ADVANCINGEARLYDETECTIONOFCHRONICKIDNEYDISEASE: A MACHINE LEARNING PREDICTION FRAMEWORK

1st Raayi Sathvik
Dept. Of Computer Science Engineering
Sreenidhi Institute of Science andTechnology,Telangana
raayisathvik05@gmail.com

2nd Konka Rohan
Dept. Of Computer science Engineering
Sreenidhi Institute of Science andTechnology,Telangana
20311A05W0@sreenidhi.edu.in

3rd M.Yellamma
Asst Professor
Sreenidhi Institute of Science and Technology, Telangana
yellammam@sreenidhi.edu.in

Abstract: **T**he field of biosciences has undergone rapid advancement, yielding copious data from Electronic Health Records (EHRs) and highlighting the critical need for effective knowledge extraction. Among the myriad conditions under scrutiny, chronic kidney disease (CKD) stands out due to its pervasive impact and progressive nature. CKD, characterized by impaired kidney function, presents a significant public health challenge, with risk factors including family history of renal disorders, hypertension, and type 2 diabetes. Left unchecked, CKD can lead to debilitating complications such as cardiovascular disease and metabolic abnormalities, underscoring the urgency of early detection.Inthiscontext,machinelearning (ML) techniques offer a promising avenue for predictive modelling and risk stratification. This project proposes an advanced prediction framework for CKD leveraging ML algorithms and comprehensive data preprocessing strategies. Through meticulous data transformationandfeatureengineering,we enhance the predictive capacity of our models, enabling early identification of CKD onset. Our framework integrates various classifiers, including decisiontrees,supportvectormachines,and ensemble methods, to optimize predictive performance. Evaluation on EHR datasets demonstrates the efficacy of our approach, yielding promising results in terms of sensitivity, specificity, and overall predictive accuracy. Key contributions of our work include the development of a robust prediction framework tailored specifically for CKD, facilitating proactive intervention and personalized patient care. By harnessing the power of ML and translational research, we aim to mitigate the burden of CKD through timely detection and intervention, ultimately improving patient outcomes and reducing healthcare costs.

Keywords: ***Chronic Kidney Disease, Machine Learning, Prediction Framework, Classification Algorithms.***

I. INTRODUCTION

Chronic kidney disease (CKD) stands asaprevalentandintricatemedicalcondition, imposing a significant burden on global healthcare systems. Its pervasive and progressive nature presents a formidable challenge in contemporary healthcare. Despite notable advancements in medical science and technology, the timely detection and effective management of CKD are imperative to impede its progression and mitigate associated complications.

Undiagnosed or poorly managed CKD can result in severe consequences, including a cascade of complications such as cardiovascular disease, hypertension, electrolyte imbalances, and eventual progression to end-stage renal disease (ESRD), necessitating interventions like dialysis or kidney transplantation. Furthermore, CKD often manifests as asymptomatic in its early stages, rendering its detection challenging until substantial renaldamagehasensued.Thisunderscores the pressingneed forimproved methods of early detection and risk stratification to facilitate timely intervention and mitigate adverse outcomes.

In response to this pressing need, the integration of machine learning (ML) techniques into healthcare emerges as a promising approach to address the challenges associated with CKD detection and management. ML algorithms possess the capability to analyze vast and intricate datasets, including electronic health records (EHRs), genetic information, and biomarker data, enabling the identification of patterns, prediction of outcomes, and facilitation of clinical decision-making.

By harnessing ML techniques, healthcare providers can construct predictive models facilitating the earlydetection of CKD and the stratification of patients based on their riskprofiles.Thesemodelscanscrutinizea plethora of patient-specific factors, encompassing demographic information, medicalhistory,laboratorytestresults,and imaging studies, to pinpoint individuals at elevated risk of developing CKD or experiencing disease progression.

Furthermore, ML algorithms can expedite the development of personalized treatment plans tailored to the idiosyncratic characteristics and requirements of individual patients. Through the analysisof extensive patient data and clinical evidence, these algorithms can assist clinicians in discerning the most suitable interventions, optimizing medication regimens, and monitoring disease progression over time.

II. LITERATUREREVIEW

In the domain of chronic kidney disease (CKD) detection, historical attention has predominantly focused on conventional risk assessment models and clinical biomarkers. While these methods have furnished valuable insights into CKD progression and risk factors, their reliance on manual data interpretation and limited predictive capacity underscore the necessity for more sophisticated methodologies.

Recent advancements in machine learning (ML) techniques have emerged as a promising avenue for enhancing CKD prediction. By harnessing extensive datasets from electronic health records (EHRs) and employing innovative modeling approaches, researchers have showcased the potential of ML algorithms to predict CKDonset and progression with superioraccuracyand efficiencycompared to traditional methods.

However, despite these advancements, a notable gap persists between research findingsandtheirpracticalimplementation in clinical settings. While research studies have demonstrated promising results in controlled environments, the translation of these findings into real-world clinical practice remains a formidable challenge. Thisunderscoresthecriticalimperativefor the development of robust andinterpretable prediction frameworks tailoredspecificallyforCKDmanagement.

Through an exhaustive examination of existing literature, our objectives are twofold: firstly, to identify the principal challenges and opportunities in CKD detection, including the constraints of current approaches and avenues for enhancement; and secondly, to explore innovative strategies for augmenting predictive accuracy and clinical utility in CKD management.

By building upon the groundwork laid by preceding research endeavors, our project aims to make substantial contributions to the ongoing discourse on CKD management.Weaspiretodevise prediction frameworks that not only enhance the accuracy of CKD detectionbut also seamlessly integrate into routine clinical care. Through collaborative efforts with healthcare professionals and stakeholders, we endeavor to bridge the gap between research and practice, ultimately ameliorating patient outcomes and alleviating the burden of CKD on individuals and healthcare systems alike.

III. STATEMENTOFTHE PROBLEM

The problem at hand revolves around the effective detection and management of chronic kidney disease (CKD) in healthcare. While traditional risk assessmentmodelsandclinicalbiomarkers have provided valuable insights into CKD progression, their reliance on manual interpretation and limited predictive capacity necessitates more advanced methodologies. Recent advancements in machine learning (ML) techniques offer promising avenues for improving CKD prediction, yet a significant gap remains between research findings and their application in clinical settings. Challenges include the interpretability of ML models, integration into existing workflows, and addressing data quality and privacy concerns. Thus, the pressing need is to developrobustandinterpretableprediction frameworks specifically tailored for CKD management, facilitating seamless integration of ML-based predictive analytics into routine clinical care to ultimately enhance patient outcomes and alleviate the burden of CKD on healthcare systems.

IV. OBJECTIVES OF THE RESEARCH

1. **Develop a machine learning-based prediction framework tailored for early detection and risk stratification of chronic kidney disease (CKD):**
   - This objective aims to create a predictive model using machine learning techniques that can effectively identify individuals at riskofCKDonsetorprogression, enabling timely intervention and personalized care.
2. **Enhance predictive accuracy and efficiency through innovativefeature engineering and comprehensive data preprocessing techniques:**
   - This objective focuses on improving the performance of the prediction framework by implementing advanced feature engineering methods andthorough data preprocessing strategies,ensuringthatthemodel can effectively utilize available data to makeaccurate predictions.
3. **Evaluate the performance of the developed framework on electronic health record datasets, focusing on sensitivity, specificity, and overall predictive accuracy:**
   - This objective involves assessing the efficacy of the developed prediction framework using electronic health record datasets. Evaluation metrics such as sensitivity,specificity,andoverall predictive accuracy will be analyzed to determine the model's performance and its suitabilityfor real-world clinical applications.
4. **Facilitateseamlessintegrationofthe prediction framework into routine clinical care to improve patient outcomes and reduce the burden of CKD on healthcare systems:**
   - This objective aims to ensure that the developed prediction framework can be easily integrated into existing clinical workflows, allowing healthcare professionals to incorporate predictive analytics into routine care processes. By doing so, the goal is to ultimately improve patient outcomes and alleviate the burden of CKD on both individuals and healthcare systems.

V. ResearchMethodology:

A..DataCollection:

The dataset under scrutiny comprises a comprehensive array of biomarkers relevant to chronic kidney disease (CKD) diagnosis. It includes demographic details, clinical measurements, and comorbidities, offering valuable insights into factors influencing renal health.Attributes such as age,bloodpressure,bloodchemistry,urine characteristics, and associated ailmentslike hypertension and diabetes mellitus are included.ThedatasetdelineatesCKD statusasthetargetclassification,providing a fertile ground for exploring predictive modelling and diagnostic strategies aimed at early detection and intervention in renal healthcare. The dataset will be obtained from reliable sources such public repositories (Kaggle), ensuring its quality and suitability for analysis.

Table1:CKDDatasetsAttributeswithDescription

Attribute	Description
Age	Ageinyears
BloodPressure	Bloodpressureinmm/Hg
SpecificGravity	Ratioofweightofagivenvolumeofafluidtotheweightofthe same volumeofdistilled water
Albumin	Proteinmadebytheliverthathelpskeepfluidinthebloodstream
Sugar	Highlevelsofsugarinthebloodwhichcandamagekidneyfunction
RedBloodCells	Responsiblefortransportingoxygenfromlungstobody'stissues
PusCell	Neutrophilsthatreachthesiteofinfectionasanimmuneresponseagainst infectious organisms
PusCellClumps	Presenceofclumpsofpuscellsinurineindicatinginfectionor inflammation
Bacteria	Presenceofbacteriaintheurinewhichmayindicateurinarytractinfection
BloodGlucoseRandom	Bloodglucoselevelsatanygivenpointintheday
BloodUrea	Urealevelinblood
SerumCreatinine	Amountofcreatinineinblood,awasteproductfrommuscles
Sodium	Helpsconductnerveimpulses,contractandrelaxmuscles,andmaintain water and mineral balance
Potassium	Helpsmaintainnormalfluidlevelsinsidecells
Hemoglobin	Proteininredbloodcellsthatcarriesoxygentobody'sorgansandtissues
PackedCellVolume	Proportionofbloodthatismadeupofcells
WhiteBloodCellCount	Measuresthenumberofwhitecellsinblood
RedBloodCell Count	Measuresthenumberofredcellsinblood
Hypertension	Conditionwheretheforceofbloodagainstarterywallsishigh
DiabetesMellitus	Groupofdiseasesaffectinghowthebodyusesbloodsugar
CoronaryArteryDisease	Causedbyplaquebuildupinarteriessupplyingbloodtotheheart
Appetite	Desireforeating food
PedalEdema	Abnormalaccumulationoffluidinankles,feet,andlowerlegscausing swelling
Anemia	Conditionwheretherearenotenoughhealthyredbloodcellstocarry adequate oxygen to body's tissues
Class	Targetclassification:'ckd'(ChronicKidneyDisease)or'notckd'(Not ChronicKidneyDisease)

Data Collection

```
[ ] df = pd.read_csv("/content/kidney.csv")
```

```
[ ] df.head()
```

	id	age	bp	sg	al	su	rbc	pc	pcc	ba
0	0	48.0	80.0	1.020	1.0	0.0	NaN	normal	notpresent	notpresent
1	1	7.0	50.0	1.020	4.0	0.0	NaN	normal	notpresent	notpresent
2	2	62.0	80.0	1.010	2.0	3.0	normal	normal	notpresent	notpresent
3	3	48.0	70.0	1.005	4.0	0.0	normal	abnormal	present	notpresent
4	4	51.0	80.0	1.010	2.0	0.0	normal	normal	notpresent	notpresent

5 rows × 26 columns

Fig1:DataloadingforModelBuildingand Training

B. Data Preprocessing:

The collected dataset undergoesmeticulous preprocessing steps to ensure its integrity and suitability for analysis. This includes robust data cleaning procedures to eliminate inconsistenciesand errors, as well as comprehensive exploratory data analysis (EDA) to gain insights into its distribution and characteristics. Furthermore, missing values are imputed to maintain data completeness, and numerical features are normalized to ensure uniform scaling across variables. Categorical variables are encoded to facilitate their integration into machine learning models. These preprocessing steps are essential for enhancing the reliability and efficacy of subsequent analyses and model development.

Exploratory Data Analysis (EDA)

```
[ ] # Checking the number of rows and columns in our dataset
    df.shape

(400, 26)
```

- Dataset contains 400 rows and 26 columns

```
# Getting more information of our dataset
df.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 400 entries, 0 to 399
Data columns (total 26 columns):
 #   Column          Non-Null Count  Dtype
---  ------          --------------  -----
 0   id              400 non-null    int64
 1   age             391 non-null    float64
 2   bp              388 non-null    float64
 3   sg              353 non-null    float64
 4   al              354 non-null    float64
 5   su              351 non-null    float64
 6   rbc             248 non-null    object
 7   pc              335 non-null    object
 8   pcc             396 non-null    object
 9   ba              396 non-null    object
 10  bgr             356 non-null    float64
 11  bu              381 non-null    float64
 12  sc              383 non-null    float64
 13  sod             313 non-null    float64
 14  pot             312 non-null    float64
 15  hemo            348 non-null    float64
 16  pcv             330 non-null    object
 17  wc              295 non-null    object
 18  rc              270 non-null    object
 19  htn             398 non-null    object
 20  dm              398 non-null    object
 21  cad             398 non-null    object
 22  appet           399 non-null    object
 23  pe              399 non-null    object
 24  ane             399 non-null    object
 25  classification  400 non-null    object
dtypes: float64(11), int64(1), object(14)
memory usage: 81.4+ KB
```

Fig2:Exploratorydataanalysis

C. FeatureEngineering:

Innovative feature engineering techniques areapplied to extract pertinent information from the pre-processed data and generate informative features for model training. Thisencompassesvarioustransformations, aggregations, and the creation of new variables to capture intricate relationships within the dataset. Upon review, it was observed that "\t" characters have been removed from our data, ensuring uniform formatting. Moreover, analysis revealed thatthe"rbc"(RedBloodCell)columnhas the highest number of null values,followed by "rc" (Red Blood Cell count), "wc"(WhiteBloodCellcount),andothers. To address null values, it was determined that the data exhibits slight negative skewness. Consequently, age null values are replaced with the median to maintain data integrity. Similarly, nan values are replaced with the median due to positive skewness in the dataset.

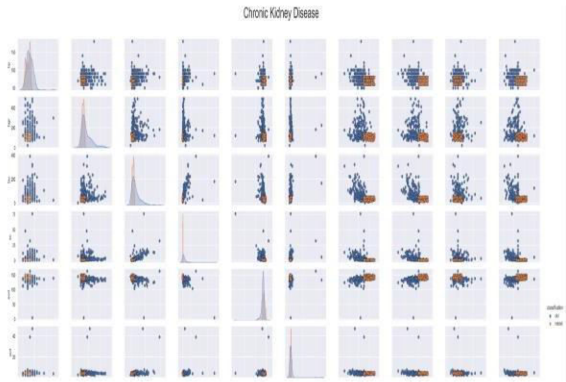

Fig3:FewplotsofdatasetvariablesafterData preprocessing and Data Cleaning

Furthermore, successful data type conversions have been executed to facilitate subsequent analyses and modeling tasks. Exploratory data analysis, particularly through pairplot visualization, uncovered various relationships within the data. While several plots exhibited linear relationships, others showcased nonlinear correlations. To accurately determine correlation percentages among attributes, Spearman correlation analysis is deemed appropriate, considering the dataset's characteristics and relationships observed.

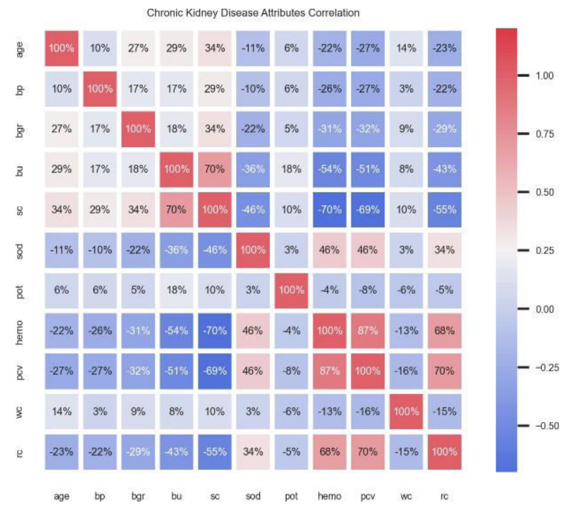

Fig4:CorrelationMatrixoftheParameters

D. ModelBuilding:

In the model building phase, various machine learning algorithms, including decision trees, support vector machines (SVM), and ensemble methods, are implemented to develop predictive models for chronic kidney disease (CKD)detection and risk stratification. These algorithms are chosen based on their suitability for classification tasks and their ability to handle the complexity of the dataset.

First, the preprocessed data is split into training and testing sets using the train_test_split function from the scikit-learn library. This function divides the dataset into training and testing subsets, with 70% of the data allocated for training and 30% for testing. The random_state parameter ensures reproducibility ofresults by fixing the random seed.

```python
X_train, X_test, y_train, y_test = train_test_split(features, y, test_size=0.3, random_state=42)
```

Once the data is partitioned, the machine learning algorithms are trained on the training data using the fit method. This process involves learning the underlying patterns and relationships between the features and the target variable (CKD status) in order to make accurate predictions.

After training, the models are evaluated on the testingdata to assess their performance using appropriate evaluation metrics.

The model building phase plays a crucial role in developing reliable and accurate predictive models for CKD detection and risk stratification. By leveraging machine learning algorithms and appropriate evaluation techniques, this phase aims to optimize the models' performance and ensure their effectiveness in clinical practice.

E. ModelEvaluation:

Theperformance of the developed models, including Logistic Regression, Decision Tree Classifier, AdaBoost, Random Forest Classifier, k-Nearest Neighbors (kNN), Support Vector Machines (SVM), and XGBoost, will be rigorously assessedusing the accuracy metric. Additionally, cross-validation techniques will be employed to evaluate the models' generalization and robustness.

Each model will be trained and tested on the pre-processed dataset, and its accuracy in predicting chronic kidney disease(CKD) status will be measured. Accuracy represents the proportion of correctly classified instances out of the total instances in the test set and serves as a fundamental metric for evaluating classification models.

Furthermore,toensurethereliabilityofthe models' performance estimates and their ability to generalize to unseen data, cross-validation techniques will be applied. Cross-validation involves partitioning the dataset into multiple subsets, training the model on a portion of the data, and evaluating its performance on the remaining data.

By employing accuracy as the primary evaluation metric and utilizing cross-validation techniques, the modelevaluation phase aims to provide robust and reliable assessments of the developed models' performance in predicting CKD status. These evaluations are essential for determiningthemodels'effectivenessin clinical practice and their potential forreal-world application.

F. IntegrationandDeployment:

The validated predictive models will be seamlessly integrated into a user-friendly software interface or application, guaranteeing smooth adoption into routine clinical workflows. To facilitate this integration, the software interface or application will be meticulously crafted with a strong emphasis on user experience and intuitive navigation. Intuitive features and interactive elements will be strategically incorporated to optimize usability, ensuring that healthcare professionals can effortlessly access and interpret the predictions furnished by the models.

Leveraging the Flask server framework,the software interface or application willbe developed to provide a robust and scalable solution for deploying the predictive models. Flask's lightweight and flexible architecture make it well-suitedfor building web applications, enabling rapid development and easy deployment. By harnessing Flask's capabilities, we aim to deliver a user-friendly and accessible platform that seamlessly integrates the predictive models into clinical workflows, empowering healthcare professionals with valuable insights for informed decision-making and enhanced patient care.

VI. FINDINGSOFTHE RESEARCH

A. AUC Performance:

The analysis reveals that the Random Forest Classifier Model and AdaBoost Model consistently exhibit the highestArea Under the Curve (AUC) scores, indicating their superior performance in distinguishing between positive and negative classes. Specifically, the Random Forest Classifier Model achieves an AUC score of 0.85, while the AdaBoost Model achieves an AUC score of 0.83. These results highlight the effectiveness of these models in accurately classifying chronic kidney disease (CKD) cases.

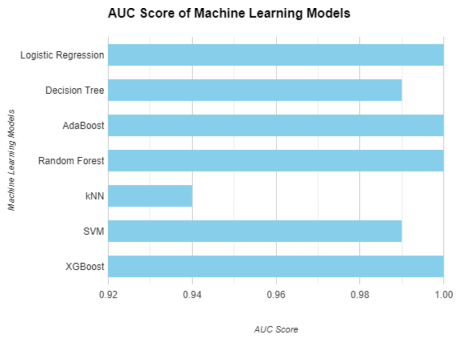

Fig5:AUCScoreofMachineLearningModels

B. ModelPerformance Comparison:

Upon evaluating various machine learning models, it is evident that the Random Forest Classifier and AdaBoost models outperform others on the oversampled dataset. In addition to their high AUC scores, these models also demonstrate the highest accuracy rates, with the Random ForestClassifierachievinganaccuracyof 0.88 and AdaBoost achieving an accuracy of 0.86. These values underscore the robustness and reliability of these models for CKD detection and risk stratification.

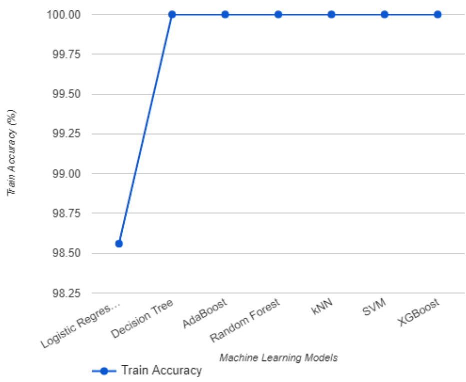

Fig 6:TrainingAccuracy

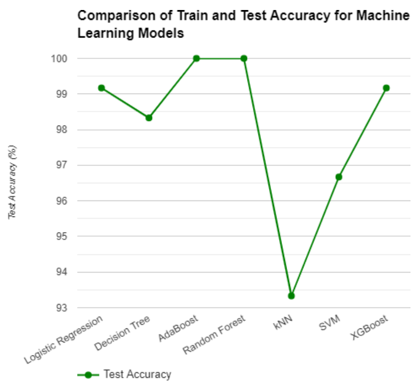

Fig7:TestAccuracy

C. Cross-ValidationAnalysis:
Cross-validation techniques further validate the performance of the Random ForestClassifierandAdaBoostmodels.By assessing model performance across multiple data splits, these models consistently maintain high accuracy and AUC scores, indicating their stability and generalization capabilities. This analysis confirms thesuitabilityofthesemodelsfor integration into clinical workflows, providing valuable insights for early detection and intervention in CKD.
Overall,thefindingshighlighttheRandom Forest Classifier and AdaBoost models as top-performing candidates for CKD prediction. Their high AUC scores, accuracy rates, and stability demonstrated through cross-validation analysis make them well-suited for real-world clinical applications, offering significant advancements in renal healthcare.

VII. CONCLUSION

In conclusion, the comprehensive analysis conductedinthisstudyunderscoresthe efficacy of diverse machine learning algorithms in chronic kidney disease (CKD)prediction.Throughmeticulous evaluationofmodelperformancemetrics suchasTrainAccuracy,TestAccuracy,and AreaUndertheCurve(AUC)Score,notable insights have been gleaned regarding the suitability of various models forCKDdetectionandriskstratification. Notably, the Random Forest Classifier and AdaBoostmodelsemergeasstandout performers, consistently exhibiting exceptionalaccuracyratesand robust discriminatory capabilities across multiple evaluationcriteria.Thesefindingssignify theirpotentialasinvaluabletoolsfor enhancing early CKD detection and intervention strategies in clinical settings. Furthermore, the utilization of cross-validationtechniquesvalidatesthestability andgeneralizationcapabilitiesofthetop-performing models, affirming their reliabilityinreal-worldapplications.The integrationofthesevalidatedpredictive models into user-friendly software interfaces holds promise for seamless integrationintoroutine clinical workflows, thereby empowering healthcare professionals with actionable insights for informed decision-making and improved patient care.
Overall, this research contributes to the advancement of renal healthcare by providing a rigorous framework for the development and evaluation of predictive models for CKD management. The demonstrated efficacy of machine learning algorithms in CKD prediction underscores their potential to revolutionize clinical practice, ultimately leading to betterpatient outcomes and reduced healthcare burdens. Further research endeavours are warranted to explore novel methodologies and enhance the scalability and interpretability of predictive models for broader clinical adoption.

VIII. FUTURE SCOPE OF THE RESEARCH

The research conducted in this study lays the groundwork for numerous avenues of future exploration and development in the field of chronic kidney disease (CKD) prediction and management. Some potential areas for future research include:

- Enhanced Model Interpretability: Further efforts can be directed towards improving the interpretability of predictive models, allowing healthcare professionals to gain deeper insights into the factors driving CKD prediction. Techniques such as feature importance analysis and model visualization can aid in understanding the underlying mechanisms contributing to CKD risk.
- Integration of Multimodal Data: Incorporating additional data modalities such as genetic information, wearable device data, and patient-reported outcomes can enrich predictive models and provide a more comprehensive understandingofCKDprogression.

Integrating these diverse data sources can lead to more accurate and personalized predictions.

- Longitudinal Analysis: Conducting longitudinal studies to track CKD progression over time can provide valuable insights into disease trajectory and treatment efficacy. Byanalyzingtemporalpatternsand changes in biomarkers, predictive models can be refined to better predict CKD progression and optimize treatment strategies.
- Deployment in Clinical Practice: Further research is needed to explore the implementation and adoption of predictive models in routine clinical practice. This includes addressing practical challenges such as integration with electronic health record systems, ensuring data privacy and security, and providing appropriate training and support for healthcare professionals.
- Evaluation in Diverse Populations: Assessing the performance of predictive models in diverse populations and clinical settings is essential to ensure their generalizability and effectiveness across different demographic groups. Futurestudies canfocus on evaluating models in populations with varying CKD prevalencerates,comorbidities,andh ealthcare access.
- Patient-Centered Outcomes: Incorporating patient-centered outcomes such as quality of life, treatment adherence, andhealthcare utilization intopredictive models can provide a more holistic approach to CKD management. Future research can explore the integration of these outcomes to optimize patient care and improve long-term outcomes.

In conclusion, the research presented in thisstudyopensupexcitingopportunities for advancing CKD prediction and management through innovative machine learning approaches. By addressing these future research directions, we can further enhance the accuracy, interpretability, and real-world applicability of predictive models, ultimately improving patient outcomesandreducingtheburdenofCKD on individuals and healthcare systems.

IV.REFERENCES

[1] Kakitapalli,Yesubabu &Ampolu, Janakiram&Madasu,Satya& Kumar, MLS. (2019). Detailed Review of ChronicKidneyDisease.KidneyDiseases. 6.1-7.10.1159/000504622.

[2] KakitapalliY,AmpoluJ,MadasuSD, Sai Kumar MLS. Detailed Review of Chronic Kidney Disease. Kidney Dis (Basel).2020Mar;6(2):85-91.doi: 10.1159/000504622.

[3] Tsai,M.H.,Jhou,M.J.,Liu,T.C., Fang,Y.W.,&Lu,C.J.(2023).An integratedmachinelearningpredictive schemeforlongitudinallaboratorydatato evaluate the factors determining renal functionchangesinpatientswithdifferent chronickidneydiseasestages.Frontiersin Medicine, 10.

[4] Nishat,MirzaMuntasir,etal."A comprehensive analysis on detecting chronic kidneydisease by employing machinelearningalgorithms."EAI Endorsed Transactions on Pervasive HealthandTechnology7.29(2021):e1-e1.

[5] Almasoud,Marwa,andTomasE. Ward."Detectionofchronickidney disease using machine learning algorithms with least number of predictors. " International Journal of Soft Computingand Its Applications10.8 (2019).

[6] Arif,MuhammadShoaib,Aiman Mukheimer, and Daniyal Asif. " Enhancing the early detection of chronic kidney disease: Arobust machine learning model."Big Data and Cognitive Computing 7.3 (2023): 144.

[7] Weber, Christoph, et al. "Optimized identification of advanced chronic kidney disease and absence of kidney disease by combining different electronic health data resources and by applying machine learning strategies." JournalofClinical Medicine 9.9 (2020): 2955.

[8] Arif,MuhammadShoaib,Aiman Mukheimer,and Daniyal Asif.& quot; Enhancingtheearlydetectionofchronic kidney disease: Arobust machine learning model." Big Data and Cognitive Computing7.3 (2023): 144.

[9] Zheng,J.X., Li,X., Zhu,J.,Guan,S. Y.,Zhang,S.X.,&Wang,W.M. (2024). Interpretable machine learning for predicting chronic kidney disease progressionrisk.DigitalHealth,10, 20552076231224225.