iJETRM
International Journal of Engineering Technology Research & Management
Published By:
https://www.ijetrm.com/

DATA-DRIVEN EARLY DIAGNOSIS OF CHRONIC KIDNEY DISEASE: DEVELOPMENT AND EVALUATION OF AN EXPLAINABLE AI MODEL

Mrs. Katuri Lalitha
Assistant Professor, Department of Information Technology,
Vidya Jyothi Institute of Technology, Hyderabad, Telangana, India

Ayush Mishra, A. Abhilash Goud, K. Sai Kishan
UG Students, Department of Information Technology,
Vidya Jyothi Institute of Technology, Hyderabad, Telangana, India

mishraayush101003@gmail.com abhilashgoud96189@gmail.com Ksaikishan08@gmail.com

ABSTRACT:
Chronic Kidney Disease (CKD) is becoming an increasingly common health concern and, if not identified early, can result in severe complications and premature death. The integration of Artificial Intelligence (AI) and Machine Learning (ML) can assist in detecting CKD at earlier stages, potentially minimizing kidney deterioration. However, the lack of interpretability in AI models often deters clinical adoption. To address this, explainable AI (XAI) methodologies are employed to enhance the transparency of CKD prediction models. This project focuses on creating a data-centric and interpretable model that quantifies the influence of specific clinical indicators in forecasting CKD. The proposed approach utilizes an Extreme Gradient Boosting classifier with three key features—hemoglobin level, specific gravity, and hypertension—validated using 5-fold cross-validation. The model demonstrates strong performance on new, unseen data, offering a cost-effective solution for early CKD diagnosis.

Keywords:
Chronic Kidney Disease, Early Detection, Artificial Intelligence, Machine Learning, XAI, Clinical Features, Prediction Model, Extreme Gradient Boosting.

1. INTRODUCTION

Chronic Kidney Disease (CKD) is recognized globally as a major public health challenge due to its increasing occurrence and potential to cause early death. It ranks among the few non-communicable diseases with a rising mortality rate over the past two decades, imposing a heavy load on healthcare systems. This burden is especially pronounced in low- and middle-income countries, where access to renal replacement therapies is often limited. CKD is commonly triggered by underlying conditions such as diabetes and high blood pressure, and it frequently coexists with cardiovascular diseases, which are leading contributors to early death among CKD patients

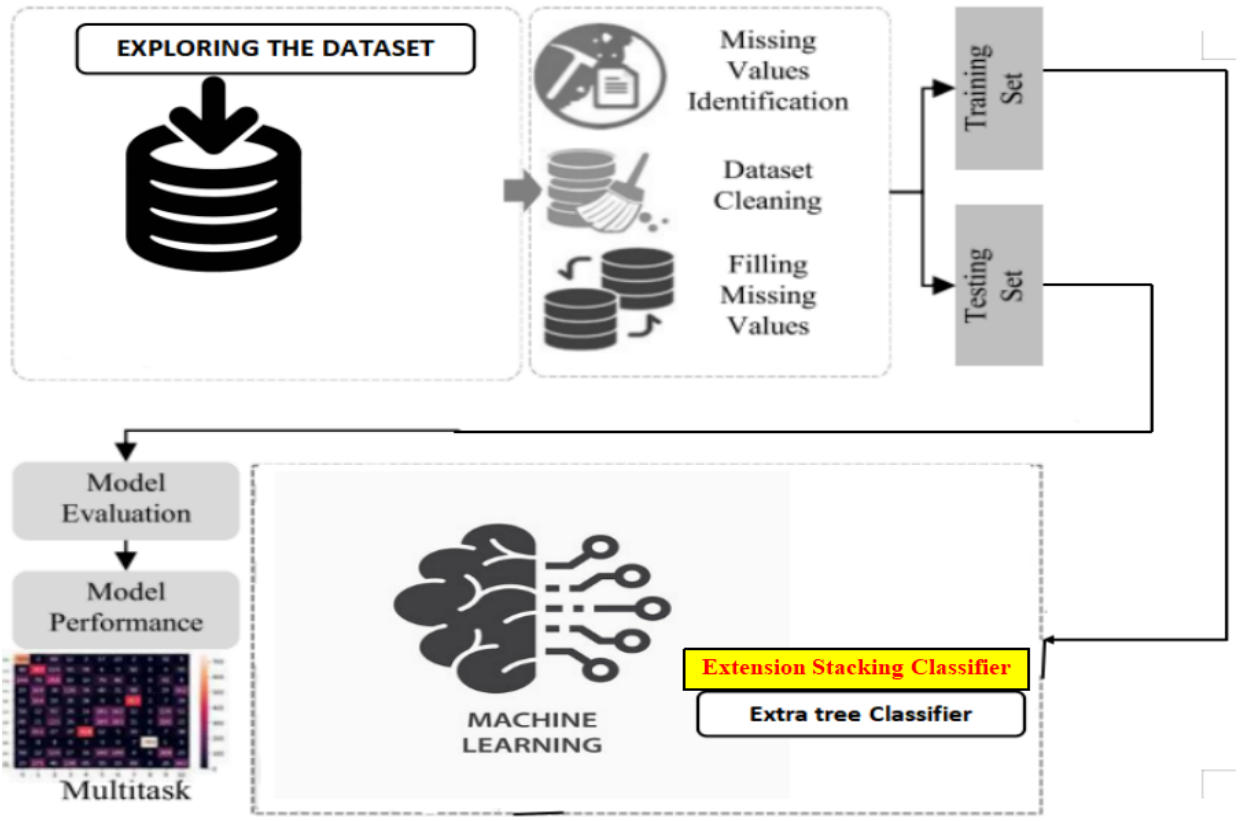
RELATED WORK

Chronic Kidney Disease (CKD) has been the subject of numerous studies aiming to improve early diagnosis using machine learning (ML) and artificial intelligence (AI) approaches. Prior research has extensively utilized classifiers such as logistic regression, support vector machines (SVM), decision trees, and ensemble methods like Random Forest and Gradient Boosting for CKD prediction. These models, while effective in identifying potential cases, often rely on a large number of clinical features, which can increase the cost and complexity of diagnosis.

SYSTEM DESIGN

This system design begins with dataset exploration and preprocessing, where missing values are identified and filled.
The cleaned dataset is split into training and testing sets for model development.
A machine learning module applies an Extension Stacking Classifier, incorporating the Extra Tree Classifier for improved performance.
The trained model undergoes evaluation using multitask performance metrics to ensure reliability and accuracy.
Finally, the results feed back into model refinement, supporting an efficient and interpretable CKD prediction system

iJETRM
International Journal of Engineering Technology Research & Management
Published By:
https://www.ijetrm.com/



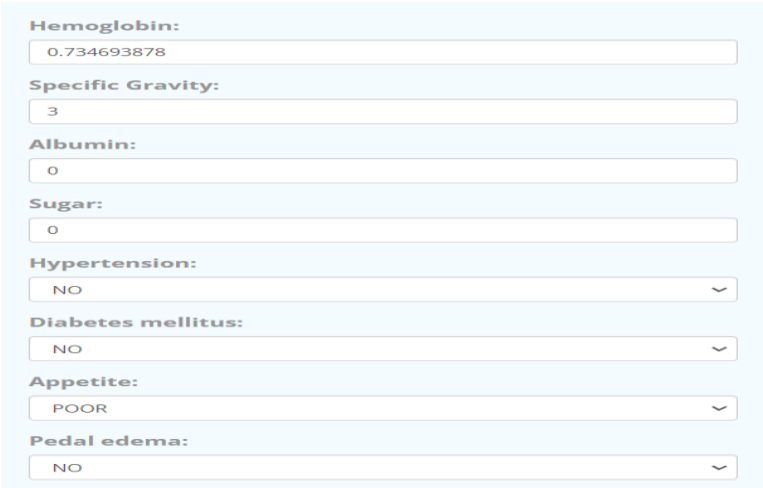
METHODOLOGIES

1. **Data Preprocessing**: The dataset is explored for missing values, followed by cleaning and filling those gaps to prepare reliable input for model training..
2. **Feature Selection**: Specific critical features (hemoglobin, specific gravity, and hypertension) are selected to minimize cost and maintain accuracy.
3. 3. **Model Building**: An **Extreme Gradient Boosting (XGBoost)** classifier is developed and trained on the cleaned data for predictive modeling.
4. **Explainability Integration**: eXplainable AI (XAI) techniques are applied to ensure interpretability of how each feature influences predictions.
5. **Validation**: A 5-fold cross-validation strategy is used to evaluate the model's generalization and accuracy on unseen data.

5. RESULTS

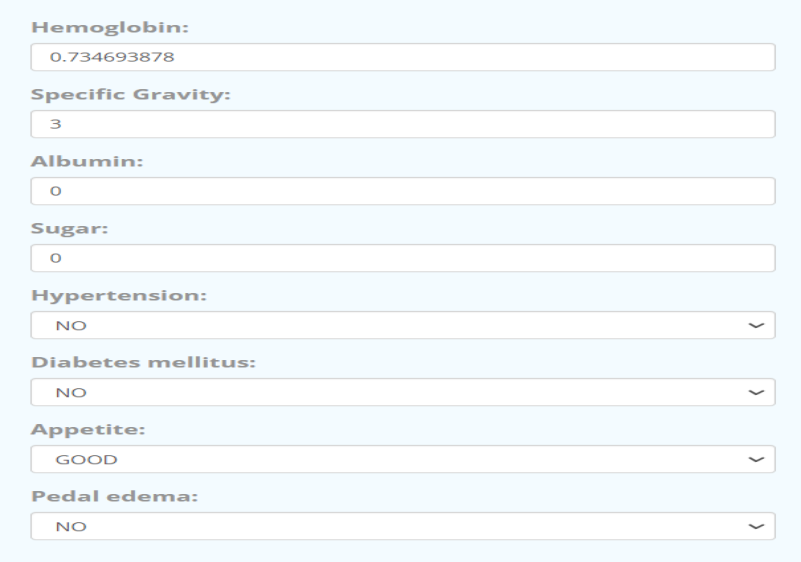
After login and successfully providing appropriate data we get the following resuts:

iJETRM
International Journal of Engineering Technology Research & Management
Published By:
https://www.ijetrm.com/

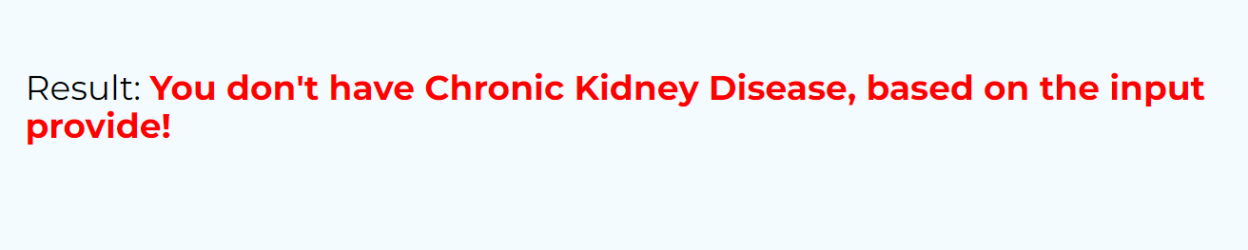


INPUT 1

RESULT 1



INPUT 1



RESULT 2

iJETRM
International Journal of Engineering Technology Research & Management
Published By:
https://www.ijetrm.com/

ACKNOWLEDGEMENT

We extend our heartfelt thanks to **Mrs. Katuri Lalitha,** Assistant Professor, for her invaluable mentorship and guidance throughout the project. Her expertise, encouragement, and insightful feedback significantly shaped our approach and methodology.

We are also thankfulto **Dr.Ramesh Cheripelli** for enriching our understanding and providing innovative solutions to the challenges we encountered for the numerous discussions, brainstorming sessions, and academic exchanges that took place within our project team. These interactions for his exceptional support and guidance throughout the project.

We wish to express our heartful thanks to **Dr.A. Obulesu, HOD**, for guiding and supporting us throughout the course.

We wish to express our heartful thanks to our Beloved Dean,Accreditation and Rankings **Dr.A.Padmaja**, whose support was indispensable to us during the course.

We wish to express our heartful thanks to the Principal **Dr.A.Srujana**,whose support was indispensable to us during the course.

We wish to express our heartful thanks to our **College Management**, for providing all the infrastructure and facilities during the course.

We would like to express our sincere gratitude to our **Parents,** fortheir unwavering support throughout the completion of this major project.

CONCLUSION AND FUTURE WORKS

This study presents the development and evaluation of an explainable prediction model for CKD early diagnosis. Using an automated optimization framework, the best combination of the ensemble tree algorithm and the number of features are selected to provide the best balanced model according to the classification and explain ability metrics. In this paper an explainable prediction model for CKD early diagnosis has been developed and evaluated using an XGBoost classifier that uses three features: hemoglobin, specific gravity, and hypertension. The model provides a good accuracy. The prediction model developed in this work achieved the classification performance of the best CKD prediction models identified in the literature with the least number of features selected compared to the other works. The paper concludes that the developed model provides a promising solution for developing countries as it results in a reduced cost of the early diagnosis of CKD.

REFERENCES

1. Kovesdy, C. P. (2022). Epidemiology of Chronic Kidney Disease: An Update. National Institutes of Health. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9073222/
2. Polat, H., Mehr, H. D., & Cetin, A. (2017). Diagnosis of Chronic Kidney Disease Based on Support Vector Machine by Feature Selection Methods. PubMed. https://pubmed.ncbi.nlm.nih.gov/28243816/
3. Webster, A. C., Nagler, E. V., Morton, R. L., & Masson, P. (2017). Chronic Kidney Disease. The Lancet. https://pubmed.ncbi.nlm.nih.gov/27887750/
4. Cockwell, P., & Fisher, L.-A. (2019). The Global Burden of Chronic Kidney Disease. The Lancet. https://www.thelancet.com/journals/lancet/article/PIIS0140-6736(19)32977-0/fulltext
5. Jeewantha, R. A., Halgamuge, M. N., Mohammad, A., & Ekici, G. (2017). Classification Performance Analysis in Medical Science: Using Kidney Disease Data. ResearchGate. https://www.researchgate.net/publication/32141427