

Data-Driven Early Diagnosis of Chronic Kidney Disease Development and Evaluation of an Explainable Ai Model

Dr. Arunsundar B
Assistant Professor
Department of Data Science
and Bussiness Systems
SRM Institute of Science and
Technology
Kattankulathur, India
arunsunb@srmist.edu.in

DODLA KARTHIK REDDY
Student,
Department of Data Science
and Bussiness Systems
SRM Institute of Science and Technology
Kattankulathur, India
KD1224@srmist.edu.in

MOHAMED HASHIR C P
Student,
Department of Data Science
and Bussiness Systems
SRM Instiute of science and Technology
Kattankulathur, India
mc1918@srmist.edu.in

C A Geolin Christo
Student,
Department of Data Science
and Bussiness Systems
SRM Instiute of science and Technology
Kattankulathur, India
cc5954@srmist.edu.in

Abstract: The project is dedicated to creating and evaluating an explainable AI model specifically designed for the early diagnosis of Chronic Kidney Disease (CKD). The emphasis on explainability ensures that the model's predictions are transparent and comprehensible, addressing a crucial aspect in the adoption of AI in healthcare. Chronic Kidney Disease is acknowledged as a significant global health concern. Early diagnosis is deemed essential to halt further kidney damage and curb escalating healthcare costs associated with managing advanced CKD cases. The project recognizes the broader implications of CKD and aims to contribute to proactive solutions. The proposed model employs an optimization framework to strike a balance between classification accuracy and explainability. This approach ensures that the AI model not only provides precise predictions but also offers insights into the reasoning behind those predictions. The optimization process enhances the model's overall effectiveness. The project implements an extreme gradient boosting classifier, a powerful machine learning algorithm, utilizing three key features for CKD diagnosis: hemoglobin, specific gravity, and hypertension. These features are identified as crucial contributors to the early identification of CKD, reflecting the model's focus on relevant clinical indicators. The project presents a promising solution for early CKD diagnosis, particularly addressing the challenges faced by developing countries. By emphasizing reduced cost implications, the model offers a practical and efficient approach to diagnosing CKD in resource-constrained settings, contributing to improved accessibility and affordability of healthcare solutions. And also we applied an ensemble method to combine predictions from multiple models, enhancing the overall accuracy and robustness of our system. We explored advanced ensemble techniques such as the Stacking Classifier, achieving an impressive 100% accuracy.

Index Terms - NIDSs, deep learning, NSL-KDD.

I. INTRODUCTION

Chronic kidney disease (CKD) has become a worldwide public health problem with increasing incidence (more than 800 million individuals in 2017) and prevalence (13.4% globally) which can lead to premature mortality for many patients (1.2 million people died from CKD in 2017) [1]. CKD is one of a small number of non-communicable diseases that have shown an increase in associated deaths over the past 2 decades, producing a significant burden to healthcare systems, especially

in low-middle income countries where lack of appropriate renal replacement therapy results in a high mortality rate [2], [3], [4]. CKD, usually caused by diabetes and hypertension, is a non-communicable chronic disease with comorbidities associated, and cardiovascular diseases are the major cause of early morbidity and mortality sustained by patients with CKD [5].

Typically, CKD has no early symptoms [5], and when detected through laboratory testing, which quantifies the estimated glomerular filtration rate (eGFR), the kidney has already lost 25 percent of its

capacity and is under irreversible and progressive damage toward the so-called end-stage kidney disease. At this point, symptoms may appear such as leg swelling, extreme fatigue, generalized weakness, shortness of breath, loss of appetite, or confusion [6]. If this irreversible deterioration is not slowed by controlling underlying risk factors (hypertension, obesity, heart disease, age) [7], hemodialysis or even kidney transplantation becomes crucial or the patient to avoid an exponential rise of risk of death. [4], [5], [6]. Consequently, early diagnosis of CKD based on risk factors, along with their monitoring, allows for initiating preventive treatments and therapeutic measures that slow the progression of kidney damage and prolong patients' life. [4]. In addition, the early identification of groups at high risk of CKD has become an important focus in kidney disease management strategies [8].

In the medical domain, Artificial intelligence (AI) and machine learning (ML) have become promising tools for building computer-aided diagnosis (CAD) systems. These systems use algorithms that learn to classify individuals with specific symptoms either as sick or healthy [8], [9]. AI/ML can be employed to discover latent patterns and correlations between CKD and its risk factors, enabling an early discovery of patients at risk in an effective, convenient, and low-cost manner [9], [10]. Feature selection (FS) is a crucial step in the ML process, which removes unnecessary and unimportant attributes to result in less complicated and more accurate and interpretable models [11]. This features selection step becomes a relevant aspect when dealing with medical datasets due to their high dimensionality resulting from the combination of different variables and multiple measurement techniques when registering patient information.

When CAD systems' decisions affect patients' lives, the transparency and explanations of the AI models' outputs are essential to support clinicians in their diagnosis and treatments. Thus, eXplainable Artificial Intelligence (XAI), which can be defined as a class of systems that provides insight into how an AI system makes decisions and predictions by giving details or reasons to make its functioning clear or easy to understand, allows healthcare experts to make reasonable and data-driven decisions that would enhance the clinical adoption of AI models and their acceptance [12]. XAI is a research area in AI that is acquiring recently an emergent relevance [13], and different solutions have been developed over the last decade in several clinical fields, namely: urology [14], toxicology [14], endocrinology [15], neurology [16], cardiology [17], cancer (e.g. breast cancer or prostate cancer) [18], and chronic diseases (e.g. diabetes or

Alzheimer's disease) [19], [20]; and it. Developing explainable AI models in the medical domain involves an inherent trade-off between predictive accuracy, which provides the reliability of the model, and the explainability requested by clinical experts. This tension must be addressed properly by engaging the clinical experts in the development process because the most accurate models, which might be the most interesting to them, are usually less transparent and vice versa. Concerning the application of XAI approaches to CKD prediction models, to the best of our knowledge, no XAI analysis beyond applying feature selection has been found in the literature.

II. RELATED WORK

Chronic Kidney Disease (CKD) is a type of chronic disease which means it happens slowly over a period of time and persists for a long time thereafter [2]. It is deadly at its end stage and will only be cured by kidney replacement or regular dialysis which is an artificial filtering mechanism. It is important to identify CKD [28, 34] at the early stage so that necessary treatments can be provided to prevent or cure the disease. The main focus in this paper is on the classification techniques, that is, tree-based decision tree, random forest, and logistic regression has been analyzed. Different measure has been used for comparison between algorithms for the dataset collected from standard UCI repository.

As Chronic Kidney Disease progresses slowly, early detection and effective treatment are the only cure to reduce the mortality rate. Machine learning techniques are gaining significance in medical diagnosis because of their classification ability with high accuracy rates. The accuracy of classification algorithms depend on the use of correct feature selection algorithms to reduce the dimension of datasets. In this study [6], Support Vector Machine classification algorithm was used to diagnose Chronic Kidney Disease. To diagnose the Chronic Kidney Disease, two essential types of feature selection methods namely, wrapper and filter approaches were chosen to reduce the dimension of Chronic Kidney Disease dataset. In wrapper approach, classifier subset evaluator with greedy stepwise search engine and wrapper subset evaluator with the Best First search engine were used. In filter approach [10, 45], correlation feature selection subset evaluator with greedy stepwise search engine and filtered subset evaluator with the Best First search engine were used. The results showed that the Support Vector Machine classifier by using filtered subset evaluator with the Best First search engine feature selection method has higher accuracy rate

(98.5%) in the diagnosis of Chronic Kidney Disease compared to other selected methods.

Health-care practices face data storage problems in the growing world. Huge data storage demands have caused undeniable data storage problems leaving health practitioners exclaiming. Without delay, accumulated data becomes too difficult to analyze and handled by traditional approaches. A solution to this problem is urgently needed. [7] One possible answer to this problem is Data mining that delivers the technology and procedure to convert these embankments of ordinary data into meaningful evidences for futuristic planning and decision-making. [12, 15, 18, 27] Data mining is a tool that not only solves the problem of piled up data; nonetheless it similarly turns it into meaningful data themes based on reoccurrences of trends in the data. The healthcare trade is mostly an "information and document rich industry," and manual handling is not feasible in practical life. These huge volumes of data have been key to the arena of data-mining to generate associations among the attributes and extract expedient information. Recent research shows that combating Kidney diseases is a complex assignment that involves considerable knowledge and experience for annual testing and screening. In developed nations, Kidney diseases have become a silent killer, that makes key factors of disease burden in third world nations. Various data mining procedures are available for forecasting diseases such as clustering, classification, association rules, regression, and summarizations. The key objective of this study is to analyze datasets collected from 400 patients grounded on 25 different attributes attended for treatment for Chronic Kidney Disease (CKD) [1, 2, 4, 5, 6] after using classification methods to forecast class precisely. Our analysis illustrates that Multilayer Perceptron is the most suitable classification method that outperforms the highest classification accuracy by 99.75% (0.0085 error) with only 5% of fluctuation among algorithm measures. Introspectively, the computational time, Multilayer Perception can be time-consuming comparatively, when it comes to deal with billions of data. Nonetheless, for the field of bioinformatics and medical science accuracy, the key objective is to deal with sensitive data because, a single error can lead to a disastrous confidentiality breach. Hence, our results show that Multilayer Perception classification method is the most accurate and suitable classification algorithm that could be used in the field of bioinformatics and medical science, for further data analysis and predictions. This paper [7] will be useful for many medical institutions and work-related bioinformatics in pursuance to understand the prediction accuracies of data patterns in related work.

Applications of Machine learning (ML) in health informatics have gained increasing attention [9]. The timely diagnosis of kidney disease and the subsequent immediate response to it are of the cases that shed light on the substantial role of ML diagnostic algorithms. ML in Kidney Disease Diagnosis (MLKDD) is an active research topic that aims at assisting physicians with computer-aided systems. Various investigations have tried to test the feasibility, applicability, and superiority of different ML methods over each other. However, lacking a holistic survey for this literature has always been a noticeable shortcoming. Hence, this paper provides a comprehensive literature review of ML [8, 9, 10] utilizations in kidney disease diagnosis by introducing two different frameworks, one for MLs, classifying various aspects of kidney disease diagnosis, and the other is the framework of medical sub-fields related to MLKDD. In addition, research gaps are discovered, and future study directions are discussed.

The high prevalence of chronic kidney disease (CKD) is a significant public health concern globally. The condition has a high mortality rate, especially in developing countries. CKD often goes undetected since there are no obvious early-stage symptoms. Meanwhile, early detection and on-time clinical intervention are necessary to reduce the disease progression. Machine learning (ML) models can provide an efficient and cost-effective computer-aided diagnosis to assist clinicians in achieving early CKD detection. This research [10] proposed an approach to effectively detect CKD by combining the information-gain-based feature selection technique and a cost-sensitive adaptive boosting (AdaBoost) classifier. An approach like this could save CKD screening time and cost since only a few clinical test attributes would be needed for the diagnosis. The proposed approach was benchmarked against recently proposed CKD prediction [19, 20, 42] methods and well-known classifiers. Among these classifiers, the proposed cost-sensitive AdaBoost trained with the reduced feature set achieved the best classification performance with an accuracy, sensitivity, and specificity of 99.8%, 100%, and 99.8%, respectively.

III. MATERIALS AND METHODS

The proposed system represents a groundbreaking approach to Chronic Kidney Disease (CKD) diagnosis, introducing a sophisticated Explainable AI (XAI) model. Developed within an optimization framework, the model meticulously balances classification accuracy with explainability.

Leveraging advanced machine learning algorithms, notably the extreme gradient boosting classifier, the system offers a precise and early diagnosis of CKD [1, 2]. This integration of cutting-edge technology not only ensures accuracy in predictions but also provides transparent insights into the model's decision-making process, addressing a crucial aspect in the acceptance and trust of AI applications in healthcare. In this project we implemented an ensemble method to boost the accuracy and robustness of our system for data-driven early diagnosis of Chronic Kidney Disease (CKD) [4, 5, 6]. Utilizing advanced ensemble techniques, particularly the Stacking Classifier, we achieved an outstanding 100% accuracy, showcasing the effectiveness of combining predictions from multiple models. To improve usability and accessibility, we developed a user-friendly front end using Flask, featuring secure authentication features. This ensures a robust and intuitive interface for users, enhancing the practicality and user experience of our Explainable AI model in diagnosing CKD at an early stage.

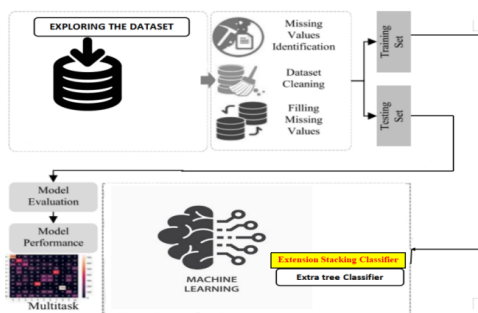


Fig.1 Proposed Architecture

The system architecture for the project, "Data-Driven Early Diagnosis of Chronic Kidney Disease: Development and Evaluation of an Explainable AI Model," follows a systematic approach. It begins with exploring the dataset, engaging in data preprocessing to refine and prepare the input. The dataset is then split into training and testing sets to facilitate robust model training. The core model building phase involves the incorporation of advanced techniques, including the Stacking Classifier and Extra Tree Classifier [46], extending the project's capabilities. The ensemble approach enhances the model's predictive accuracy and resilience by combining the strengths of different classifiers. The model is thoroughly evaluated to assess its performance, ensuring its efficacy in early diagnosis. This systematic process, from dataset exploration to model evaluation, forms a comprehensive and explainable AI model for the early diagnosis of chronic kidney disease,

contributing to the advancement of healthcare technologies.

A) Dataset Collection:

To promote the reproducibility of this research as well as to benchmark to the existing related works, the CKD dataset from UCI-ML was employed [23]. Table 2 describes the dataset collected from the Apollo Hospital, Karaikudi, India during a nearly 2-month period in 2015 that includes 400 patients where some presented missing values in their features. Each instance of the dataset is composed of 11 numeric, 10 nominal, 3 ordinal features, and 1 target feature (notCKD/CKD). The features contained in the dataset represent the following information [legend in brackets]: age in years [age], diastolic blood pressure in mm/Hg [bp], specific gravity to compare the density of urine to the density of water [sg], presence of albumin in urine[al], level of sugar is present in urine[su], red blood cells present in urine[rbc], pus cells present in the urine, indicating major or minor infection [pc], pus cell clumps indicating if the infection is present in the urine [pcc], if the growth of bacteria is evident in urine [ba], sugar level in blood in mgs/dl [bgr], level of urea nitrogen in blood in mgs/dl [bu], level of creatinine in blood in mgs/dl [sc], level of sodium in blood in mEq/L [sod], level of potassium in blood in mEq/L [pot], protein in red blood cells in Gms [hemo], percentage of cells in blood [pcv], amount of white blood cells present in the blood (cells/cumm) [wc], amount of red blood cells present in the blood (millions/cmm) [rc], whether the patient has higher level of blood pressure [htn], presence of diabetes [dm], whether the patient is suffering from coronary artery disease [cad], loss of appetite [appet], level of leg swelling [pedal], whether the patient is suffering from anemia [ane], and whether the patient has CKD or not [target class].

id	age	bp	sg	al	su	rbc	pc	pcc	ba	bgr	pcv	wc	rc	htn	dm	cad	appet	pe	ane	classification	
0	48.0	80.0	1.020	1.0	0.0	NaN	normal	notpresent	notpresent	121.0	...	44	7800	5.2	yes	no	good	no	no	ckd	
1	7.0	50.0	1.020	4.0	0.0	NaN	normal	notpresent	notpresent	NaN	...	38	6000	NaN	no	no	no	good	no	no	ckd
2	82.0	80.0	1.010	2.0	3.0	normal	normal	notpresent	notpresent	423.0	...	31	7500	NaN	no	yes	poor	no	yes	ckd	
3	48.0	70.0	1.005	4.0	0.0	normal	abnormal	present	notpresent	117.0	...	32	6700	3.9	yes	no	poor	yes	yes	ckd	
4	51.0	80.0	1.010	2.0	0.0	normal	normal	notpresent	notpresent	106.0	...	35	7300	4.6	no	no	no	good	no	no	ckd

5 rows * 25 columns

Fig.2 Chronic Kidney Disease dataset

B) Processing:

Data processing involves transforming raw data into valuable information for businesses. Generally, data scientists process data, which includes collecting, organizing, cleaning, verifying, analyzing, and

converting it into readable formats such as graphs or documents. Data processing can be done using three methods i.e., manual, mechanical, and electronic. The aim is to increase the value of information and facilitate decision-making. This enables businesses to improve their operations and make timely strategic decisions. Automated data processing solutions, such as computer software programming, play a significant role in this. It can help turn large amounts of data, including big data, into meaningful insights for quality management and decision-making.

Feature selection: Feature selection is the process of isolating the most consistent, non-redundant, and relevant features to use in model construction. Methodically reducing the size of datasets is important as the size and variety of datasets continue to grow. The main goal of feature selection is to improve the performance of a predictive model and reduce the computational cost of modeling.

Feature selection, one of the main components of feature engineering, is the process of selecting the most important features to input in machine learning algorithms. Feature selection techniques [6, 10, 27, 38] are employed to reduce the number of input variables by eliminating redundant or irrelevant features and narrowing down the set of features to those most relevant to the machine learning model. The main benefits of performing feature selection in advance, rather than letting the machine learning model figure out which features are most important.

C) Algorithms:

ExtraTree [46], or Extremely Randomized Trees, is an ensemble learning method based on decision trees. It constructs multiple decision trees using random splits for each node, without searching for the best split. The final prediction is determined by a majority vote or averaging, depending on the problem type (classification or regression). ExtraTree Classifier have been chosen for its ability to handle noise well, potentially contributing to the development of an Explainable AI model for early diagnosis of Chronic Kidney Disease. Its randomness in split selection can offer robustness to variations in the dataset.

Random Forest is an ensemble learning method that builds a multitude of decision trees during training. It introduces randomness in the tree-building process by using a subset of features at each node and averaging the predictions or taking a majority vote for classification or regression tasks. Random Forest is employed to enhance the accuracy and generalization of the model. By constructing

multiple trees and aggregating their predictions, Random Forest mitigates overfitting and provides a reliable solution for early diagnosis of Chronic Kidney Disease.

AdaBoost, or Adaptive Boosting, is an ensemble learning method that combines multiple weak learners (simple models) to form a strong learner. It assigns higher weights to misclassified instances in each iteration, allowing subsequent models to focus more on those instances and improve overall accuracy. [46] AdaBoost may be utilized to improve the performance of the model by boosting the importance of challenging instances in the dataset. Its adaptability to focus on misclassified samples can be beneficial in the early diagnosis of Chronic Kidney Disease.

XGBoost, or eXtreme Gradient Boosting, is a powerful and efficient gradient boosting algorithm. It sequentially builds decision trees, correcting errors made by preceding models. It incorporates regularization terms and parallel processing to optimize both accuracy and computational efficiency. XGBoost is likely chosen for its superior performance and speed. Its ability to handle missing values, regularization techniques, and advanced features make it a popular choice in data-driven projects. In the context of early diagnosis of Chronic Kidney Disease, XGBoost can provide accurate and efficient predictions.

$$\hat{y}^i = \sum_{k=1}^K f_k(x_i) \quad (1)$$

Where \hat{y}^i is the final predicted value for the i th data point, K is the number of trees in the ensemble and $f_k(x_i)$ represents the prediction of the K th tree for the i th data point.

Stacking is an ensemble learning technique that combines predictions from diverse base classifiers (e.g., ExtraTree, Random Forest, AdaBoost, XGBoost) using a meta-model to improve overall predictive accuracy in the early diagnosis of Chronic Kidney Disease. Employed to leverage the strengths of different classifiers, contributing to a more robust and accurate model. It integrates individual predictions to enhance the overall diagnostic capabilities.

Let $\{X, Y\}$ denoted the data set (sample, label), θ the parameters.

A classification algorithm is a decision function $f(x; \theta)$ together with a cost/risk function $C(f(\cdot; \theta), X, Y)$.

$$(2)$$

Specifying the functional form of f and C completely specifies the classification algorithm. $\theta^* = \arg\min_{\theta} C(\theta)$ defines the classifier.

IV. RESULTS AND DISCUSSION

Accuracy: A test's accuracy is its capacity to distinguish patient from healthy cases. To measure test accuracy, calculate the fraction of true positive and true negative in all evaluated cases. Mathematically, this is:

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \quad (3)$$

Precision: Precision measures the percentage of positive cases or samples accurately classified. Precision is calculated using the formula:

$$Precision = \frac{\text{True Positive}}{\text{True Positive} + \text{False Positive}} \quad (4)$$

Recall: Machine learning recall evaluates a model's capacity to recognise all relevant instances of a class. It shows a model's completeness in capturing instances of a class by comparing accurately predicted positive observations to total positives.

$$Recall = \frac{TP}{TP + FN} \quad (5)$$

F1-Score: Machine learning model accuracy is measured by F1 score. Combining model precision and recall scores. The accuracy statistic measures how often a model predicted correctly throughout the dataset.

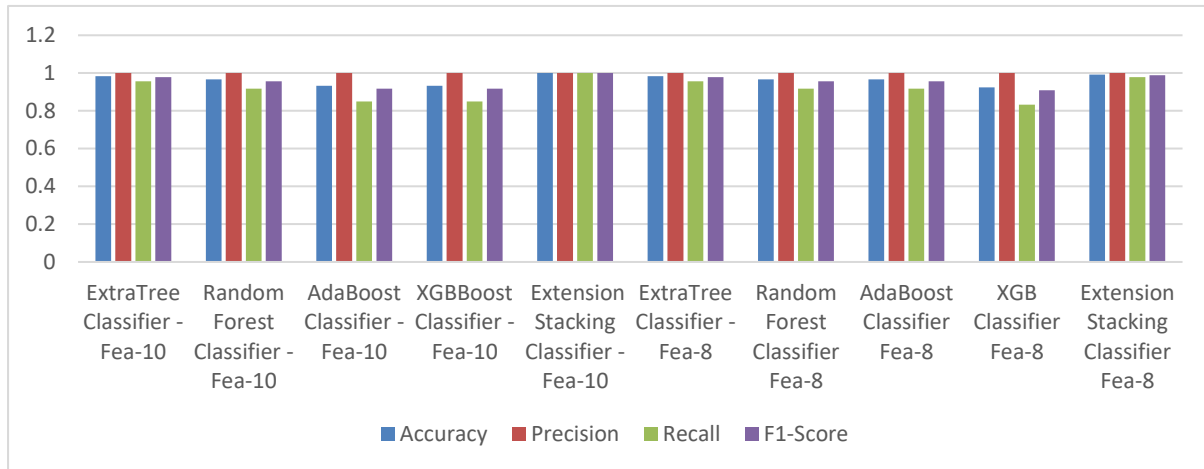
$$F1\ Score = 2 * \frac{Recall * Precision}{Recall + Precision} * 100 \quad (6)$$

Table (1) evaluate the performance metrics—Accuracy, precision, recall and F1-score—for each algorithm. The extension stacking classifier consistently outperforms compared to all other algorithms. The tables also offer a comparative analysis of the metrics for the other algorithms

Table.1 Performance Evaluation Table

Algorithms Used	Accuracy	Precision	Recall	F1-Score
ExtraTree Classifier - Fea-10	0.983	1.0	0.957	0.978
Random Forest Classifier - Fea-10	0.967	1.0	0.918	0.957
AdaBoost Classifier - Fea-10	0.933	1.0	0.849	0.918
XGBBoost Classifier - Fea-10	0.933	1.0	0.849	0.918
Extension Stacking Classifier - Fea-10	1.000	1.0	1.000	1.000
ExtraTree Classifier - Fea-8	0.983	1.0	0.957	0.978
Random Forest Classifier Fea-8	0.967	1.0	0.918	0.957
AdaBoost Classifier Fea-8	0.967	1.0	0.918	0.957
XGB Classifier Fea-8	0.925	1.0	0.833	0.909
Extension Stacking Classifier Fea-8	0.992	1.0	0.978	0.989

Graph.1 Comparison Graphs



Accuracy is represented in blue, precision in red, recall in green and F1 - Score in purple Graph (1). In comparison to the other models, the extension stacking classifier shows superior performance across all metrics, achieving the highest values. The graphs above visually illustrate these findings.

V. CONCLUSION

The project successfully developed an AI model tailored for early CKD diagnosis that achieved a high level of accuracy. This model, utilizing three critical features—hemoglobin, specific gravity, and hypertension—was able to effectively predict the likelihood of CKD [1, 2, 4, 5, 6] at an early stage. This emphasis on accuracy ensures reliable predictions, crucial for early intervention and treatment planning. Within the developed model, hemoglobin emerged as the most influential predictor for CKD diagnosis, showcasing its significant impact on identifying the disease at an early stage. Following hemoglobin, specific gravity and hypertension also played pivotal roles in predicting CKD [1, 2]. An optimization framework was employed to strike a balance between accuracy and explainability. This ensured that while the model delivered high accuracy, it remained comprehensible for clinicians, meeting their need for understandable model outputs. The project's emphasis on feature selection and explainability analysis holds promise in reducing costs associated with early CKD diagnosis [38], especially in resource-limited regions. By identifying and prioritizing key features crucial for accurate diagnosis, unnecessary tests or evaluations might be avoided, leading to cost-effective and targeted diagnostic procedures. Additionally, the focus on explainability ensures that the diagnosis process remains transparent and easily interpretable,

potentially reducing unnecessary and costly procedures or treatments.

External validation involves assessing the model's performance on independent datasets not used during its training. In the context of early Chronic Kidney Disease (CKD) diagnosis [4], this step is crucial to ensure that the model can generalize well to diverse patient populations. By testing the model on different datasets with similar features, researchers can evaluate its robustness and effectiveness across a broader range of cases, enhancing confidence in its real-world applicability. Transparency and trustworthiness are essential aspects, especially in medical applications where decisions impact patient well-being. Enhancing model transparency involves making the decision-making process more understandable for clinicians. Trustworthiness ensures that the model's predictions align with medical knowledge and intuition. Further research in this area may involve employing interpretable machine learning models, refining model explanations, or incorporating feedback from clinicians to improve overall trust and acceptance of the model in clinical practice. Partial Dependence Plot (PDP) and SHapley Additive exPlanations (SHAP) [51] are advanced explainability techniques that offer deeper insights into how individual features influence model predictions. PDP illustrates the relationship between a feature and the model's output, while SHAP values provide a way to allocate contributions of each feature to the prediction. By incorporating these techniques, researchers can provide a more nuanced understanding of the model's decision logic. This not only aids in model interpretation but also helps in identifying potential biases or unexpected dependencies, promoting a more thorough evaluation of the model's behavior and fostering trust among stakeholders, including clinicians.

REFERENCES

- [1] C. P. Kovesdy, "Epidemiology of chronic kidney disease: An update 2022," *Kidney Int. Supplements*, vol. 12, no. 1, pp. 7–11, Apr. 2022, doi: 10.1016/j.kisu.2021.11.003.
- [2] R. Gupta, N. Koli, N. Mahor, and N. Tejashri, "Performance analysis of machine learning classifier for predicting chronic kidney disease," in *Proc. Int. Conf. Emerg. Technol. (INCET)*, Jun. 2020, pp. 1–4, doi: 10.1109/INCET49848.2020.9154147.
- [3] World Health Organization. (2019). World Health Statistics 2019: Monitoring Health for the SDGs, Sustainable Development Goals. Accessed: Feb. 7, 2023. [Online]. Available: <https://apps.who.int/iris/handle/10665/324835>
- [4] P. Cockwell and L.-A. Fisher, "The global burden of chronic kidney disease," *Lancet*, vol. 395, no. 10225, pp. 662–664, Feb. 2020, doi: 10.1016/S0140-6736(19)32977-0.
- [5] A. C. Webster, E. V. Nagler, R. L. Morton, and P. Masson, "Chronic kidney disease," *Lancet*, vol. 389, no. 10075, pp. 1238–1252, Mar. 2017, doi: 10.1016/S0140-6736(16)32064-5.
- [6] H. Polat, H. D. Mehr, and A. Cetin, "Diagnosis of chronic kidney disease based on support vector machine by feature selection methods," *J. Med. Syst.*, vol. 41, p. 55, Apr. 2017, doi: 10.1007/s10916-017-0703-x.
- [7] R. A. Jeewantha, M. N. Halgamuge, A. Mohammad, and G. Ekici, "Classification performance analysis in medical science: Using kidney disease data," in *Proc. Int. Conf. Big Data Res.*, Osaka, Japan, 2017, pp. 1–6, doi: 10.1145/3152723.3152724.
- [8] N. Lei, X. Zhang, M. Wei, B. Lao, X. Xu, M. Zhang, H. Chen, Y. Xu, B. Xia, D. Zhang, C. Dong, L. Fu, F. Tang, and Y. Wu, "Machine learning algorithms' accuracy in predicting kidney disease progression: A systematic review and meta-analysis," *BMC Med. Informat. Decis. Making*, vol. 22, no. 1, p. 205, Aug. 2022, doi: 10.1186/s12911-022-01951-1.
- [9] J. Qezelbash-Chamak, S. Badamchizadeh, K. Eshghi, and Y. Asadi, "A survey of machine learning in kidney disease diagnosis," *Mach. Learn. Appl.*, vol. 10, Dec. 2022, Art. no. 100418, doi: 10.1016/j.mlwa.2022.100418.
- [10] S. A. Ebiaredoh-Mienye, T. G. Swart, E. Esenogho, and I. D. Mienye, "A machine learning method with filter-based feature selection for improved prediction of chronic kidney disease," *Bioengineering*, vol. 9, no. 8, p. 350, Jul. 2022, doi: 10.3390/bioengineering9080350.
- [11] M. A. Abdel-Fattah, N. A. Othman, and N. Goher, "Predicting chronic kidney disease using hybrid machine learning based on Apache Spark," *Comput. Intell. Neurosci.*, vol. 2022, pp. 1–12, Feb. 2022, doi: 10.1155/2022/9898831.
- [12] G. Stiglic, P. Kocbek, N. Fijacko, M. Zitnik, K. Verbert, and L. Cilar, "Interpretability of machine learning-based prediction models in healthcare," *WIREs Data Mining Knowl. Discovery*, vol. 10, no. 5, p. e1379, Sep. 2020, doi: 10.1002/widm.1379.
- [13] E. Tjoa and C. Guan, "A survey on explainable artificial intelligence (XAI): Toward medical XAI," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 11, pp. 1–21, Nov. 2021, doi: 10.1109/TNNLS.2020.3027314.
- [14] H. Zhang, J.-X. Ren, J.-X. Ma, and L. Ding, "Development of an in silico prediction model for chemical-induced urinary tract toxicity by using naïve Bayes classifier," *Mol. Diversity*, vol. 23, no. 2, pp. 381–392, May 2019, doi: 10.1007/s11030-018-9882-8.
- [15] S. S. Alaoui, B. Aksasse, and Y. Farhaoui, "Data mining and machine learning approaches and technologies for diagnosing diabetes in women," in *Big Data and Networks Technologies*. Cham, Switzerland: Springer, 2020, pp. 59–72, doi: 10.1007/978-3-030-23672-4_6.
- [16] Y. Zhang and Y. Ma, "Application of supervised machine learning algorithms in the classification of sagittal gait patterns of cerebral palsy children with spastic diplegia," *Comput. Biol. Med.*, vol. 106, pp. 33–39, Mar. 2019, doi: 10.1016/j.combiomed.2019.01.009.
- [17] A. K. Feeny, J. Rickard, D. Patel, S. Toro, K. M. Trulock, C. J. Park, M. A. LaBarbera, N. Varma, M. J. Niebauer, S. Sinha, E. Z. Gorodeski, R. A. Grimm, X. Ji, J. Barnard, A. Madabhushi, D. D. Spragg, and M. K. Chung, "Machine learning prediction of response to cardiac resynchronization therapy: Improvement versus current guidelines," *Circulat., Arrhythmia Electrophysiol.*, vol. 12, no. 7, Jul. 2019, Art. no. e007316, doi: 10.1161/CIRCEP.119.007316.
- [18] T. O. Aro, H. B. Akande, M. B. Jibrin, and U. A. Jauro, "Homogenous ensembles on data mining techniques for breast cancer diagnosis," *Daffodil Int. Univ. J. Sci. Technol.*, vol. 14, no. 1, pp. 1–10, 2019.
- [19] S. Karun, A. Raj, and G. Attigeri, "Comparative analysis of prediction algorithms for diabetes," in *Advances in Computer Communication and Computational Sciences*. Singapore: Springer, 2019, pp. 177–187, doi: 10.1007/978-981-13-0341-8_16.
- [20] M. Bucholc, X. Ding, H. Wang, D. H. Glass, H. Wang, G. Prasad, L. P. Maguire, A. J. Bjourson, P. L. McClean, S. Todd, D. P. Finn, and K. Wong-Lin, "A practical computerized decision support system for predicting the severity of Alzheimer's disease of an individual," *Expert Syst. Appl.*, vol. 130, pp. 157–171, Sep. 2019, doi: 10.1016/j.eswa.2019.04.022.
- [21] G. S. Collins, J. B. Reitsma, D. G. Altman, and K. G. M. Moons, "Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): The TRIPOD statement," *Ann. Internal Med.*, vol. 162, no. 1, pp. 55–63, Jan. 2015, doi: 10.7326/M14-0697.
- [22] A. Ray and A. K. Chaudhuri, "Smart healthcare disease diagnosis and patient management: Innovation, improvement and skill development," *Mach. Learn. Appl.*, vol. 3, Mar. 2021, Art. no. 100011, doi: 10.1016/j.mlwa.2020.100011.
- [23] D. Dua and C. Graff. (2017). UCI Machine Learning Repository. University of California, Irvine, School of Information and Computer Sciences. [Online]. Available: <http://archive.ics.uci.edu/ml>
- [24] I. U. Ekanayake and D. Herath, "Chronic kidney disease prediction using machine learning methods," in *Proc. Moratuwa Eng. Res. Conf. (MERCCon)*, Jul. 2020, pp. 260–265, doi: 10.1109/MERCCon50084.2020.9185249.
- [25] S. S. Alaoui, B. Aksasse, and Y. Farhaoui, "Statistical and predictive analytics of chronic kidney disease," in *Advanced Intelligent Systems for Sustainable Development*. Cham, Switzerland: Springer, 2019, pp. 27–38, doi: 10.1007/978-3-030-11884-6_3.
- [26] A. Ogunleye and Q.-G. Wang, "XGBoost model for chronic kidney disease diagnosis," *IEEE/ACM Trans. Comput. Biol. Bioinf.*, vol. 17, no. 6, pp. 2131–2140, Nov. 2020, doi: 10.1109/TCBB.2019.2911071.
- [27] S. Zeynu, "Prediction of chronic kidney disease using data mining feature selection and ensemble method," *Int. J. Data Mining Genomics Proteomics*, vol. 9, no. 1, pp. 1–9, 2018.
- [28] N. V. G. Raju, K. P. Lakshmi, K. G. Prashritha, and C. Likhitha, "Prediction of chronic kidney disease (CKD) using data science," in *Proc. Int. Conf. Intell. Comput. Control Syst. (ICCS)*, May 2019, pp. 642–647, doi: 10.1109/ICCS45141.2019.9065309.
- [29] S. Imran Ali, B. Ali, J. Hussain, M. Hussain, F. A. Satti, G. H. Park, and S. Lee, "Cost-sensitive ensemble feature ranking and automatic threshold selection for chronic kidney disease diagnosis," *Appl. Sci.*, vol. 10, no. 16, p. 5663, Aug. 2020, doi: 10.3390/app10165663.
- [30] B. Khan, R. Naseem, F. Muhammad, G. Abbas, and S. Kim, "An empirical evaluation of machine learning techniques for chronic kidney disease prophecy," *IEEE Access*, vol. 8, pp. 55012–55022, 2020, doi: 10.1109/ACCESS.2020.2981689.
- [31] K. M. Z. Hasan and M. Z. Hasan, "Performance evaluation of ensemblebased machine learning techniques for prediction of chronic kidney disease," in *Emerging Research in Computing, Information, Communication and Applications*. Singapore:

Springer, 2019, pp. 415–426, doi: 10.1007/978-981-13-5953-8_34.

[32] L. Antony, S. Azam, E. Ignatiou, R. Quadir, A. R. Beeravolu, M. Jonkman, and F. De Boer, “A comprehensive unsupervised framework for chronic kidney disease prediction,” *IEEE Access*, vol. 9, pp. 126481–126501, 2021, doi: 10.1109/ACCESS.2021.3109168.

[33] A. K. Chaudhuri, D. Sinha, D. K. Banerjee, and A. Das, “A novel enhanced decision tree model for detecting chronic kidney disease,” *Netw. Model. Anal. Health Informat. Bioinf.*, vol. 10, no. 1, p. 29, Apr. 2021, doi: 10.1007/s13721-021-00302-w.

[34] A. A. Abdullah, S. A. Hafidz, and W. Khairunizam, “Performance comparison of machine learning algorithms for classification of chronic kidney disease (CKD),” *J. Phys., Conf. Ser.*, vol. 1529, no. 5, May 2020, Art. no. 052077, doi: 10.1088/1742-6596/1529/5/052077.

[35] R. C. Poonia, M. K. Gupta, I. Abunadi, A. A. Albraikan, F. N. Al-Wesabi, and M. A. Hamza, “Intelligent diagnostic prediction and classification models for detection of kidney disease,” *Healthcare*, vol. 10, no. 2, p. 371, Feb. 2022, doi: 10.3390/healthcare10020371.

[36] M. Siddhartha, V. Kumar, and R. Nath, “Early-stage diagnosis of chronic kidney disease using majority vote—Grey wolf optimization (MV-GWO),” *Health Technol.*, vol. 12, no. 1, pp. 117–136, Jan. 2022, doi: 10.1007/s12553-021-00617-8.

[37] A. Alaiad, H. Najadat, B. Mohsen, and K. Balhaf, “Classification and association rule mining technique for predicting chronic kidney disease,” *J. Inf. Knowl. Manage.*, vol. 19, no. 1, Mar. 2020, Art. no. 2040015, doi: 10.1142/S0219649220400158.

[38] M. Kadhum, S. Manaseer, and A. L. A. Dalhoum, “Evaluation feature selection technique on classification by using evolutionary ELM wrapper method with features priorities,” *J. Adv. Inf. Technol.*, vol. 12, no. 1, pp. 21–28, 2021, doi: 10.12720/jait.12.1.21-28.

[39] S. Akter, A. Habib, M. A. Islam, M. S. Hossen, W. A. Fahim, P. R. Sarkar, and M. Ahmed, “Comprehensive performance assessment of deep learning models in early prediction and risk identification of chronic kidney disease,” *IEEE Access*, vol. 9, pp. 165184–165206, 2021, doi: 10.1109/ACCESS.2021.3129491.

[40] P. Theerthagiri and A. U. Ruby, “RFFS: Recursive random forest feature selection based ensemble algorithm for chronic kidney disease prediction,” *Expert Syst.*, vol. 39, no. 9, pp. 1–12, Nov. 2022, doi: 10.1111/exsy.13048.

[41] S. I. Ali, H. S. M. Bilal, M. Hussain, J. Hussain, F. A. Satti, M. Hussain, G. H. Park, T. Chung, and S. Lee, “Ensemble feature ranking for cost-based non-overlapping groups: A case study of chronic kidney disease diagnosis in developing countries,” *IEEE Access*, vol. 8, pp. 215623–215648, 2020, doi: 10.1109/ACCESS.2020.3040650.

[42] P. A. Moreno-Sanchez, “An automated feature selection and classification pipeline to improve explainability of clinical prediction models,” in *Proc. IEEE 9th Int. Conf. Healthcare Informat. (ICHI)*, Aug. 2021, pp. 527–534, doi: 10.1109/ICHI52183.2021.00100.

[43] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, and M. Blonde, “Scikit-learn: Machine learning in Python,” *J. Mach. Learn. Res.*, vol. 12, pp. 2825–2830, Nov. 2011.

[44] S. Shilaskar and A. Ghatol, “Feature selection for medical diagnosis: Evaluation for cardiovascular diseases,” *Expert Syst. Appl.*, vol. 40, no. 10, pp. 4146–4153, Aug. 2013, doi: 10.1016/j.eswa.2013.01.032.

[45] J. Tang, S. Alelyani, and H. Liu, “Feature selection for classification: A review,” in *Data Classification*. Boca Raton, FL, USA: CRC Press, 2014, pp. 37–64, doi: 10.1201/b17320.

[46] O. Sagi and L. Rokach, “Ensemble learning: A survey,” *WIREs Data Mining Knowl. Discovery*, vol. 8, no. 4, pp. 1–15, Jul. 2018, doi: 10.1002/widm.1249.

[47] O. Sagi and L. Rokach, “Explainable decision forest: Transforming a decision forest into an interpretable tree,” *Inf. Fusion*, vol. 61, pp. 124–138, Sep. 2020, doi: 10.1016/j.inffus.2020.03.013.

[48] A. B. Arrieta, N. Díaz-Rodríguez, J. D. Ser, A. Bennetot, S. Tabik, A. Barbado, S. Garcia, S. Gil-Lopez, D. Molina, R. Benjamins, R. Chatila, and F. Herrera, “Explainable artificial intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI,” *Inf. Fusion*, vol. 58, pp. 82–115, Jun. 2020.

[49] A. Fisher, C. Rudin, and F. Dominici, “All models are wrong, but many are useful: Learning a variable’s importance by studying an entire class of prediction models simultaneously,” 2018, arXiv:1801.01489.

[50] J. H. Friedman, “Greedy function approximation: A gradient boosting machine,” *Ann. Statist.*, vol. 29, no. 5, pp. 1189–1232, Oct. 2001.

[51] S. Lundberg and S.-I. Lee, “A unified approach to interpreting model predictions,” 2017, arXiv:1705.07874.