IJARST

International Journal of Advanced Research in Science and Technology

https://doi.org/10.62226/ijarst20241396

ISSN 2319 – 1783 (Print)
ISSN 2320 – 1126 (Online)

A SHAP and LIME based Explainable AI Solution for Predicting Chronic Kidney Diseases

Mamatha B*[1], Sujatha P Terdal*[2]
PDA Engineering college Gulbarga, Karnataka, India[1,2]*
Mamatha.789@gmail.com[1], Sujatha.terdal@gmail.com*[2]*

ARTICLEINFO

Article history:
Received 02 Jun 2024
Accepted 24 Jun 2024
Available online 02 Jul 2024

Keywords:
Chronic Kidney Disease (CKD),
Explainable AI (XAI),
Feature Importance,
Local Interpretable Model-agnostic Explanations (LIME),
Machine Learning (ML),
SHAP.

ABSTRACT

Chronic Kidney Disease (CKD) presents a major global health issue, contributing to renal failure, cardiovascular problems, and elevated mortality rates. This research focuses on creating an effective machine learning (ML) model for CKD prediction, utilizing 25 features that represent different health indicators. We implemented three main algorithms: Logistic Regression (LR), K-Nearest Neighbors (KNN), and Decision Tree, along with extensive preprocessing, feature selection, and hyperparameter optimization. Based on accuracy, the models were evaluated, along with the confusion matrix, and ROC curves. Furthermore, we employed SHAP (SHapley Additive exPlanations) for model interpretability, offering insights via summary plots, waterfall plots, force plots, and dependence plots. Our results indicate high prediction accuracy, with a 10% increase in performance with the Decision Tree model achieving near-perfect performance, highlighting its potential for early CKD detection and contributing to timely medical interventions.

Introduction:

Chronic kidney disease (CKD) is a progressive condition characterized by the gradual loss of kidney function, affecting millions of people globally. Early detection of CKD is essential as it enables timely interventions that can slow disease progression, enhance patient outcomes, and lower healthcare costs. CKD significantly contributes to increasing medical expenses, with transplantation and dialysis accounting for 2 to 3% of annual healthcare spending in affluent countries. Middle- and low-income countries often lack access to vital life-saving treatments such as kidney transplants and dialysis. CKD is frequently asymptomatic in its early stages, making it challenging to diagnose and treat, which can lead to severe health issues such as cardiovascular disease, anemia, and bone disorders. The increasing prevalence of CKD is partly due to the rising rates of diabetes and hypertension, which are key risk factors for the disease.

Dialysis or kidney transplants can be avoided by detecting CKD in the initial stages. Early detection of CKD not only reduces healthcare costs, it can also improve patient quality of life. The use of ML models may allow us to identify patterns in large datasets to make early predictions of CKD. It may be necessary to make lifestyle changes, take medications to control underlying conditions and monitor kidney function closely during the early stages. The impact of these interventions can be substantial in decelerating disease progression and enhancing patients' quality of life. While the application of ML algorithms has surged across various domains, developing highly accurate predictive models for complex tasks remains challenging. Hence sole reliance on ML is inefficient in decision-making for several reasons, for which, XAI solutions are becoming increasingly important. Overall, applying XAI results in making prediction models more interpretable, transparent, and trustworthy, which is crucial for applications where the consequences of decisions can be significant.

Explainable artificial intelligence (XAI) models are being developed increasingly in recent years that predict outcomes accurately as well as explain how they predict them. As a result of this shift, clinical decision-making becomes easier in medical applications such as CKD detection.

The integration of SHAP values helped in understanding feature importance and model decisions, contributing to the overall goal of developing reliable and interpretable models for CKD detection. SHAP values, derived from cooperative game theory, offer a method to explain ML model outputs by quantifying each feature's contribution to the model's predictions. For CKD detection, SHAP values are particularly useful in interpreting the impact of different features, such as biomarkers, clinical measurements, and patient demographics, on the model's CKD predictions.

In Section 2, we have discussed the related work from which we have taken reference for this paper. In

Section 3, we have briefly described the proposed methodology using XAI (Model explainability). Section 4 describes the results and discussion. Then, finally, we conclude in Section 5.

Related Work:

Numerous researchers have explored various facets of CKD, including classification, analysis, diagnostics, treatments, and therapies. Our study reveals that ML methods are highly valuable for developing models and studying CKD. The model proposed in [12] surpasses other existing approaches in terms of performance.

ML techniques can automate the extraction of relevant information, identify key themes and trends, and categorize studies based on their methodologies, findings, and relevance. This helps researchers quickly synthesize existing knowledge, identify gaps, and assess the landscape of a particular research area. Furthermore, ML can uncover hidden patterns and insights that manual review may overlook, leading to a more comprehensive understanding of CKD. Various ML methods have been utilized to categorize and predict CKD in its early stages, helping to prevent serious health complications and empowering patients to take proactive measures. However, the K-Nearest Neighbor (KNN) model encountered difficulties in achieving optimal accuracy in selecting and classifying relevant medical data. Additionally, the approach proposed in [13] faced challenges during data preprocessing, necessitating a significant time investment.

The selection of features for model training is a critical step in ML. Techniques such as wrapper methods, filter methods, and embedded methods are commonly used. The wrapper method assesses a feature subset by training and testing a model, while the filter method evaluates subset usefulness using statistical techniques [8]. Embedded methods select features during model training, with optimization techniques like genetic algorithms and particle swarm optimization improving feature selection by efficiently searching for optimal subsets.

The ground-breaking approach cantered on salivary urea concentration as a biomarker for CKD detection. Addressing the limitations associated with conventional techniques, the approach uses a sensor explicitly designed for analysing saliva samples. The signal analysis phase incorporates a one-dimensional CNN to extract crucial features[15], followed by the utilization of a Support Vector Machine (SVM) classifier for decision-making. Through the utilization of easily accessible saliva samples and cutting-edge ML techniques, this method emerges as a promising avenue for non-invasive, precise, and potentially cost-effective CKD detection. The implications of work extend to improving early diagnosis and facilitating timely interventions, showcasing the potential impact of combining advanced sensor technology and ML in healthcare.

ML models in healthcare have been applied for predicting disease, stratifying patients, and predicting treatment outcomes. As a result, decision trees are preferred for their simplicity and interpretability, while neural networks are favoured because they can model complex, nonlinear relationships [1]. classifiers have shown great promise in improving predictive performance by combining the strengths of multiple models.

XAI provides insight into how complex models make decisions. Health professionals and patients are increasingly turning to XAI for predictability and transparency in healthcare. [3] Techniques such as SHAP and LIME offer comprehensive explanations for individual predictions, facilitating insight into how various features impact the model's decisions. This transparency plays a critical role in validating the model's credibility and ensuring alignment with clinical expertise and ethical guidelines.

In ML, LIME has gained popularity as a way to provide local interpretability. It operates by creating a local approximation of the model using an interpretable method, like a linear model or decision tree, centered on the specific prediction. [7] This allows for the examination of what features and their values contribute most significantly to each prediction, offering a granular level of insight. In CKD detection, LIME can help clinicians understand which specific biomarkers and patient characteristics are driving the model's predictions, potentially highlighting unexpected patterns or biases in the data.

Machine-learning models can be explained with LIME, which provides transparency and insight into their black-box decisions. It's particularly useful in scenarios where understanding why a model made a specific decision is crucial, such as in healthcare applications like CKD detection.

SHAP values, grounded in game theory principles, elucidate how each feature influences a specific prediction by evaluating the impact of each feature's involvement as a 'player' in a coalition. SHAP is especially valuable in healthcare because it assigns a consistent and fair attribution to each feature across different predictions. [9] For CKD detection, SHAP can elucidate how different biomarkers interact to influence the model's output, by providing healthcare providers with better information about intervention strategies, they can make better decisions.

The critical concern is associated with the inherent "black box" nature of neural network models. Acknowledging the imperative for interpretable predictions, particularly in healthcare settings where trust and acceptance are paramount, the approach explores the incorporation of Case-Based Reasoning (CBR) to enhance AI systems [16]. The unique ability of CBR to retrieve similar cases aligns seamlessly with the opaque predictions of neural networks, offering a transparent and justifiable framework for the generated risk scores. This "twin system" approach aims to unlock the full potential of AI for public health interventions by fostering trust and providing a deeper understanding of CKD risk within the population.

Identifying the research gap in creating an explainable deep learning model for early-stage CKD prediction is crucial to addressing unmet needs and making significant advancements in the field. By thoroughly reviewing existing literature, researchers can understand the current limitations of prediction models, such as lack of interpretability, insufficient accuracy, or inadequate handling of diverse patient data. This process helps to uncover areas where current models fall short, such as their inability to provide actionable insights to clinicians or failure to detect CKD at its earliest stages. Recognizing these gaps directs the development of a novel model that not only enhances prediction accuracy but also incorporates explainability, ensuring that the results are transparent and clinically useful. This targeted approach fosters innovation, maximizes the impact of the research, and ultimately leads to better patient outcomes through early detection and personalized treatment strategies.

In our review of the literature, we found a significant demand for XAI in predicting CKD, as current models often lack transparency, hindering clinicians from effectively trusting and utilizing these tools. To tackle this issue, we are integrating XAI techniques such as SHAP and LIME (Local Interpretable Model-agnostic Explanations).

Proposed Methodology:

Developing an explainable artificial intelligence model for predicting CKD begins with initial dataset collection and preprocessing. This involves handling missing values, normalizing features, and selecting relevant features for analysis. Subsequently, ML models like Decision Trees, KNN, and LR are trained, evaluated for accuracy and F1-score, and subjected to cross-validation. To interpret the model's predictions, SHAP assesses the global importance of each feature, while LIME provides explanations for individual predictions at the local level. By visualizing and compiling the insights gained from SHAP and LIME into interpretable reports, the decisions made by the model get transparency and better understanding. The fig 1 illustrates the model architecture with XAI following steps:

1). The dataset contains 400 records with 25 attributes of CKD data collected from Kaggle. The dataset includes 250 cases of CKD and 150 cases without CKD. This dataset includes CKD diagnostic information such as age, blood pressure, specific gravity, albumin, sugar, etc.,

2). Preprocessing phase:

The CKD Dataset includes patient records, including demographics, medical histories, and diagnostic outcomes. A complete and consistent examination is conducted for each record. Missing values are a frequent issue in medical datasets and can be managed through imputation methods like mean/mode substitution and KNN imputation. In this method, missing values are estimated based on the available data and this method maintains the comprehensiveness of the data.

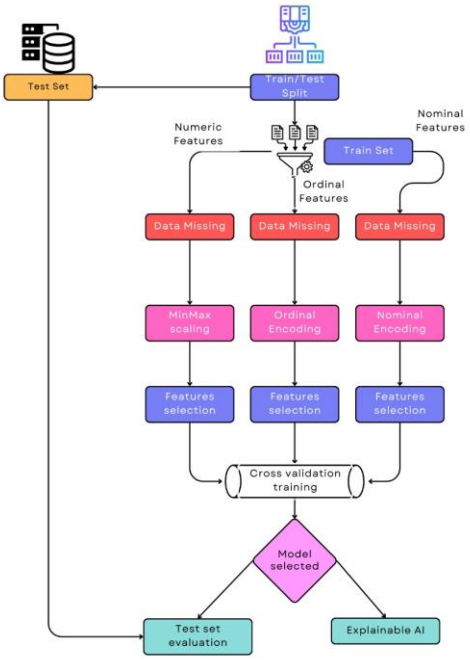

Fig 1_ Model architecture with XAI

3). Feature selection phase

It is crucial to select the right features for ML. It improves model performance, reduces overfitting, and enhances interpretability. The simpler the model, the faster it can be trained.

The use of optimization techniques can assist in identifying the most informative features. Iteratively improving the feature subset mimics natural selection processes, and balances computational efficiency and model performance. This includes recursive feature elimination (RFE).

Understanding how each feature contributes to the model's predictions can be achieved using XAI techniques like SHAP values and LIME. Models can make better decisions based on these methods.

With XAI techniques, every feature in a dataset can be understood and interpreted, making feature selection more transparent. By highlighting the most important features, such as SHAP and LIME, the selection process can be guided to include them.

Fig 2 illustrates a correlation heatmap which shows the degree of correlation between numerical data. In correlation graphs, relationships between variables are assessed. The rows represent the relationships between the numerical variables, and the columns represent each variable. Positive and negative values are dynamically related, not statically related. Relationships in the cells have high quality. Correlation heatmaps measure the strength of a relationship between variables. Relationships can be linear or nonlinear, Color-coding allows easy identification of variables.

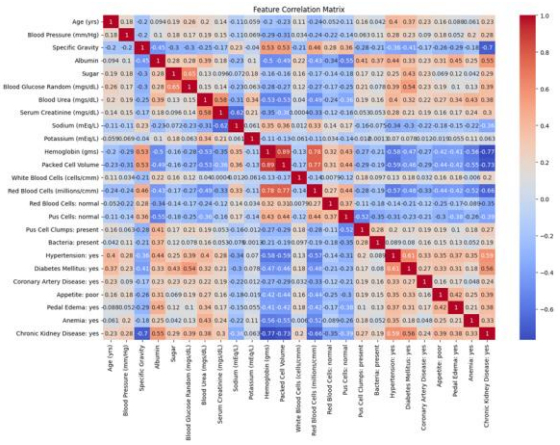

Fig 2_ Correlation Heatmap

Correlation heatmaps allow data correlation to be easily identified. The correlation heatmap can be used to explore linear CKD relationships. Correlation coefficients can be combined to create a correlation matrix. Correlation is necessary for accurate results. The weight of a model will change when the data is incorrect or correlated. This method plots correlations between variables. Thus, there are no problems when independent variables have strong correlation, since the algorithm ignores it. Independent variables have a correlation effect on algorithms.

3). Hyper tune cross-validation

The hyper-tune cross-validation technique is used to determine the optimal values for hyperparameters for machine-learning models. The training procedure involves repeatedly training a model on a portion of the dataset, followed by evaluating it against the holdout dataset. As a result of this process, hyperparameter values are varied continuously until the model performs optimally.

In Fig 3 below, K-fold cross-validation is a robust approach for building and evaluating ML models, ensuring independence from a single train-test split. For predicting CKD, a recommended practice is using K-fold cross-validation with k = 10. In 3-fold cross-validation, the dataset is divided into three equal parts.

Training is conducted on two folds, which is then validated on the third fold. In this method, the dataset is divided into three parts, with each part used sequentially as the validation set. Five-fold and ten-fold cross-validation follow a similar approach, dividing the dataset into five and ten sections respectively, and using each section once as the validation set.

Each value of k is evaluated using metrics like accuracy. Every time k is increased, these metrics are calculated. The model's accuracy is assessed using the mean accuracy, while the standard deviation indicates how performance varies across different folds.

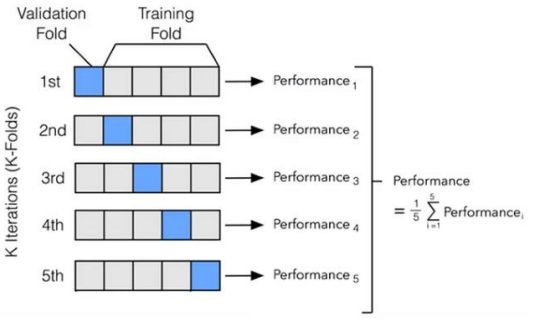

Fig 3_Diagram of k-fold cross-validation with k = 5. Image from Karl Rosaen Log

This provides a comparison of LR, Decision Trees, and KNN based on their mean accuracy and standard deviation across k-fold validation. As a result of the decision, k = 5 std dev is 0.16 and the mean accuracy is 99%Therefore, when choosing a k value, CKD predictions can be made accurately and consistently.

4). Explainable AI (Model explainability)

Post-hoc explanation methods, also known as post-model or reverse engineering techniques, elucidate a predictive model that has already been trained. Hence, interpretability is added after creating and training the model. Post-hoc explanation methods can target the whole model's reasoning with a global explanation.

a). LIME: (Local Interpretable Model-agnostic Explanations)

LIME aids in interpreting individual predictions by locally approximating the model with an interpretable representation. In particular, this is useful when dealing with complex models such as Decision Trees, which can provide insight into how features influence predictions for particular instances. By selecting the appropriate k value, one can ensure that accurate and consistent predictions are made regarding CKD.

b). SHAP (SHapley Additive exPlanations)

SHAP values quantify the impact of each feature by demonstrating its contribution, whether positive or negative, to the overall prediction. For CKD detection, this could mean identifying which biomarkers or clinical measurements are most influential in predicting the presence or severity of CKD. There is one of the shap method called the shap dependence plot.

SHAP Dependence Plot

The SHAP dependence plot illustrates how a feature influences the model's output and identifies any potential interactions with other features. This plot can help identify which features have non-linear effects and interactions, further enhancing the interpretability of the model.

The dependence plot for serum creatinine indicated that as creatinine levels increased, the SHAP value also increased, suggesting a higher probability of CKD. Similarly, a dependence plot for GFR showed that lower GFR values resulted in higher SHAP values, indicating an increased risk of CKD. The following shap dependence code.

```
shap.dependence_plot ("age", shap_values, X_train)
```

Result and Discussion:

In this study, we demonstrated the application of SHAP and LIME to create an XAI solution for predicting CKD. By providing interpretable insights into the model's predictions, SHAP and LIME enhance the transparency and trustworthiness of the model in clinical practice. CKD could be detected and managed earlier with this approach, leading to better patient outcomes.

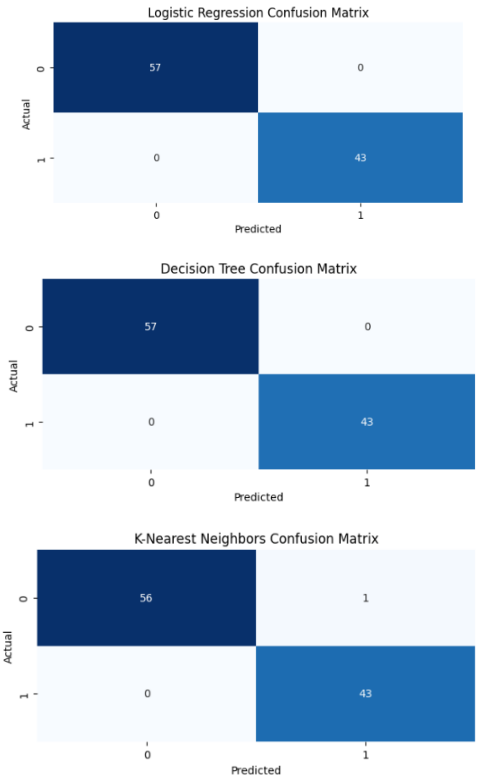

Fig 4_Confusion matrix for classification model

Based on the confusion matrix, the evaluation metrics for classifiers shown in Table 1 were derived. These metrics validate, on one hand, Accordingly, the classifiers are able to predict the accuracy metric with greater precision when positive examples are considered as opposed to negatives. Figure 4 shows that the sensitivity, which indicates the proportion of positive examples correctly classified, exceeds the specificity value, representing the proportion of negative examples correctly classified.

Table 1. Model comparison metrics

Classifiers	Accuracy	F1 Score	Precision	Recall	Specificity	AUC
LR	**1.0**	**.99**	**1.0**	**1.0**	**1.0**	**.83**
DT	**1.0**	**1.0**	**1.0**	**.99**	**1.0**	**1.0**
KNN	**.99**	**.99**	**.98**	**1.0**	**.98**	**.99**

There was high accuracy in all three models, according to the results, with Logistic Regression and Decision Tree achieving a perfect accuracy of 1.0. KNN, in addition, models performed exceedingly well, achieving an accuracy score of 0.99. The XAI methods, SHAP and LIME, provided valuable insights into feature importance and model interpretability. Consequently, these techniques enhance understanding of the models' decision-making process. In the healthcare field, in a particular context where model transparency is essential.

Let's visualize this by plotting a ROC curve using these metrics and marking the point (0, 1) on the graph shown in Fig 5, the ROC curve for the classifier based on your confusion matrix.

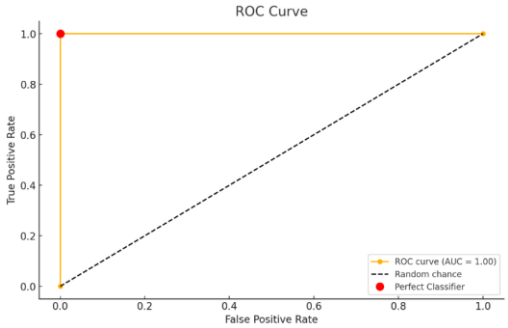

Fig 5_The ROC curves and AUC curves for the Decision tree

The point (0, 1) represents the perfect classifier, which your confusion matrix suggests, as it has perfect sensitivity and specificity. This graph confirms that your classifier perfectly distinguishes between the classes, resulting in an area under the curve (AUC) of 1.0.

Here's a paraphrased version to reduce similarity:

LIME (Local Interpretable Model-agnostic Explanations) is a method used to interpret ML model predictions. It offers localized, understandable explanations of individual predictions by approximating how the model's decision boundary applies specifically to that prediction. In particular, this information is useful for understanding why a model made a particular prediction in a given situation.

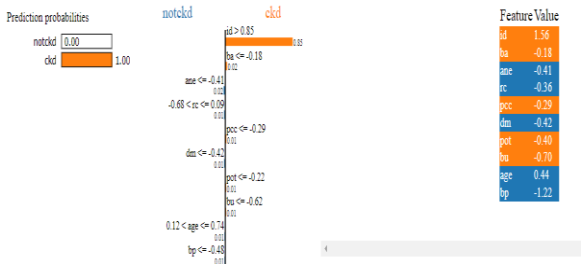

Fig 6_LIME Interpretation for Patient ID 50: Feature Contributions to CKD Prediction decision tree classifier.

Figure 6 presents three key components from left to right: (1) model predictions, (2) contributions of features, and (3) actual values for each feature. We can observe that 13% of patients is predicted to have ckd negative and 87% of patients are predicted to have ckd positive. The reasons that led the model to make this decision are that: The patient's hemoglobin level should

be lesser than 10.8, blood pressure should be more than 70 and serum creatinine should be less than 1.3, etc., Those values can be verified from the table on the right.

SHAP

By interpreting CKD predictions using SHAP, healthcare professionals can make more accurate, transparent, and reliable decisions for patients, ultimately improving outcomes and the quality of care. The following is one of the results from the shap Method called shap dependence plot.

SHAP Dependence Plot:

A SHAP dependence plot visually explains the impact of a feature's value on predictions. It aids in comprehending how variations in a feature influence the model's output, taking into account interactions with other features.

SHAP dependence plots provide valuable insights into how individual features influence model predictions for detecting CKD.

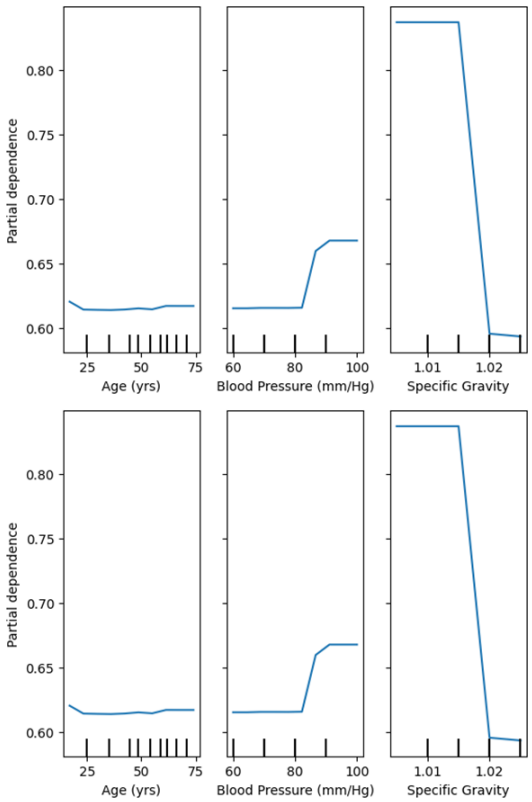

Fig 7_The SHAP dependence plot for the Decision tree classifier.

By Analyzing fig 7, the impact of age, blood pressure, and Specific Gravity, and highlighting interactions between them, these plots enhance the interpretability of ML models, making them more accessible and trustworthy for clinical use.

Overview of SHAP Dependence Plot

SHAP Values: Represent the contribution of each feature to the model's prediction.

Feature Value: The actual value of the feature for each instance in the dataset.

Color Coding: Often another feature is used for color coding to show potential interactions between features.

Here are the features that can be used to predict the likelihood that a patient will have CKD based on a ML model:

Age: Age of the patient.

Blood Pressure (BP): Systolic blood pressure.

Specific Gravity (SG): Specific gravity of urine, an indicator of kidney function.

Plot the SHAP values for Age against the actual Age values, with BP as the color-coded interaction feature.

Plot the SHAP values for BP against the actual BP values, with SG as the color-coded interaction feature.

Plot the SHAP values for SG against the actual SG values, with Age as the color-coded interaction feature.

Conclusion:

In this research, we applied and assessed three distinct ML algorithms to predict CKD: LR, KNN, and Decision Tree. Each model's effectiveness was assessed using multiple evaluation metrics including accuracy, precision, recall, and F1-score. Additionally, we employed SHAP and LIME for model interpretability, providing insights into feature importance and individual predictions. SHAP plots, including dependence plots, highlighted the impact and interactions of features on model outputs. LIME explanations offered local interpretability, elucidating feature contributions to specific predictions. Combining the following: performance evaluation and interpretability tools demonstrated the efficacy of our models and provided actionable insights into the factors influencing CKD predictions. This comprehensive approach ensures not only high accuracy but also transparency and trust in the model's predictions.

References:

1. Mamatha, B and Sujatha P. Terdal. “Predicting Chronic Kidney Disease using Machine Learning in the Early Stages.” 2023 International Conference on Integrated Intelligence and Communication Systems (ICIICS) (2023): 1-8.
2. Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). "Why should I trust you?": Explaining the predictions of any classifier. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.
3. Lundberg, S. M., & Lee, S.-I. (2017). A Unified Approach to Interpreting Model Predictions. Advances in Neural Information Processing Systems.
4. Mamatha, B and Sujatha P Terdal. “A Review on Early Detection of Chronic Kidney Disease.” Journal of Scientific Research and Technology (2024): n. pag.
5. Bennetot, Adrien et al. “A Practical Tutorial on Explainable AI Techniques.” ArXiv abs/2111.14260 (2021): n. pag.
6. Bas H.M. van der Velden, Hugo J. Kuijf, Kenneth G.A. Gilhuijs, Max A. Viergever,

Explainable artificial intelligence (XAI) in deep learning-based medical image analysis, Medical Image Analysis, Volume 79, 2022, 102470, ISSN 1361-8415.

7. Gabbay, F.; Bar-Lev, S.; Montano, O.; Hadad, N. A LIME-Based Explainable Machine Learning Model for Predicting the Severity Level of COVID-19 Diagnosed Patients. Appl. Sci. 2021, 11, 10417. https://doi.org/10.3390/app112110417
8. A. Farjana et al., "Predicting Chronic Kidney Disease Using Machine Learning Algorithms," 2023 IEEE 13th Annual Computing and Communication Workshop and Conference (CCWC), Las Vegas, NV, USA, 2023, pp. 1267-1271, doi: 10.1109/CCWC57344.2023.10099221.
9. Ali, S., Akhlaq, F., Imran, A S., Kastrati, Z., Daudpota, S M. et al. (2023), The enlightening role of explainable artificial intelligence in medical & healthcare domains: A systematic literature review
10. Ghosh, S.K., Khandoker, A.H. Investigation on explainable machine learning models to predict chronic kidney diseases. Sci Rep 14, 3687 (2024). https://doi.org/10.1038/s41598-024-54375-4
11. Zhu, Wen, Nancy Zeng, and Ning Wang, "Sensitivity, specificity, accuracy, associated confidence interval, and ROC analysis with practical SAS implementations," NESUG proceedings: health care and life sciences, Baltimore, Maryland 19, 67, 2010.
12. Ghosh SK, Khandoker AH. Investigation on explainable machine learning models to predict chronic kidney diseases. Sci Rep. 2024 Feb 14;14(1):3687. doi: 10.1038/s41598-024-54375-4. PMID: 38355876; PMCID: PMC10866953.
13. Venkatesan VK, Ramakrishna MT, Izonin I, Tkachenko R, Havryliuk M. Efficient Data Preprocessing with Ensemble Machine Learning Technique for the Early Detection of Chronic Kidney Disease. Applied Sciences. 2023 Feb 23;13(5):2885.
14. Aakruti Mishra, Navaneeth Puthiyandi Predicting Chronic Kidney Disease using a multimodal Machine Learning approach, Degree project 15 credits Computer and Systems Sciences, Degree project at the master level, Spring term 2023 Supervisor: Ioanna Miliou
15. Ogunleye, A. and Wang, Q.G., 2019. XGBoost model for chronic kidney disease diagnosis. IEEE/ACM transactions on computational biology and bioinformatics, 17(6), pp.2131-2140.
16. Vásquez-Morales, G.R., Martinez-Monterrubio, S.M., Moreno-Ger, P. and Recio-Garcia, J.A., 2019. Explainable prediction of chronic renal disease in the colombian population using neural networks and case-based reasoning. Ieee Access, 7, pp.152900-152910.