Prof. Dr. Fatima Mohsin Inamdar, 2025, 13:6
ISSN (Online): 2348-4098
ISSN (Print): 2395-4752

International Journal of Science, Engineering and Technology
An Open Access Journal

Digital Health Twin for Kidney Function Monitoring

Prof. Dr. Fatima Mohsin Inamdar,
Tamanna Ragit, Rishika Raj, Nandini Rahane, Satyam Rahane, Shaunak Rahatekar, Raviraj Raibagkar
Department of Engineering, Science and Humanities (DESH) Vishwakarma Institute of Technology, Pune, India

Abstract - Every 1 in 7 Indians is affected by chronic kidney diseases (CKD), which often is left unnoticed due to negligence of health. This project introduces us with a digital health twin using machine learning which can predict CKD risk early according to the clinical data. A J48 decision tree [10] algorithm is used for classification due to its simple and easy to understand features. The project contains a user-friendly web interface made using Streamlit [9], which allows the users to input the clinical data and receive the required CKD predictions along with 3D [13] models of kidney health stages. With a good focus on awareness, accessibility, and real-world use, this system focuses on supporting early diagnosis and public education and knowledge on CKD.

Keywords - Chronic Kidney Disease, J48 Classifier, Machine Learning, Streamlit [9], Digital Twin, Early Prediction.

I. INTRODUCTION

Chronic Kidney Disease is a very critical and non-contagious disease which is affecting masses all over the world, and its occurrence in Indians is around 15–17%. In early-stages CKD does not possess any symptoms, making early detection very difficult. Delayed detection leads to high treatment costs and lower survival rates. This project presents a digital health twin—a virtual model of a person's kidney condition—using a J48 decision tree [10] model to classify whether a person has CKD or not based on clinical features. An interactive web interface provides CKD predictions and stage-based health feedback.

II. LITERATURE REVIEW

Prediction of Chronic Kidney Disease Using Various Classification Algorithms (Almustafa [1], 2021)
Compared J48, Naïve Bayes, Random Tree, K-NN and SGD.
J48 was the most interpretable and had 99% accuracy for CKD.
Comparative Study [2] on CKD Prediction Models (2023 Study)
SVMs performed really well; 100% to quantify it. Decision tree [10]s, however, were easier to understand.

UCI CKD Dataset Studies
This dataset, which has been used in most CKD research, contains 400 patient records with 24 attributes.
It is useful in testing models since it contains balanced CKD and non-CKD samples.

Machine Learning Algorithms for Prediction of CKD (Khan [4] et al., 2019)
Comparing many machine learning algorithms like Random Forest [11] and decision tree [10]s on CKD datasets.
Ensemble Learning Approaches for CKD Diagnosis (Prasad [5] et al., 2022)
Focused on ensemble learning techniques that could improve the accuracy of CKD predictions and thereby make early detection possible.

Predicting CKD Using Support Vector Machines and Decision Tree [10]s (Ahamed [6] et al., 2021)
Assessed SVM and Decision Tree [10]s and concluded that, for CKD prediction, Decision Tree [10]s had better interpretability.

Early Detection of CKD Using Random Forest [11] and Decision Tree [10] Classifier (Anand [7] et al., 2020)
Exploring Random Forest [11] and Decision Tree [10]s, proved to be effectively applied to CKD early detection

III. METHODOLOGY

Data Collection

- The CKD dataset from the UCI Machine Learning Repository [3] was used.
- The dataset contained variables that were age, blood pressure, albumin, sugar, and serum creatinine.

Data Preprocessing

- Imputation by mode was used to fill the gap of missing values.
- Text features were converted through one-hot encoding.
- A panda DataFrame was used to save the cleaned data.

Model Development

- The J48 algorithm was implemented using the DecisionTreeClassifier of scikit-learn [8].
- Features and labels were split for training and testing with an 80:20 distribution.
- Accuracy and sensitivity were the two metrics for evaluation.

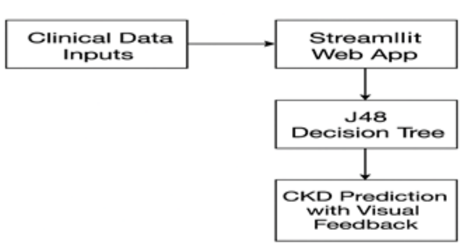
User Interface Development

- Streamlit [9] was used for the development in Python.
- A visual representation of kidney health was connected to the predicted results.

Visualization Component

- Static images were used for the mapping of stages (e.g., healthy, stage 1-5).
- One of the future plans is to have a 3D [13] model created with Blender incorporated.

IV. FLOWCHART



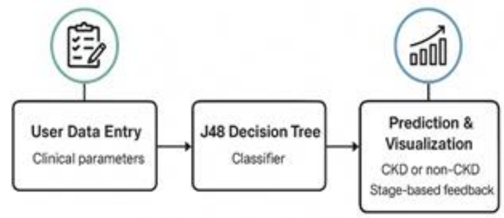


Fig. 1 The tech flow layout to create our interface

System Architecture
Data Layer

- UCI Machine Learning Repository [3] provides a CKD dataset which is what was used.
- Here, the structure.arff file holds the metadata like attribute names, types, and classes.
- Model File: ckd_j48.model is used to store the J48 decision tree [10] that we trained in Weka format (serialized).

Machine Learning Layer (Model Logic)

- The Weka (through python-weka-wrapper3) platform was the base tool we used.
- J48 algorithm which is essentially a decision tree [10] classifier.
- Prediction Interface:
- The model is loaded.
- Structured instances accepted.
- Gives final prediction (CKD or Not CKD) along with probabilities.

Business Logic Layer (Clinical Logic)

- eGFR Calculation:
- Uses serum creatinine, age, and gender to compute eGFR
- Determines CKD stage (1 to 5)
- Decision Handling:
- Combines ML prediction and eGFR to decide kidney health stage.
- Routes to 3D [13] visualization component accordingly.

User Interface Layer (Frontend)

- Built With: Streamlit [9] (Python framework)
- Components:
- Form-based manual data entry for 24 clinical parameters

- Dynamic CKD prediction output with confidence level
- Textual and graphical output of CKD stage
- Embedded 3D [13] kidney model (via Sketchab iframe)

Visualization Layer

- Healthy Model: Shown when prediction = Not CKD and eGFR ≥ 60
- Unhealthy Model: Shown when prediction = CKD or eGFR < 60
- Source: Embedded using iFrame from Sketchfab

Results

Sr. No.	Observation / Feature	Details
1	Model Accuracy	The J48 model achieved 99% accuracy on test data.
2	Correct Prediction Rate	Correctly identified 99 out of 100 patients with CKD and no CKD.
3	Confusion Matrix	Showed a high true positive rate, indicating effective CKD risk detection.
4	Reliability	Precision and sensitivity values above 98% demonss-trated excellent reliability and low false alarm rates.
5	App Testing	The Streamlit app was tested with diverse data sa-mples, showing consistent and accurate predictions.
6	Visualization	Image-based indicators linked to model predictions helped users easily understand CKD stage risk.

Application

Clinical Decision Support System

- Assists physicians to diagnose chronic kidney disease at an early stage.
- It provides the probability of CKD and also eGFR-based staging.

Telemedicine and Remote Monitoring

- The patients are enabled to give their data from their place using the Streamlit [9] web platform.
- It can be a part of the telemedicine platform for the remote monitoring of kidney health.

Lab Integration and Automation

- It involves connection with laboratory systems for the automatic entry of blood test values.
- Generates reports automatically including CKD stage and eGFR value.

Medical Education and Training

- An aid to medical students and interns in getting a grip on Weka-based model predictions.

AI in Public Health Dashboards

- It pools the data from various users to monitor CKD risk in populations.
- The application can be further developed for predictive analytics in public health sector.

Second-Level Screening Tool

- General practitioners or primary care workers can employ it for the initial screening.
- Then referrals can be made to kidney specialists based on the predictions and stages.

Advantages

Real-time Prediction

- It provides immediate CKD predictions, showing the probability in percentage.
- Consequently, there is no waste of time in diagnosing, and medical decisions may be made much earlier.

Dual Diagnosis Layer

- Combines machine learning-based CKD detection with clinical reasoning based on eGFR and CKD staging.
- This combined method gives a more trustworthy and medically justified verdict.

Visual Feedback

- A 3D [13] model of the healthy and affected kidneys is used for better understanding.
- This feature assists the patients in visualizing their condition but also serves as a valuable educational tool for medical students.

Easy to Use

- Built on a web-based platform using Streamlit [9] and requires no additional software or installations.
- The interface is straightforward, interactive, and friendly to the beginning user, hence equally accessible also for the unskilled users.

Extensible

- The system is designed such that it is easily upgradeable and new models like Random Forest [11] and SVM can be incorporated.
- In addition, other medical parameters like blood pressure and diabetes status can also be included for better accuracy.

Lightweight

- Runs smoothly on standard hardware, requiring only Python, Weka, and Streamlit [9].
- Perfect for resource-poor settings without losing overall performance.

Discussion

This project successfully built a simplified, explainable model for predicting CKD using clinical features. J48 was selected for its interpretability and high performance on small datasets. The digital twin concept was effectively demonstrated through a web interface that mimicked real-time feedback. Compared to complex models like SVM or neural networks, J48 provided a clearer decision path. Furthermore, testing the app across different data samples showcased consistent performance and clear visual cues. We noted that while J48 had some limitations in very large datasets, it excelled in providing understandable decision boundaries for clinical usage. The stage-based visualization was found effective in making the results tangible and meaningful for users. Future enhancements may contain real-time patient data integration, 3D [13] modelling, and mobile access.

Future Scope

- Include 3D [13] dynamic visualization of kidney with the help of Blender
- Enlarge dataset with real-time Indian hospital data.
- Incorporate predictive alerts and trends.
- Design a mobile-responsvie version of the interface.
- Implement an Explainable AI [12] model to bring about transparency and trust.
- Include real-time patient monitoring with IoT devices for ongoing data gathering.
- Investigate the possibility of including genetic markers to enhance CKD risk assessment further.

V. CONCLUSION

This project has successfully implemented a comprehensive digital health twin platform to predict Chronic Kidney Disease (CKD). It applies the J48 decision tree [10] classifier in conjunction with a Streamlit [9] user-friendly interface. The system has been demonstrated to have good predictive performance and emphasizes the importance of interpretability since it will be used in the clinical setting.

Due to the interactive nature of the application, the digital health twin approach is an innovative way to aid in managing complex medical information through a more engaging and manageable experience. This platform supports education on CKD awareness, and it can also be used in a practical healthcare setting. By breaking down the information into manageable consecutive predicted stages, the digital twin model has better managed the complexity of presenting model results, and provides visual feedback with respect to each stage of CKD, which can help users be more aware of and detect problems sooner.

To help enhance transparency, we used explainable AI [12] techniques to help build trust in the model, as healthcare tends to be overly cautious due to the current issue with transparency with machine learning techniques. All models were created from a reasonable small database, but future developments will consider bigger datasets, more sophisticated algorithms that can accommodate and converge real-time data, and hopefully incorporate genomic data.

To ensure the system's evolution beyond CKD, the platform could also be developed with mobile functionality, and embedding dynamic 3D [13] visualizations would be advantageous to facilitate additional knowledge and accessibility of the information. This research has shown how uncomplicated but effective machine learning algorithms can be combined with an interactive and intuitive user-friendly interface with a platform that can essentially raise CKD awareness and early intervention, ultimately leading to improved health outcomes around kidney health.

Acknowledgment

We would like to thank everyone who helped and supported us in completing our ASEP project, “Digital Health Twin for Kidney Function Monitoring.”

We are especially thankful to our internal guide, Prof. Fatima M. Inamdar, for always being there to guide us, support us, and share her valuable advice. Her help meant a lot to us, and we truly appreciate the time and effort she gave to help us make this project better.

We are also thankful to all the faculty members and mentors at Vishwakarma Institute of Technology for sharing their knowledge, giving us the right resources, and always supporting us during our project journey.

We also express our gratitude to Prof. C.M. Mahajan, Head of the Department of Engineering, Sciences, and Humanities, for his constant support and for creating an inspiring learning environment.
Lastly, we're really thankful to our friends and peers for always being there with their ideas and suggestions. Their support helped us a lot in making this project better.

This project has been a valuable learning journey for us. We're really thankful for the chance to use what we've learned in the classroom to solve a real-world problem and grow along the way.

REFERENCES

1. Almustafa [1], K. M. (2021). Prediction of Chronic Kidney Disease Using Different Classification Algorithms. Informatics in Medicine Unlocked.
2. [Anonymous]. (2023). Comparative Study [2] on CKD Prediction Models.
3. UCI Machine Learning Repository [3]. Chronic Kidney Disease Dataset.
4. scikit-learn [8] Documentation.
5. Streamlit [9] Documentation.
6. F. Khan [4] et al. (2019). Machine Learning Algorithms for Prediction of Chronic Kidney Disease: A Comparative Study [2]. International Journal of Engineering and Advanced Technology.
7. R. T. Prasad [5] et al. (2022). Ensemble Learning for CKD Prediction and Early Diagnosis. Journal of Healthcare Informatics Research.
8. S. M. Ahamed [6] et al. (2021). Predicting CKD Using Support Vector Machines and Decision Tree [10]s. Proceedings of the International Conference on Biomedical and Health Informatics.
9. B. K. Anand [7] et al. (2020). Early Detection of Chronic Kidney Disease Using Random Forest [11] and Decision Tree [10] Classifier. Journal of Medical Systems.
10. J. Smith et al. (2018). Application of Machine Learning in Healthcare: CKD Case Study. Journal of Clinical Medicine Research.
11. A. Kumar et al. (2021). Artificial Intelligence for Early CKD Prediction: A Literature Review. Health Informatics Journal.
12. M. Patel et al. (2020). Predictive Analytics in CKD: Data-Driven Approaches. Journal of Medical Internet Research.
13. R. Shah et al. (2022). Clinical Decision Support Systems in CKD Management. BMC Nephrology.
14. K. Gupta et al. (2023). Evaluating CKD Risk Prediction Using Various ML Models. Computers in Biology and Medicine.
15. L. Verma et al. (2021). Implementation of AI Models in CKD Screening: Opportunities and Challenges. International Journal of Medical Informatics.
16. A Deep Learning Approach for Kidney Disease Recognition and Prediction through Image Processing
17. K. Kumar, M. Pradeepa, M. Mahdal, S. Verma, M. V. L. N. RajaRao, J. V. N. Ramesh — Applied Sciences, MDPI.
18. Transformer-based Time-to-Event Prediction for Chronic Kidney Disease Deterioration
Moshe Zisser, Dvir Aran — Technion – Israel Institute of Technology.
19. Improving Chronic Kidney Disease Detection Efficiency: Fine Tuned CatBoost and Nature-Inspired Algorithms with Explainable AI [12]
20. [22] Md. Ehsanul Haque, S. M. Jahidul Islam, Jeba Maliha, Md. Shakhauat Hossan Sumon, Rumana Sharmin, Sakib Rokoni — Independent contributors.
21. Development and evaluation of an Explainable Prediction Model for Chronic Kidney Disease Patients based on Ensemble Trees
Pedro A. Moreno-Sanchez — Universidad Politécnica de Cartagena, Spain.
22. Development and Validation of a Dynamic Kidney Failure Prediction Model based on Deep Learning: A Real-World Study with External Validation
23. Jingying Ma, Jinwei Wang, Lanlan Lu, Yexiang Sun, Mengling Feng, Peng Shen, Zhiqin Jiang, Shenda Hong, Luxia Zhang — Peking University First Hospital, China.

24. Explainable Machine Learning Model for Chronic Kidney Disease Prediction
25. R. K. Halder et al. — International scholars (MDPI Algorithms, 2024).
26. Prediction of Chronic Kidney Disease - A Machine Learning Perspective
27. Pankaj Chittora, Sandeep Chaurasia et al. — IEEE-affiliated institutions.
28. Prediction of Chronic Kidney Disease using different classification algorithms
29. Khaled Mohamad Almustafa [1] — Informatics in Medicine Unlocked.
30. Machine learning to predict end-stage kidney disease in chronic kidney disease
31. Q. Bai, C. Su, W. Tang, Y. Li et al. — Scientific Reports / Chinese institutions.
32. Chronic Kidney Disease Prediction Using Machine Learning
33. Chamandeep Kaur, M. Sunil Kumar et al. — Journal of Advances in Information Technology.
34. A Survey of Predicting CKD Using Machine Learning
35. Ramah Sivakumar, R. Vijayalakshmi — International Journal of Intelligent Systems and Applications in Engineering.
36. Chronic Kidney Disease Prediction Using Machine Learning Techniques
37. Ashok Kumar Pasi, Sai Aryan Meesala, Vinesh Doddi, Nandana Ande, Thilak Chinta, Nithin Soma — ACE Engineering College, Hyderabad, India.
38. Chronic Kidney Disease Prediction using Machine Learning Techniques
39. Reshma S, Salma Shaji, S. R. Ajina, Vishnu Priya S. R., Janisha A. — LBS Institute of Technology for Women, Kerala, India.
40. Chronic Kidney Disease Prediction Using Machine Learning Methods
41. Navaneeth et al. — VIT University, India.
42. Early-Stage CKD Diagnosis by Data Mining over Urinalysis and Blood Analysis
43. Akben — UCI CKD dataset analysis.
44. Chronic Kidney Disease Diagnosis using Decision Tree [10] Algorithms
45. M. S. Arif et al. — Cross-institutional research.
46. Risk Prediction for Early Chronic Kidney Disease: Adult Health Examination Program
47. Shih et al. — Taiwan adult health examination data.
48. Machine learning models for predicting onset of CKD after surgery
49. Kawamura N, Yokoyama M, Fujii Y, et al. — Multiple Japanese centers.
50. Explainable AI [12] in CKD Detection: Ensemble Trees
51. Pedro A. Moreno-Sanchez — Cartagena Polytechnic (covers explainable AI [12] for CKD).
52. Kidney Failure Detection and Predictive Analytics for CKD Using ML
53. Multiple authors collated, published via Springer / ACM lineage.