning from vast datasets, reducing human errors in diagnosis, This ensures more reliable CKD diagnosis, minimizing the risk of misdiagnosis and improving clinical decision-making.

Early Detection : By identifying CKD in its early stages, the system helps prevent the disease from progressing, improving patient outcomes and quality of life, Timely interventions can delay or prevent kidney failure, reducing the burden on patients and healthcare systems.

Data Integration : The system can integrate data from multiple sources (lab results, EHRs, genetic tests) for comprehensive CKD risk assessment and monitoring, This comprehensive view allows for a more personalized and accurate diagnosis by considering a broader range of factors.

Automated Feature Extraction : The system automatically extracts relevant features from patient data, such as lab values and lifestyle factors, simplifying the diagnostic process for healthcare professionals, This reduces the need for manual data processing, saving time and reducing potential human error in feature selection.

Adaptability : The neural network model can be retrained with new patient data, making it adaptable to different populations and evolving healthcare needs, This flexibility ensures the system remains up-to-date with new findings, making it effective across various patient demographics.

Scalability : The system can handle increasing amounts of data and patients as it grows, making it suitable for both small clinics and large healthcare facilities, Its ability to scale ensures that healthcare systems of all sizes can benefit from the technology without compromising performance.

Cost-Effectiveness : Early detection and timely interventions reduce the need for costly treatments like dialysis or transplants, leading to overall cost savings in healthcare, By preventing CKD progression, the system can significantly cut down long-term medical expenses for both patients and healthcare providers.

Personalized Patient Care: The combined system leverages machine learning to provide tailored dietary recommendations specific to the patient's CKD stage, age, and preferences, offering a personalized approach to care that aligns with individual health goals and restrictions.

Disadvantage:

Data Quality Dependence : The accuracy of machine learning models is heavily reliant on the quality and completeness of the input data. Poor data can lead to inaccurate predictions, This limitation can hinder the system's effectiveness in areas with incomplete or inconsistent patient data.

Complexity and Interpretability : Neural networks and other advanced algorithms can act as "black boxes," making it difficult for healthcare professionals to interpret the results and understand the decision-making process, This lack of transparency can make it challenging for clinicians to trust or explain the system's recommendations to patients.

Need for Expertise : Requires specialized knowledge in data science and machine learning, which may not be readily available in all healthcare settings, The necessity for technical expertise can limit the adoption of such systems in smaller or resource-limited healthcare environments

9. CONCLUSION

The integration of machine learning into chronic kidney disease (CKD) detection presents a transformative opportunity to enhance patient care and outcomes. By leveraging advanced algorithms, healthcare providers can achieve improved accuracy in diagnoses, facilitate early detection, and create personalized treatment plans based on a comprehensive analysis of diverse patient data. The ability to automatically extract features and adapt to new information makes machine learning particularly suited for the dynamic nature of healthcare.

However, the implementation of such technologies also comes with challenges, including data quality dependency, complexity, and the need for ethical considerations regarding patient privacy. Ensuring the interpretability of models and mitigating potential biases are essential to foster trust among healthcare professionals and patients alike.

10. REFERENCES

[1] Imesh Udara Ekanayake, et al., "A Machine Learning-Based System for Predicting Chronic Kidney Disease (CKD) Focused on Early Detection through Clinical Data," *Journal of Medical Systems*, 2021.

[2] Nikhila, et al., "Machine Learning-Based System for Predicting Chronic Kidney Disease (CKD) UsingEnsemble Algorithms," *International Journal of Engineering and Advanced Technology*, 2020.

[3] Pankaj Chittora, et al., "Exploring Machine Learning Techniques to Predict Chronic Kidney Disease (CKD)with a Deep Neural Network," *International Journal of Computer Applications*, 2021.

[4] Asif Salekin, John Stankovic, "A Machine Learning-Based Approach to Detect Chronic Kidney Disease(CKD) Using 24 Predictive Attributes," *Health Information Science and Systems*, 2021.

[5] Rajeshwari, H.K. Yogish, "Predicting Chronic Kidney Disease (CKD) Using Machine Learning Techniques for Early Diagnosis," *Journal of King Saud University - Computer and Information Sciences*, 2020.

[6] Chilakamarthi Prem Kashyap, Gollapudi Sai Dayakar Reddy, M. Balamurugan, "Timely Diagnosis of Chronic Kidney Disease (CKD) Using Machine Learning Techniques," *Biomedical Engineering Letters*, 2022.

[7] Zhang Qiu-Li, and Rothenbacher Dietrich, "Prevalence of chronic kidney disease in population-based studies: Systematic review," BMC Public Health, Vol. 8, PP. 117, April 2008.

[8] W. de Souza, L. C. de Abreu, L. G. da Silva, and I. M. P. Bezerra, "Incidence of chronic kidney disease hospitalisations and mortality in Espírito Santo between 1996 to 2017," PLOS ONE, vol. 14, pp. e0224889, Nov. 2019.

[9] Singh, V., Asari, V.K., Rajasekaran, R., “A Deep Neural Network for Early Detection and Prediction of Chronic Kidney Disease,” *Diagnostics*, Vol. 12, pp. 116, January 2022

[10] N. Park, E. Kang, M. Park, H. Lee, H.-G. Kang, H.-J. Yoon, and U. Kang, “Predicting acute kidney injury in cancer patients using heterogeneous and irregular data,” PLOS ONE, Vol.13, PP. 1-21, July 2018.

[11] Dibaba Adeba Debal and Tilahun Melak Sitote, “Chronic Kidney Disease Prediction Using Machine Learning Techniques”, Journal of Big Data, Vol.9, PP.109, 2022.

[12]Abdulhamit Subasi, Emina Alickovic, and Jasmin Kevric, "Diagnosis of Chronic Kidney Disease by Using Random Forest," *IFMBE Proceedings*, vol. 62, pp. 589-594, 2017. doi: 10.1007/978-981-10-4166-2_89

[13]Qin, J.; Chen, L.; Liu, Y.; Liu, C.; Feng, C.; Chen, B. A Machine Learning Methodology for Diagnosing Chronic Kidney Disease. IEEE Access 2019, 8, 20991–21002.

[14]Ilyas, H.; Ali, S.; Ponum, M.; Hasan, O.; Mahmood, M.T.; Iftikhar, M.; Malik, M.H. Chronic kidney disease diagnosis using decision tree algorithms. BMC Nephrol. 2021, 22, 273.
[15]Drall, S.; Drall, G.S.; Singh, S.; Naib, B.B. Chronic kidney disease prediction using machine learning: A new approach. Int. J. Manag. Technol. Eng. 2018, 8, 278–287.
[16]Elhoseny, M.; Shankar, K.; Uthayakumar, J. Intelligent diagnostic prediction and classification system for chronic kidney disease. Sci. Rep. 2019, 9, 9583.
[17]Vasquez-Morales, G.R.; Martinez-Monterrubio, S.M.; Moreno-Ger, P.; Recio-Garcia, J.A. Explainable Prediction of Chronic Renal Disease in the Colombian Population Using Neural Networks and Case-Based Reasoning. IEEE Access 2019, 7, 152900–152910.