Chronic Kidney Disease Detection Augmented with Hybrid Explainable AI

Mohamed Desoky
Faculty of Computer Science and Engineering, Galala University, Egypt
mdi103213@gu.edu.eg

Nada Gamal ElAdl
Faculty of Computer Science and Engineering, Galala University, Egypt
ngi212202@gu.edu.eg

Tamer Abuhmed
College of Computing and Informatics, Sungkyunkwan University, South Korea
tamer@skku.edu

Shaker El-Sappagh
Faculty of Computer Science and Engineering, Galala University, Egypt
shaker.elsappagh@gu.edu.eg

2025 15th International Conference on Electrical Engineering (ICEENG) | 979-8-3315-1901-8/25/$31.00 ©2025 IEEE | DOI: 10.1109/ICEENG64546.2025.11031335

***Abstract*—This study proposes an explainable machine learning (ML)-based framework for early prediction of chronic kidney disease (CKD), addressing the critical challenge of model interpretability in AI-driven healthcare applications. Various ML algorithms (random forest, support vector machines, gradient boosting, and logistic regression) are optimized using feature selection techniques (i.e., RFE and SelectKBest) to enhance diagnostic accuracy. Ensemble models achieve 100% accuracy, demonstrating the effectiveness of ML in CKD detection. To ensure transparency, explainable AI (XAI) techniques such as SHAP, fuzzy rule-based systems, and decision trees are applied to identify key biomarkers (e.g., hemoglobin, serum creatinine, and specific gravity), making predictions clinically interpretable. The integration of fuzzy logic further aligns model decisions with medical reasoning, enhancing clinician trust. This research bridges the gap between AI predictions and clinical decision-making, contributing to the development of transparent, data-driven clinical decision support systems for early CKD detection, personalized treatment, and improved patient outcomes.**

Keywords*—*Chronic Kidney Disease, Machine Learning, Explainable AI, SHAP, Feature Selection, Hyperparameter Optimization.

I. Introduction

Chronic kidney disease (CKD) is a major global health issue, affecting over 800 million people and causing 1.2 million deaths in 2017 [1][2]. It often progresses without early symptoms, leading to delayed diagnosis and irreversible renal damage [3][4]. The burden is particularly severe in low- and middle-income countries due to limited access to renal replacement therapies. Early diagnosis is crucial to initiate preventive interventions that can slow disease progression [3]. Machine learning (ML) techniques have shown promise in early CKD detection by identifying hidden patterns in clinical data [5][6]. Feature selection is essential to improve model accuracy and interpretability, especially when handling complex, high-dimensional medical datasets [7].

The black-box nature of many ML models hinders their adoption in clinical settings, where transparency and explainability are vital for building physician trust. Explainable Artificial Intelligence (XAI) techniques help address this limitation by making model decisions interpretable and aligned with medical reasoning [8][9]. While XAI has been applied in other clinical domains [10–15], its integration into CKD prediction remains underexplored, with most studies limited to basic feature selection [16][17].

To address these gaps, this study proposes a novel and explainable ML framework for early CKD prediction. The contributions include (1) the integration of robust feature selection and hyperparameter optimization to enhance model performance, (2) the development of a hybrid XAI approach combining SHAP, decision trees, and fuzzy logic for multi-perspective interpretability, and (3) the provision of a clinically intuitive and transparent explanation of model decisions. This approach advances the field by improving prediction accuracy while ensuring model decisions are understandable and trustworthy for clinical use.

II. Methods

A. Dataset Description

The CKD dataset, obtained from the UCI machine learning repository, comprises 400 patient records with 25 attributes encompassing clinical, biochemical, and demographic data. These attributes include blood and urine test results, electrolyte levels, diabetes status, hypertension, and other health indicators associated with CKD. The dataset contains both numerical and categorical variables, requiring appropriate preprocessing techniques for effective machine learning applications. The target variable (class) indicates whether a patient has CKD or not, making this dataset a valuable resource for predictive modeling, early detection, and medical decision support in CKD diagnosis.

Table I. Dataset Description.

Original Name	Abb.	Datatype	Range/Unique Values
Age	age	Numerical	(2.0–90.0)
Blood pressure	bp	Numerical	(50.0–180.0)
Specific gravity	sg	Numerical	(1.005–1.025)
Albumin	al	Numerical	(0.0–5.0)
Sugar	su	Numerical	(0.0–5.0)
Red blood cells	rbc	Categorical	[nan 'normal' 'abnormal']
Pus cell	pc	Categorical	[normal, abnormal]
Pus cell clumps	pcc	Categorical	[notpresent, present]
Bacteria	ba	Categorical	[notpresent, present]
Blood glucose random	bgr	Numerical	(22.0–490.0)
Blood urea	bu	Numerical	(1.5–391.0)
Serum creatinine	sc	Numerical	(0.4–76.0)
Sodium	sod	Numerical	(4.5–163.0)
Potassium	pot	Numerical	(2.5–47.0)
Hemoglobin	hemo	Numerical	(3.1–17.8)
Packed cell volume	pcv	Categorical	[44, 38, 31, 32, 35]
White blood cell count	wc	Categorical	[7800, 6000, 7500, 6700, 7300]
Red blood cell count	rc	Categorical	[5.2, 3.9, 4.6, 4.4]
Hypertension	htn	Categorical	[Yes, No]
Diabetes mellitus	dm	Categorical	[Yes, No]
Coronary artery disease	cad	Categorical	[Yes, No]
Appetite	appet	Categorical	[Good, Poor]
Pedal edema	pe	Categorical	[Yes, No]
Anemia	ane	Categorical	[Yes, No]
Class	Class	Categorical	[CKD, Not CKD]

B. Proposed Architecture

The proposed framework for CKD prediction consists of several well-defined phases, each contributing to the development of an accurate, interpretable, and clinically relevant ML model. The framework encompasses data preprocessing, feature engineering, hyperparameter optimization, model selection, and model explainability to ensure robust predictions while maintaining clinical transparency. Below is a detailed breakdown of each phase:

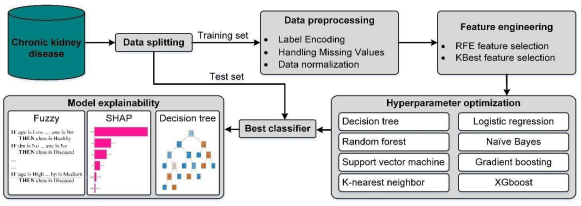

Fig. 1. The proposed architecture.

1. Data Splitting

To prevent data leakage, the first step in the framework involves data partitioning into training and test sets to facilitate supervised learning. Given a raw dataset D consisting of N instances and M features, it is divided into a training set D_{Train} used for model learning and optimization, and a test set D_{Test} used to evaluate the generalization performance. The data split ratio is typically set as stratified 70:30 for train:test, ensuring the model is trained on a sufficiently large sample while retaining an unbiased subset evaluation.

2. Data Preprocessing

The data preprocessing pipeline is designed to handle missing values, apply feature scaling, and encode categorical variables, ensuring the dataset is optimally prepared for machine learning models. This step is only performed on the training dataset, and the trained operators are used to transform the testing set. For numerical features, missing values are imputed using the mean strategy. Feature scaling is applied using z-score normalization, which normalizes the distribution by centering around zero with a standard deviation of one, improving model stability and performance, i.e., $\hat{x} = x - \mu/\sigma$, where μ and σ represent the mean and standard deviation of the feature, respectively. For categorical features, missing values are handled using the mode, replacing them with the most frequently occurring category to retain categorical integrity, and feature encoding is performed using one-hot encoding. These preprocessing steps are seamlessly integrated using the ColumnTransformer operator in Sklearn, which applies numerical and categorical transformations independently while maintaining a structured and unified data representation.

3. Feature engineering and hyperparameter optimization

To enhance model performance, reduce computational complexity, and improve generalization, we applied two feature selection techniques: Recursive Feature Elimination (RFE) and SelectKBest, ensuring that only the most informative features were retained for model training. Feature selection is a crucial step to enhance model interpretability and reduce dimensionality while maintaining predictive power. The framework employs two feature selection techniques:

1. Recursive Feature Elimination: RFE, a wrapper-based approach, recursively eliminates the least significant features by training a Random Forest Classifier as the base estimator, iteratively refining the selection until only the 10 most influential features remain. This method recursively removes the least important features based on model coefficients. A feature subset is selected by optimizing a scoring function: $RFE(F) = \underset{\acute{F}}{\operatorname{argmin}}\, \mathcal{L}(M(\acute{F}))$, where $\mathcal{L}$ is the loss function, $M(\acute{F})$ is the model trained on subset $\acute{F}$.
2. KBest Feature Selection: SelectKBest is a univariate filter method. Features are ranked using a statistical test (e.g., ANOVA F-score), selecting the top k features that maximize discrimination, selecting the top 10 features (k=10) to maximize correlation with the target variable while minimizing noise and redundancy.

By systematically identifying the most relevant features, this feature selection pipeline reduces the risk of overfitting, enhances interpretability, optimizes computational efficiency, and ensures a robust learning process for improved predictive performance. We independently applied both RFE and SelectKBest to the preprocessed dataset, followed by training multiple machine learning models with Bayesian optimization for hyperparameter tuning. Model performance was evaluated using accuracy scores on the test set.

4. Model Optimization

To enhance model performance, hyperparameter tuning is conducted using Bayesian optimization and cross-validation. The framework optimizes parameters for multiple classifiers. (1) Decision Tree: Optimized for maximum depth, minimum samples per split; (2) Random Forest: Optimized for the number of estimators, max depth; (3) Support Vector Machine: Optimized for kernel type and regularization parameter (C); (4) K-Nearest Neighbor: Optimized for the number of neighbors (k); (5) Logistic Regression: Optimized for solver type and regularization; (6)Boosting: Optimized for learning rate, number of estimators.

Each model undergoes evaluation using a 10-fold cross-validation strategy, where performance metrics such as accuracy, precision, recall, and F1-score are computed. Following hyperparameter tuning, the best-performing classifier is selected based on classification accuracy and explainability metrics. The optimal model is then used for final validation on the test dataset. The classification decision function for a given instance x is defined as $\hat{y} = f_\theta(x)$, where f_θ is the trained model with parameters θ, and $\hat{y}$ is the predicted class.

5. Model Explainability

To enhance clinical trust and adoption, the framework integrates XAI techniques:

1. Fuzzy Logic: Converts model decisions into human-readable rules. For example: IF age is high AND … AND blood pressure is medium THEN class = Diseased. This ensures logical consistency between model predictions and medical knowledge.
2. SHAP: Measures the impact of each feature on the prediction using the Shapley value formula: $\phi_j = \sum_{S \subseteq F \setminus \{j\}} \frac{|S|!(|F|-|S|-1)!}{|F|!} \big(v(S \cup \{j\}) - v(S)\big)$, where $v(S)$ represents the model's output when using a feature subset S.
3. Decision Tree Visualization: Constructs an interpretable tree-based representation of model decisions, showing which features contribute to classification.

III. RESULTS

This section presents the findings of the study, including the analysis of ML models' results and the discussion of model XAI features. Table 2 shows the testing results of the optimized ML models using the SelectK features. It indicates that all tested models have achieved high accuracy, precision, and F1-scores, demonstrating strong performance in classification tasks. Gradient Boosting and SVM yielded the highest accuracy (98.75%), while Logistic Regression, AdaBoost, and Random Forest performed lower compared to SVM and gradient boosting, each achieving 98.12% accuracy. Despite having the lowest accuracy (97.50%), KNN maintained a competitive F1-score of 98%. Notably, all models exhibited high precision (100%), suggesting their robustness in minimizing false positives. However, variations in recall values highlight differences in sensitivity, with KNN showing the lowest recall (96%), indicating potential

challenges in capturing all positive instances. These findings provide insights into model selection for optimal predictive performance in the given dataset.

TABLE II. RESULT OF THE MODELS BASED ON THE SELECTK FEATURES.

Model	Accuracy	Precision	Recall	F1-Score
Random Forest	98.12%	100%	97%	98%
Gradient Boosting	98.75%	100%	98%	99%
AdaBoost	98.12%	100%	98%	99%
SVM	98.75%	100%	97%	98%
Logistic Regression	98.12%	100%	98%	99%
Decision Tree	98.12%	100%	97%	98%
KNN	97.50%	100%	96%	98%

Table 3 highlights the effectiveness of RFE in optimizing model performance. As noticed, random forest and gradient boosting achieved perfect resting accuracy (100%), demonstrating their robustness in classification tasks when feature selection was applied. These models also exhibited perfect testing precision and F1-scores (100%) scores, indicating their strong predictive capabilities without misclassifications. Other models, such as AdaBoost, Decision Tree, and Logistic Regression, maintained high accuracy (98.75%–98.12%) and F1-scores (99%), showing their reliability after feature selection. SVM and KNN had slightly lower accuracy (97.50%), with KNN exhibiting the lowest recall (96%), suggesting potential limitations in capturing all positive cases. Overall, the RFE technique improved model efficiency by enhancing accuracy and interpretability, reinforcing its importance in feature selection for optimal predictive performance.

Results of Table 2 and Table 2 could raise overfitting concern. However, the proposed pipeline prevented any possibility of overfitting. This is because dataset has been divided into training and testing from the first beginning, the data preprocessing steps have been applied on the training set alone and been used to transform the testing set, and the models have been validated using 10-fold-cross validation technique. The possible reasons for good performance can be justified by the accurate data preprocessing steps including feature selection optimization, the hyperparameters optimization process using the powerful Bayesian optimizer, and the high quality of the used dataset. Note that literature has many studies that achieve comparable performance [18].

TABLE III. RESULT OF THE MODELS BASED ON THE RFE FEATURES.

Model	Accuracy	Precision	Recall	F1-Score
Random Forest	100%	100%	97%	98%
Gradient Boosting	100%	100%	100%	100%
AdaBoost	98.75%	100%	98%	99%
SVM	97.50%	100%	98%	99%
Logistic Regression	98.12%	100%	98%	99%
Decision Tree	98.75%	100%	98%	99%
KNN	97.50%	100%	96%	98%

IV. EXPLAINABLE AI

To enhance model transparency and interpretability, XAI techniques were employed. These methods provided insights into the decision-making process of machine learning models, ensuring that their outputs were comprehensible and trustworthy. SHAP values were computed to measure the impact of each feature on model predictions. This technique, based on cooperative game theory, provides a unified measure of feature importance by distributing the prediction difference between the actual output and the expected output across all input features. By leveraging SHAP values, a more granular understanding of feature contributions was obtained, aiding in the identification of the most influential variables affecting the model's decisions. Additionally, SHAP summary plots were used to visualize the overall impact of each feature across multiple instances, while SHAP force plots were employed to explain individual predictions by illustrating how features push the model's output higher or lower. Decision Tree-Based Explainability which inherently offers high interpretability due to their rule-based structure has been explored. Fuzzy rule-based systems were employed to enhance decision-making by modeling uncertainty and imprecision in data. This approach enables a more human-like interpretation of the relationships between input features and predictions. The fuzzy rule set was constructed as follows: Fuzzification: Continuous input variables were transformed into fuzzy linguistic terms (e.g., low, medium, high) using predefined membership functions. A set of fuzzy IF-THEN rules were derived based on data-driven approaches. By integrating these explainability techniques, the study ensured that model decisions were interpretable, transparent, and aligned with domain knowledge, thereby enhancing trust in the predictive framework. Figure 2 presented the decision tree model which identified the following critical decision rules for predicting CKD. Each rule represented a sequence of feature-based conditions leading to a 100% probability of CKD diagnosis:

- **Rule 1**: IF hemo $\leq 13.05 \wedge$ sc $\leq 1.15 \wedge$ rbc ≤ 1.50 THEN CKD.
- **Rule 2**: IF hemo $\leq 13.05 \wedge$ sc $\leq 1.15 \wedge$ rbc $> 1.50 \wedge$ al $\leq 0.50 \wedge$ pcv ≤ 23.50 THEN CKD.
- **Rule 3**: IF hemo $\leq 13.05 \wedge$ sc $\leq 1.15 \wedge$ rbc $> 1.50 \wedge$ al > 0.50 THEN CKD.
- **Rule 4**: IF hemo $\leq 13.05 \wedge$ sc $> 1.15 \wedge$ and hemo ≤ 12.95 THEN CKD.
- **Rule 5**: IF hemo $\leq 13.05 \wedge$ sc $> 1.15 \wedge$ hemo $> 12.95 \wedge$ age > 55 THEN CKD.
- **Rule 6**: IF hemo $> 13.05 \wedge$ sg ≤ 1.02 THEN CKD.
- **Rule 7**: IF hemo $> 13.05 \wedge$ sg $> 1.02 \wedge$ htn > 1.50 THEN CKD.
- **Rule 8**: IF hemo $\leq 13.05 \wedge$ sc $\leq 1.15 \wedge$ rbc $\leq 1.5 \wedge$ al $\leq 0.5 \wedge$ pcv ≤ 23.5 THEN Not CKD.
- **Rule 9**: IF hemo $> 13.05 \wedge$ sg $\leq 1.017 \wedge$ htn ≤ 1.5 THEN Not CKD
- **Rule 10**: IF hemo $\leq 13.05 \wedge$ sc $\leq 1.15 \wedge$ rbc $\leq 1.5 \wedge$ al ≤ 0.5 THEN Not CKD.

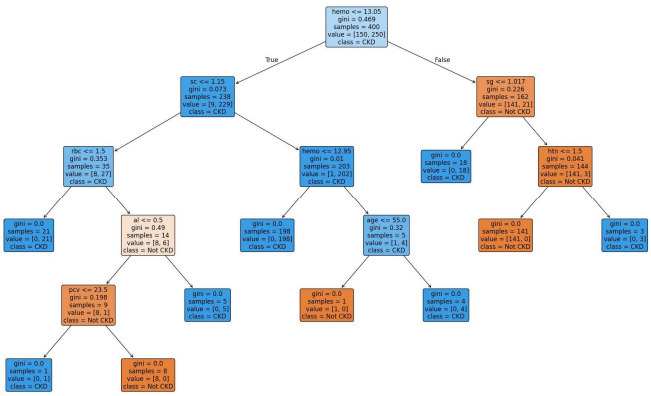

Fig. 2. XAI based on the decision tree glass-box model.

The three figures of Figure 3, Figure 4, and Figure 5 illustrate feature importance rankings. They provided insights into the most significant attributes influencing CKD classification, derived from different feature selection and explainability techniques. Figure 3 presents the top 10 important features identified by the decision tree model. The scores of feature importance highlight the most influential factors in predicting chronic disease. In Figure 3, the top-

ranked features were included where features like age, pe, and appet have minimal influence on the model.

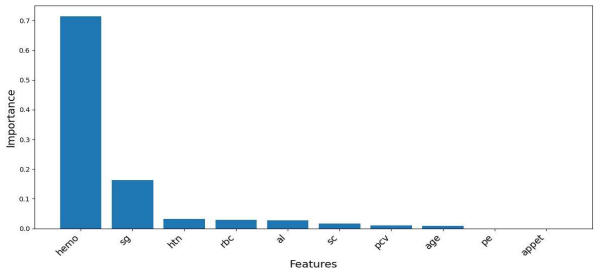

Fig. 3. Top important features (decision tree).

- hemo: The most significant predictor, reinforcing its clinical importance.
- sg: Retains a high ranking, emphasizing its role in kidney function monitoring.
- htn: Appears as a key factor in CKD.
- rbc, al, and sc: All relevant in kidney pathology.

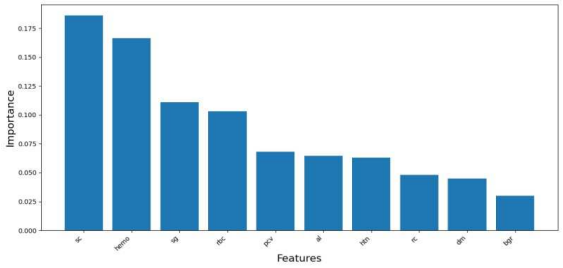

Fig. 4. Top important features (random forest).

Figure 4 presents the top 10 important features identified by the random forest model. The feature importance scores revealed the key factors influencing chronic disease prediction, with sc and hemo being the most significant. In Figure 4, the most significant features were:

- sc: The most important predictor, aligning with clinical knowledge that increased creatinine indicates reduced kidney function.
- hemo: CKD patients often suffer from anemia, making hemoglobin a crucial biomarker.
- sg: Related to kidney filtration efficiency, lower values indicate CKD.
- rbc and pcv: Important for detecting anemia and hydration levels.

The presence of albumin (al), hypertension (htn), and blood urea (bu) further highlighted the connection between kidney function and systemic conditions. This ranking aligned with medical knowledge, as CKD patients typically exhibit abnormalities in serum creatinine, hemoglobin, and specific gravity due to impaired renal function. Figure 5 presents the feature importance ranking based on SHAP values. SHAP values measured how much each feature contributes to model predictions, making this an interpretability-driven feature ranking. The key features identified included:

- dm: The top feature, indicating that non-diabetic status may reduce CKD risk.
- Hemo and sg: Highly influential, reinforcing their clinical relevance.
- htn: Patients with hypertension are at higher risk for CKD progression.
- pcv_52: Affects oxygen transport and hydration status.
- bu and al: Elevated levels indicate kidney dysfunction.
- sc and bgr: Core indicators of kidney filtration ability and metabolic health.
- appet: A symptom associated with CKD progression.

Unlike Figure 3, SHAP analysis emphasized the categorical feature of diabetes mellitus (dm) as the most impactful, which suggested that CKD risk was highly influenced by diabetic status, an important consideration for predictive models.

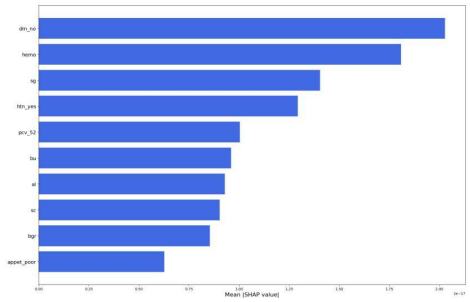

Fig. 5. Features important based on SHAP values.

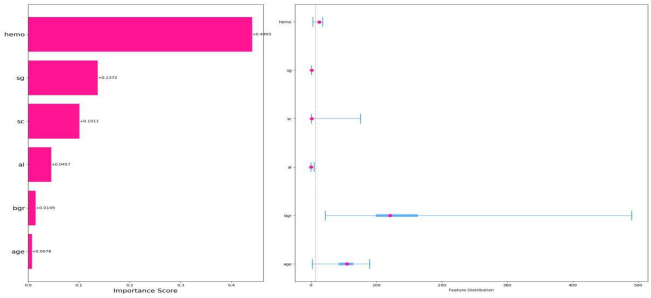

Fig. 6. Global feature importance.

Figure 6 presents the global feature importance analysis. The left chart displayed the most influential features in the model, with hemo having the highest impact. The right plot summarizes the local distribution of feature effects, providing insights into individual predictions. This visualization helped in understanding both overall and case-specific feature contributions in chronic disease prediction.

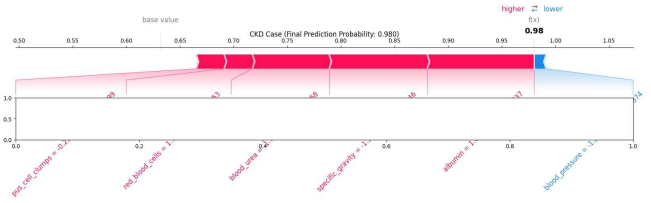

Fig. 7. Force plot for CKD case.

Figure 7 presents a SHAP force plot illustrating the impact of individual features on a CKD Case. This SHAP's force plot explained the final CKD prediction probability of 0.98, illustrating how individual features contribute to the classification. The base value represented the average model output, and features either increase (red) or decrease (blue) the probability of CKD. Key features pushing the prediction higher included pus cell clumps (0.2), red blood cells (1), blood urea (92.4 mg/dL), specific gravity (1.01), and albumin (1), all of which were strong clinical indicators of CKD. Conversely, blood pressure (21.4) slightly decreased the probability but was insufficient to alter the classification. The model predicted CKD due to significant markers such as albuminuria, low specific gravity, and high blood urea, reinforcing their clinical importance. This visualization enhances explainability in AI-driven diagnosis, making predictions transparent and interpretable for clinicians, thereby fostering trust in ML-based CKD detection.

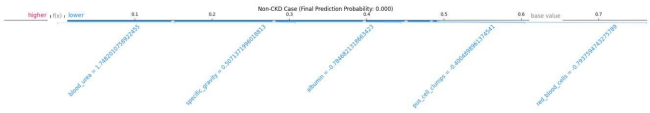

Fig. 8. Force plot for non-CKD case.

Figure 8 presented a SHAP force plot illustrating the impact of individual features on a Non-CKD Case. This SHAP force plot explained the final prediction probability of 0.000, indicating a non-CKD case with high confidence. The base value represented the model's average probability before considering feature contributions. In this case, all contributing features pushed the probability lower, as shown by the blue arrows, reinforcing a strong classification as a non-CKD patient. Key features that lowered the probability of CKD included blood urea (1.74), specific gravity (1.02), albumin (0), pus cell clumps (0.4), and red blood cells (0.29). These values suggested normal kidney function since low blood urea indicates proper filtration, higher specific gravity reflected healthy urine concentration ability, and the absence of albuminuria and pus cell clumps supported normal kidney health. The low red blood cell count also did not significantly affect classification. The model confidently predicted no CKD, as the key clinical markers remained within normal ranges. This interpretability provided by SHAP enhanced trust in AI-driven diagnosis by offering a transparent explanation of why the model has classified the patient as non-CKD.

Figure 9 presented a SHAP summary plot displaying the impact of individual features on model predictions. Each point represented a feature's SHAP value for a specific instance, with color indicating feature value (red for high and blue for low). This SHAP summary plot provided a detailed interpretability analysis of the model's predictions for CKD by illustrating the impact of different features. The Y-axis ranked features by importance, while the X-axis represented SHAP values, indicating whether a feature increases (positive SHAP value) or decreases (negative SHAP value) the likelihood of CKD.

The color gradient (red for high values, blue for low values) further contextualized feature influence. Key predictors included hemoglobin (hemo), serum creatinine (sc), specific gravity (sg), and albumin (al), which aligned with established clinical markers for CKD. Low hemoglobin, high serum creatinine, and low specific gravity strongly contributed to CKD classification, reinforcing the model's reliability. Diabetes (dm_yes), hypertension (htn_yes), and elevated blood urea (bu) also push predictions toward CKD, while normal red blood cell count (rbc_normal), normal packed cell volume (pc_normal), and stable blood pressure (bp) decreased CKD risk. The plot highlighted the interplay of multiple biomarkers, making AI-driven CKD prediction transparent and clinically interpretable, thereby fostering trust and adoption in medical decision-making.

The last XAI methodology was the fuzzy knowledge base. Fuzzy systems can be used to represent XAI knowledge using natural language and human understandable format. The following rule set defined the relationship between different medical attributes and the classification outcome. These rules provided the knowledge base of Mamdani fuzzy rule-based system. The fuzzification process was based on a set of predefined membership functions which can be learned from the data. Figure 10 presented the Gaussian fuzzy membership functions that were learned from our dataset where x-axis represents membership degree and y-axis represents the features. The following was the rule base or knowledge base that can be used to classify a patient using fuzzy variables and fuzzy values such as low, medium, and high based on the corresponding membership functions.

Fig. 9. SHAP summary plot.

- **RULE 1:** IF age is Low ∧ p is Low ∧ sg is Low ∧ al is Low ∧ su is Low ∧ hb is Low ∧ pcv is Low ∧ wc is Low ∧ rc is Low ∧ rbc is No ∧ pc is No ∧ pcc is No ∧ ba is No ∧ htn is No ∧ dm is No ∧ cad is No ∧ appet is No ∧ pe is No ∧ ane is No THEN Non-CKD.
- **RULE 2:** IF age is High ∧ bp is High ∧ sg is High ∧ al is High ∧ su is High ∧ hb is High ∧ pcv is High ∧ wc is High ∧ rc is High ∧ rbc is Yes ∧ pc is Yes ∧ pcc is Yes ∧ ba is Yes ∧ htn is Yes ∧ dm is Yes ∧ cad is Yes ∧ appet is Yes ∧ pe is Yes ∧ ane is Yes THEN CKD.
- **RULE 3:** IF age is Medium ∧ bp is Medium ∧ sg is Medium ∧ al is Medium ∧ su is Medium ∧ hb is Medium ∧ pcv is Medium ∧ wc is Medium ∧ rc is Medium ∧ rbc is No ∧ pc is No ∧ pcc is No ∧ ba is No ∧ htn is No ∧ dm is No ∧ cad is No ∧ appet is No ∧ pe is No ∧ ane is No THEN CKD.
- **RULE 4:** IF age is Low ∧ bp is Medium ∧ sg is High ∧ al is Low ∧ su is Medium ∧ hb is High ∧ pcv is Low ∧ wc is Medium ∧ rc is High ∧ rbc is No ∧ pc is Yes ∧ pcc is No ∧ ba is No ∧ htn is Yes ∧ dm is No ∧ cad is Yes ∧ appet is No ∧ pe is Yes ∧ ane is No THEN CKD.
- **RULE 5:** IF age is High ∧ bp is Low ∧ sg is Medium ∧ al is High ∧ su is Low ∧ hb is Medium ∧ pcv is High ∧ wc is Low ∧ rc is Medium ∧ rbc is Yes ∧ pc is No ∧ pcc is Yes ∧ ba is Yes ∧ htn is No ∧ dm is Yes ∧ cad is No ∧ appet is Yes ∧ pe is No ∧ ane is Yes THEN Non-CKD.
- **RULE 6:** IF age is Medium ∧ bp is Low ∧ sg is High ∧ al is Medium ∧ su is Low ∧ hb is High ∧ pcv is Medium ∧ wc is Low ∧ rc is High ∧ rbc is No ∧ pc is Yes ∧ pcc is No ∧ ba is No ∧ htn is Yes ∧ dm is No ∧ cad is Yes ∧ appet is No ∧ pe is Yes ∧ ane is No THEN CKD.
- **RULE 7:** IF age is Low ∧ bp is High ∧ sg is Medium ∧ al is Low ∧ su is High ∧ hb is Medium ∧ pcv is Low ∧ wc is High ∧ rc is Medium ∧ rbc is Yes ∧ pc is No ∧ pcc is Yes ∧ ba is Yes ∧ htn is No ∧ dm is Yes ∧ cad is No ∧ appet is Yes ∧ pe is No ∧ ane is Yes THEN Non-CKD.
- **RULE 8:** IF age is High ∧ bp is Medium ∧ sg is Low ∧ al is High ∧ su is Medium ∧ hb is Low ∧ pcv is High ∧ wc is Medium ∧ rc is Low ∧ rbc is No ∧ pc is Yes ∧ pcc is No ∧ ba is No ∧ htn is Yes ∧ dm is No ∧ cad is Yes ∧ appet is No ∧ pe is Yes ∧ ane is No THEN CKD.
- **RULE 9:** IF age is Medium ∧ bp is High ∧ sg is Low ∧ al is Medium ∧ su is High ∧ hb is Low ∧ pcv is Medium ∧ wc is High ∧ rc is Low ∧ rbc is Yes ∧ pc is No ∧ pcc is Yes ∧ ba is Yes ∧ htn is No ∧ dm is Yes ∧ cad is No ∧ appet is Yes ∧ pe is No ∧ ane is Yes THEN Non-CKD.
- **RULE 10:** IF age is Low ∧ bp is Low ∧ sg is High ∧ al is Low ∧ su is Low ∧ hb is High ∧ pcv is Low ∧ wc is Low ∧ rc is High ∧ rbc is No ∧ pc is Yes ∧ pcc is No ∧ ba is No ∧ htn is Yes ∧ dm is No ∧ cad is Yes ∧ appet is No ∧ pe is Yes ∧ ane is No THEN CKD.

Although the prediction of CKD using ensemble ML models has been studied in literature, this study provided a significant contribution to the literature. These contributions can be summarized as follows. First, the study provided an accurate and consistent ML pipeline for predicting chronic diseases by integrating feature selection and hyperparameter optimization. Second, the study explored three of the most popular XAI techniques, including decision tree, SHAP, and fuzzy reasoning. The combination and fusion of the three techniques provided a consistent and complementary method to justify the model's decision from different perspectives.

The resulting hybrid XAI provided a medically intuitive interpretability of model decisions, which improved the medical trustworthiness of model decisions. Third, CKD's literature has not discussed the smoothness and compatibility of the provided XAI approaches. In addition, the deep justification of the resulting XAI and its linking with the medical literature provided a nice and accurate medical justification of the decisions. These issues have not been explored in the CKD literature. For reproducibility purposes, this is the GitHub of the code: https://github.com/NadaGamalElAdl/CKD_XAI.

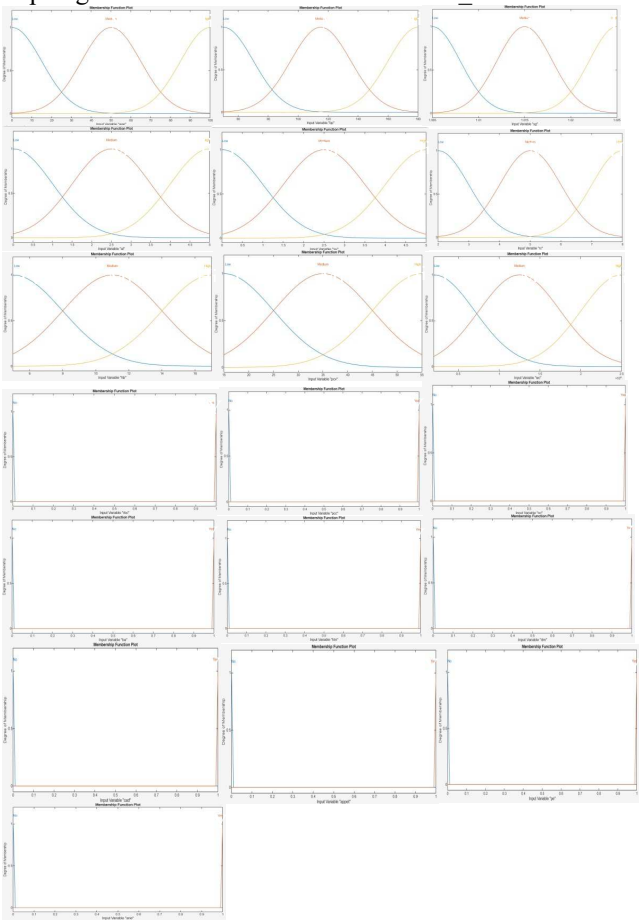

Fig. 10. Fuzzy membership functions.

V. CONCLUSION

This study proposed an explainable machine learning framework for early CKD detection, integrating feature selection (RFE, SelectKBest) and ML models (Random Forest, Gradient Boosting, SVM, Logistic Regression) to enhance accuracy, with ensemble models achieving up to 100% classification performance. To ensure clinical transparency, XAI techniques such as SHAP, decision trees, and fuzzy rule-based systems identified key biomarkers (hemoglobin, serum creatinine, specific gravity, albumin), aligning with medical knowledge and fostering clinician trust. Despite its effectiveness, future work should focus on external validation on larger datasets, clinical workflow integration, hybrid AI models, advanced explainability methods, CKD progression prediction, and ethical bias mitigation. This research underscores the potential of AI-driven decision support systems in nephrology, contributing to accurate, transparent, and scalable CKD diagnosis in real-world healthcare settings.

ACKNOWLEDGMENT

This study was funded by the MSIT (Ministry of Science and ICT), Korea, under the ICT Creative Consilience Program (1ITP-2021-2020-0-01821) supervised by the IITP (Institute for Information & Communications Technology Planning & Evaluation).

REFERENCES

[1] Kovesdy, C. P. (2022). Epidemiology of chronic kidney disease: an update 2022. Kidney international supplements, 12(1), 7-11.

[2] Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. *Advances in neural information processing systems*, *30*.

[3] Webster, A. C., Nagler, E. V., Morton, R. L., & Masson, P. (2017). Chronic kidney disease. The lancet, 389(10075), 1238-1252.

[4] Polat, H., Danaei Mehr, H., & Cetin, A. (2017). Diagnosis of chronic kidney disease based on support vector machine by feature selection methods. Journal of medical systems, 41, 1-11.

[5] Lei, N., Zhang, X., Wei, M., Lao, B., Xu, X., Zhang, M., ... & Wu, Y. (2022). Machine learning algorithms' accuracy in predicting kidney disease progression: a systematic review and meta-analysis. BMC Medical Informatics and Decision Making, 22(1), 205.

[6] Qezelbash-Chamak, J., Badamchizadeh, S., Eshghi, K., & Asadi, Y. (2022). A survey of machine learning in kidney disease diagnosis. Machine Learning with Applications, 10, 100418.

[7] Abdel-Fattah, M. A., Othman, N. A., & Goher, N. (2022). Predicting chronic kidney disease using hybrid machine learning based on apache spark. Computational Intelligence and Neuroscience, 2022(1), 9898831.

[8] Stiglic, G., Kocbek, P., Fijacko, N., Zitnik, M., Verbert, K., & Cilar, L. (2020). Interpretability of machine learning-based prediction models in healthcare. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 10(5), e1379.

[9] Tjoa, E., & Guan, C. (2020). A survey on explainable artificial intelligence (XAI): Toward medical XAI. IEEE transactions on neural networks and learning systems, 32(11), 4793-4813.

[10] Zhang, H., Ren, J. X., Ma, J. X., & Ding, L. (2019). Development of an in silico prediction model for chemical-induced urinary tract toxicity by using naïve Bayes classifier. Molecular diversity, 23, 381-392.

[11] Sossi Alaoui, S., Aksasse, B., & Farhaoui, Y. (2020). Data mining and machine learning approaches and technologies for diagnosing diabetes in women. In Big Data and Networks Technologies 3 (pp. 59-72). Springer International Publishing.

[12] Zhang, Y., & Ma, Y. (2019). Application of supervised machine learning algorithms in the classification of sagittal gait patterns of cerebral palsy children with spastic diplegia. Computers in biology and medicine, 106, 33-39.

[13] Feeny, A. K., Rickard, J., Patel, D., Toro, S., Trulock, K. M., Park, C. J., ... & Chung, M. K. (2019). Machine learning prediction of response to cardiac resynchronization therapy: improvement versus current guidelines. Circulation: Arrhythmia and Electrophysiology, 12(7), e007316.

[14] Aro, T. O., Akande, H. B., Jibrin, M. B., & Jauro, U. A. (2019). Homogeneous ensembles on data mining techniques for breast cancer diagnosis. Daffodil international university journal of science and technology, 14(1), 9-12.

[15] Bucholc, M., Ding, X., Wang, H., Glass, D. H., Wang, H., Prasad, G., & Alzheimer's Disease Neuroimaging Initiative. (2019). A practical computerized decision support system for predicting the severity of Alzheimer's disease of an individual. Expert systems with applications, 130, 157-171.

[16] Jha, V., Garcia-Garcia, G., Iseki, K., Li, Z., Naicker, S., Plattner, B., & Yang, C. W. (2013). Chronic kidney disease: global dimension and perspectives. The Lancet, 382(9888), 260-272.

[17] Arif, M. S., Mukheimer, A., & Asif, D. (2023). Enhancing the early detection of chronic kidney disease: a robust machine learning model. Big Data and Cognitive Computing, 7(3), 144.

[18] Dritsas, E., & Trigka, M. (2022). Machine learning techniques for chronic kidney disease risk prediction. Big Data and Cognitive Computing, 6(3), 98.