A Hybrid CNN–LSTM–GRU Deep Learning Model for the Accurate Classification of Chronic Kidney Disease

Ahmed Fahim
Department of Computer Science, College of Computer Engineering and Sciences, Prince Sattam Bin Abdulaziz University, Al-Kharj, Saudi Arabia | Department of Computer Science, Faculty of Computers and Information, Suez University, Suez, Egypt
a.abualeala@psau.edu.sa (corresponding author)

Ahmed M. Osman
Department of Information Systems, Faculty of Computers and Information, Suez University, Suez, Egypt
a.osman@suezuni.edu.eg

Zahraa Tarek
Department of Computer Engineering and Information, College of Engineering, Wadi Ad Dwaser, Prince Sattam Bin Abdulaziz University, Saudi Arabia | Computer Science Department, Faculty of Computers and Information, Mansoura University, Mansoura, Egypt
z.elmana@psau.edu.sa

Ahmed M. Elshewey
Department of Computer Science, Faculty of Computers and Information, Suez University, P.O. Box: 43221, Suez, Egypt | Applied Science Research Center, Applied Science Private University, Amman, Jordan
ahmed.elshewey@fci.suezuni.edu.eg

Received: 20 August 2025 | Revised: 31 August 2025 and 7 September 2025 | Accepted: 9 September 2025

 DOI: https://doi.org/10.48084/etasr.14206

ABSTRACT

Chronic Kidney Disease (CKD) is a progressive and often undiagnosed condition that poses a significant global health risk due to its silent progression and typical detection at a late stage. This study presents an advanced hybrid deep learning framework that integrates Convolutional Neural Networks (CNN), Long Short-Term Memory (LSTM), and Gated Recurrent Unit (GRU) architectures to improve the early prediction and classification of CKD. The framework employs a preprocessing pipeline that includes data cleaning, normalization, and class balancing using the Synthetic Minority Oversampling Technique (SMOTE) before performing deep feature extraction and sequence modeling. The hybrid CNN-LSTM-GRU model demonstrated outstanding performance, achieving an accuracy of 98.75%, a precision of 100%, a recall of 97.56%, an F1-score of 98.77%, and an Area Under the ROC Curve (AUC) of 0.988. These results significantly outperform conventional models such as LSTM, GRU, DNN, and 1D-CNN. The proposed framework has strong potential to support clinical decision-making systems for accurate, early CKD diagnosis, thereby improving patient outcomes and reducing the burden on healthcare systems.

Keywords-chronic kidney disease; hybrid neural network; health informatics; CNN–LSTM–GRU; deep learning

I. INTRODUCTION

CKD is a major non-communicable disease that contributes to a growing number of premature deaths. The condition progresses silently over time, often undetected until its later stages when irreversible kidney damage has already occurred. The disease imposes a substantial health burden, particularly in low- and middle-income countries such as Egypt and Bangladesh. For example, Egypt has recorded a CKD-related

mortality rate exceeding 25%, more than double the global average of 10%. The lack of routine screening, limited awareness, and absence of early diagnostic tools further complicate efforts to mitigate the progression of CKD. CKD arises from the kidneys' diminished ability to effectively filter blood, leading to the accumulation of waste products and fluids in the body. Symptoms, such as fatigue, vomiting, high blood pressure, and abdominal discomfort, typically appear only in advanced stages. This latent progression makes early detection critical to reduce complications and improve long-term survival. Traditional diagnostic approaches, however, are often limited by their reliance on manual analysis and static thresholds that fail to account for complex interactions in clinical data.

Advances in artificial intelligence, especially deep learning, have enabled the creation of automated, highly precise diagnostic systems that can analyze medical data in real time. The proposed hybrid model is designed to identify both spatial patterns and sequential dependencies in clinical features, providing a thorough assessment of CKD risk. Numerous machine learning and deep learning approaches have been explored for the early detection and classification of CKD. Traditional supervised learning techniques, such as Support Vector Machine (SVM), Decision Tree (DT), Naïve Bayes (NB), and Logistic Regression (LR), have shown promise in disease classification tasks. Deep Neural Networks (DNNs) have also gained attention due to their strong performance on medical and biomedical datasets. Research has focused on CKD-specific prediction models. Authors in [1] developed a hybrid CNN-SVM model for CKD prediction, outperforming standalone models like SVM and Random Forest (RF). The model achieved 96.8% accuracy and a perfect recall, but the computational load remains high, necessitating further optimization before clinical deployment. Authors in [2] proposed a Density-based Feature Selection method optimized with Ant Colony Optimization (D-ACO), which improved model efficiency by using only 14 of 24 features, achieving 95% accuracy.

Authors in [3] employed a DNN to outperform conventional classifiers, such as NB, LR, RF, and Adaboost, achieving 96.6% accuracy. Authors in [4] developed a machine learning framework for early CKD prediction, addressing late diagnosis and treatment costs. They used a dataset of 400 patients and evaluated six models, with CatBoost achieving the best performance with 97.5% accuracy. Authors in [6] studied RF and ANNs for classifying CKD and predicting creatinine levels. They found that RFs outperformed ANNs for at-home feature-based classification, achieving 92.5% accuracy. Both models showed over 98% accuracy with monitoring and laboratory features, and RFs improved the predictability of creatinine levels when using laboratory features. Authors in [7] developed machine learning models to predict the Length Of Stay (LOS) and hospital billing costs for CKD patients. The RF model performed best for billing prediction, while LR was the best for LOS classification. The study highlights the importance of machine learning for predicting hospital resource utilization and financial planning in CKD care. Authors in [8] developed a multiclass CKD stage prediction model for diabetic patients using longitudinal data from the CRIC study. The TabNet-based model outperformed traditional classifiers and identified key predictors, including serum creatinine, cystatin C, age, and sex. This study offers a promising tool for early detection of CKD at an early stage in diabetic populations, potentially improving interventions and patient care outcomes. Authors in [9] used machine learning to predict Renal Anemia (RA) in patients with chronic kidney disease. They used an optimized FT-Transformer model enhanced with Bayesian Optimization, achieving an accuracy of 91.81%. This study demonstrates the effectiveness of transformer-based architectures for early RA prediction

II. METHODOLOGY

A. Dataset

This study utilized a publicly available CKD dataset obtained from the Kaggle repository [10]. The dataset (N=400) contains 24 clinical attributes, including demographic data (age), vital signs (blood pressure), lab results (specific gravity, albumin, sugar), and binary indicators for comorbidities (hypertension, diabetes, etc.).

TABLE I. STATISTICAL ANALYSIS OF THE PROPOSED DATASET

Attribute	Mean	Std	Min	Max
age	51.48	16.97	2.00	90.00
bp	76.47	13.48	50.00	180.00
sg	1.017	0.005	1.005	1.025
al	1.01	1.27	0.00	5.00
su	0.45	1.03	0.00	5.00
bgr	148.04	74.78	22.00	490.00
htn	0.37	0.48	0.00	1.00
dm	0.34	0.47	0.00	1.00
cad	0.08	0.27	0.00	1.00
appet	0.79	0.40	0.00	1.00
pe	0.19	0.39	0.00	1.00
ane	0.15	0.36	0.00	1.00
Classification	0.62	0.48	0.00	1.00

Before model training, extensive preprocessing was applied to ensure data quality. This included handling missing values using median imputation, standardizing numerical features, encoding categorical variables, and addressing class imbalance using SMOTE. These preprocessing steps improved the dataset's suitability for deep learning classification. To avoid information leakage and preserve class balance, the present study applied stratified splitting: 80% train / 20% test at the patient level, and then carved a 20% validation fraction from the training set for early stopping and model selection. This yields 256/64/80 samples for train/eval/test, respectively (class-balanced: train = 160 CKD / 96 non-CKD; eval = 40 CKD / 24 non-CKD; test = 50 CKD / 30 non-CKD). All preprocessing (imputation, scaling, encoding, and SMOTE on the training fold only) was fitted exclusively on the training data and applied to validation/test using the saved transform parameters.

B. Proposed Methodology

The proposed framework introduces a hybrid deep learning architecture that combines CNN, LSTM, and GRU to enhance the classification of CKD. This design aims to capture both spatial feature relationships and temporal dependencies within the clinical data, thereby providing a robust and comprehensive

predictive model. The pipeline begins with extensive data preprocessing, including normalization, imputation of missing values, and class balancing using SMOTE. The cleaned and standardized dataset is reshaped into a format suitable for deep learning input.

In the feature extraction stage, the CNN module is responsible for learning local feature representations from the input vectors. Its output is then passed sequentially through LSTM and GRU layers, which are optimized to model the temporal and sequential nature of patient data. Dropout layers are introduced to prevent overfitting, followed by dense layers and a sigmoid-activated output layer for binary classification (CKD versus No CKD). Figure 1 displays the workflow diagram of the proposed methodology.

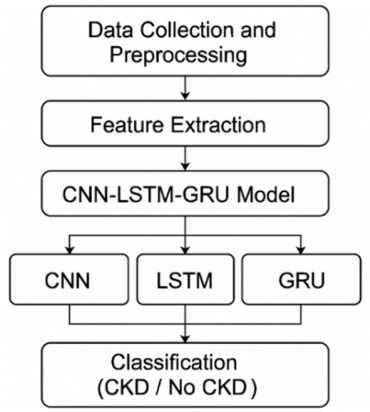


Fig. 1. Workflow diagram of the proposed method.

```
Algorithm 1 – Pseudocode for CNN–LSTM–GRU-Based Hybrid Deep Learning Framework for CKD Classification:
Input: CKD dataset D with N samples and M clinical features
Output: Predicted class label (CKD / No CKD) and performance metrics
1. Preprocess the dataset:
   – Handle missing values using median imputation.
   – Normalize all numerical features.
   – Encode categorical variables.
   – Apply SMOTE to balance class distribution.
2. Prepare data for training:
– Reshape the dataset to match input requirements of CNN layers.
   – Split the dataset into training and testing subsets.
3. Construct the hybrid CNN–LSTM–GRU model:
– Apply a convolutional layer to extract local spatial features.
– Use ReLU activation and max pooling for feature downsampling.
– Flatten or reshape the CNN output to a sequential format.
– Pass the output to an LSTM layer for long-term dependency modeling.
– Pass the LSTM output to a GRU layer to capture short-term dependencies.
   – Apply dropout to reduce overfitting.
   – Add a dense layer with sigmoid activation for binary classification.
4. Compile the model:
   – Use binary cross-entropy as the loss function.
   – Select Adam as the optimizer.
   – Define evaluation metrics: Accuracy, Precision, Recall, F1-Score, AUC.
5. Train the model:
   – Fit the model using the training dataset.
   – Optionally implement early stopping and validation split.
6. Evaluate the model:
   – Predict using the test dataset.
   – Calculate the confusion matrix and evaluation metrics.
   – Plot the ROC curve and compute AUC.
7. Output the final classification results and model performance metrics.
```

III. RESULTS

A comparative analysis of the proposed CNN–LSTM–GRU hybrid deep learning model's performance against several baseline models, including LSTM, Bi-LSTM, GRU, DNN, and 1D-CNN, was conducted for the task of CKD classification. The evaluation is based on a range of metrics: accuracy, precision, recall, F1-score, and AUC. All models were trained and tested on the preprocessed dataset.

The classification performance of each model is summarized in Table III. The CNN–LSTM–GRU hybrid model achieved the highest accuracy of 98.75%, indicating its superior overall predictive power. It recorded a perfect precision of 100%, meaning that the model produced no false positives, and a high recall of 97.56%, confirming its strong ability to correctly identify CKD cases. Its F1-score of 98.77% reflects an excellent balance between precision and recall. In comparison, the DNN and 1D-CNN models also performed well, with accuracies of 96.25% and 95.00%, and F1-scores of 96.30% and 95.00%, respectively, but fell short in either sensitivity or specificity. Conversely, the standalone GRU model exhibited the lowest performance, with an accuracy of 91.25% and F1-score of 91.14%, primarily due to reduced recall and precision. Overall, the CNN–LSTM–GRU model demonstrated the most reliable and robust performance, making it the best candidate for effective and accurate CKD classification. To further interpret the models' behavior, confusion matrices were generated for each architecture. The confusion matrix of the CNN–LSTM–GRU model is shown in Figure 2, revealing that only one CKD case was misclassified (FN = 1), and no false positives occurred. This indicates excellent recall and perfect precision. The confusion matrices for the remaining models are presented in Figure 3.

TABLE II. PERFORMANCE EVALUATION METRICS OF ALL MODELS

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
LSTM	93.75	94.87	92.50	93.67
Bi-LSTM	93.75	92.68	95.00	93.83
GRU	91.25	92.31	90.00	91.14
DNN	96.25	95.12	97.50	96.30
1D-CNN	95.00	95.00	95.00	95.00
CNN-LSTM-GRU	98.75	100.00	97.56	98.77

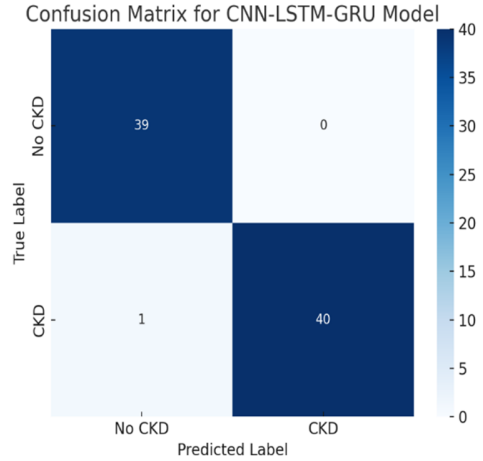
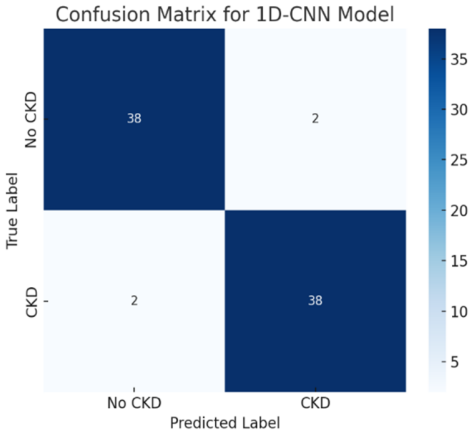
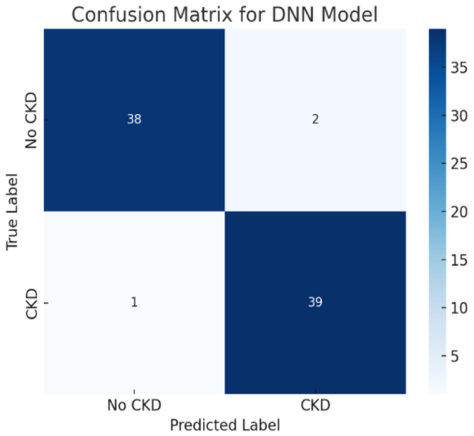


Fig. 2. Confusion matrix of CNN–LSTM–GRU.

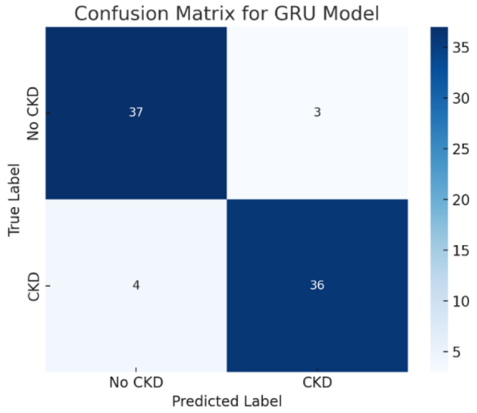
(a)



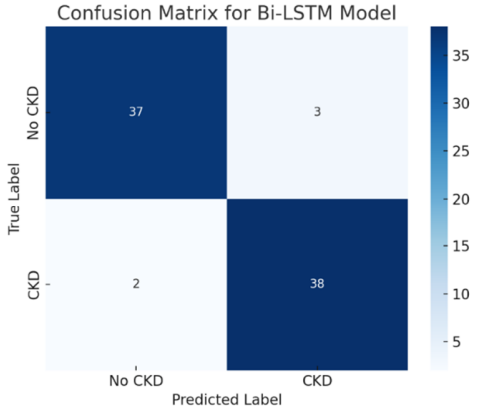
(b)



(c)



(d)



(e)

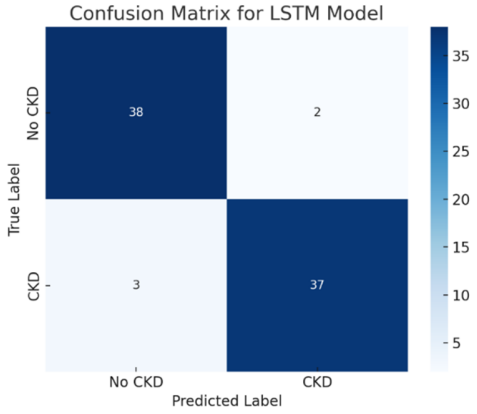


Fig. 3. Confusion matrices for: (a) 1D-CNN, (b) DNN, (c) GRU, (d) Bi-LSTM, and (e) LSTM.

Figure 4 displays the ROC curves of all evaluated models on the independent test set. The proposed CNN-LSTM-GRU model achieved the highest AUC of 0.988, followed by DNN (AUC = 0.962) and 1D-CNN (AUC = 0.950). LSTM and Bi-LSTM obtained AUC values of 0.938, and GRU obtained an AUC value of 0.913.

The integration of CNN, LSTM, and GRU layers allows the proposed model to utilize local spatial features, long-term dependencies, and efficient sequential processing. This synergy helps the model outperform all standalone deep learning methods. Its outstanding performance, especially in reducing false positives and negatives, demonstrates its strong potential

as a clinical decision support tool for early-stage CKD screening.

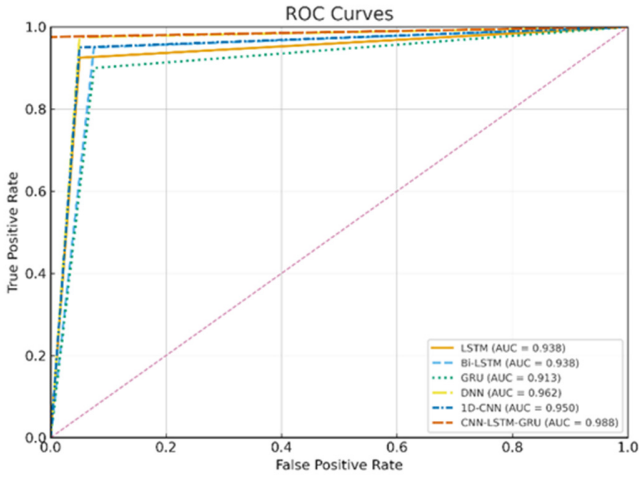


Fig. 4. ROC curve comparison of all models.

A. *Comparative Analysis*

As depicted in Table IV, the proposed model achieves a higher accuracy than previously reported benchmarks on the same CKD dataset.

TABLE III. COMPARATIVE ANALYSIS ON THE SAME DATASET

Study	Best model	Best accuracy (%)
This work	Hybrid CNN–LSTM–GRU	98.75
[11]	LR / SVM (tie)	97.50
[12]	SVM + filter FS	98.50
[13]	XGBoost	93.29

Table V summarizes the key hyperparameters used to configure and optimize the proposed CNN–LSTM–GRU model, including the Adam optimizer (learning rate = 0.001), 64 filters and units for CNN, LSTM, and GRU layers, a dropout rate of 0.3, binary cross-entropy loss, and early stopping with a validation split of 20%.

TABLE IV. SUMMARY OF KEY HYPERPARAMETERS FOR THE PROPOSED MODEL

Optimizer	Adam
Learning rate	0.001
Loss function	Binary Cross-Entropy
Batch size	32
Epochs	100
CNN filters	64
LSTM units	64
GRU units	64
Dropout rate	0.3
Output activation	Sigmoid
Validation split	0.2
Early stopping	Enabled (patience=10)

IV. CONCLUSIONS AND FUTURE WORK

This study presented a hybrid deep learning framework that integrates Convolutional Neural Networks (CNN), Long Short-Term Memory (LSTM), and Gated Recurrent Unit (GRU) for the accurate classification of Chronic Kidney Disease (CKD). Through extensive evaluation on a publicly available CKD dataset, the model achieved outstanding performance, recording an accuracy of 98.75%, a precision of 100%, a recall of 97.56%, an F1-score of 98.77%, and an Area Under the ROC Curve (AUC) of 0.988, outperforming all baseline models. These results highlight the potential of deep hybrid architectures in enhancing early CKD detection and supporting clinical decision-making. Future work will expand this framework by incorporating explainable AI (XAI) techniques, such as SHAP or LIME, to improve model interpretability for clinical practitioners. Additionally, the model will be evaluated on larger and more diverse datasets, including multi-center or real-time clinical data, to assess its generalizability. Integration with Electronic Health Record (EHR) systems and deployment in a real-world diagnostic setting will also be explored to validate its clinical utility. Future extensions of this work may also explore federated learning approaches, such as the method proposed in [14], to enable secure, privacy-preserving CKD classification across multiple clinical sites.

The present evaluation relies on a single, relatively small cohort (N = 400) collected in a limited time window, which may limit generalizability to other settings and populations. Although the current work used stratified splits and fitted preprocessing only on the training data, the utilization of SMOTE to mitigate class imbalance can inflate minority-class performance if not carefully contained within each training fold. In addition, the performed task is formulated as binary CKD detection rather than stage-wise stratification, while this study did not perform external validation on data from an independent site or through prospective collection. Finally, while the proposed hybrid network achieved strong test-set metrics, further work should examine calibration, threshold selection, and clinical utility curves under real-world prevalence.

ACKNOWLEDGMENT

This study is supported via funding from Prince sattam bin Abdulaziz University project number (PSAU/2025/R/1447).

REFERENCES

[1] K. Ramu *et al*., "Hybrid CNN-SVM model for enhanced early detection of Chronic kidney disease," *Biomedical Signal Processing and Control*, vol. 100, Feb. 2025, Art. no. 107084, https://doi.org/10.1016/j.bspc.2024.107084.

[2] M. Elhoseny, K. Shankar, and J. Uthayakumar, "Intelligent Diagnostic Prediction and Classification System for Chronic Kidney Disease," *Scientific Reports*, vol. 9, no. 1, July 2019, Art. no. 9583, https://doi.org/10.1038/s41598-019-46074-2.

[3] H. Kriplani, B. Patel, and S. Roy, "Prediction of Chronic Kidney Diseases Using Deep Artificial Neural Network Technique," in *Computer Aided Intervention and Diagnostics in Clinical and Medical Images*, 2019, pp. 179–187, https://doi.org/10.1007/978-3-030-04061-1_18.

[4] S. R. Kalupukuru and K. Natarajan, "Machine Learning Methods for Predicting the Prognosis of Chronic Kidney Disease," *Procedia*

Computer Science, vol. 258, pp. 1372–1382, Jan. 2025, https://doi.org/10.1016/j.procs.2025.04.370.

[5] P. Chittora *et al.*, "Prediction of Chronic Kidney Disease - A Machine Learning Perspective," *IEEE Access*, vol. 9, pp. 17312–17334, 2021, https://doi.org/10.1109/ACCESS.2021.3053763.

[6] B. Metherall, A. K. Berryman, and G. S. Brennan, "Machine learning for classifying chronic kidney disease and predicting creatinine levels using at-home measurements," *Scientific Reports*, vol. 15, no. 1, Feb. 2025, Art. no. 4364, https://doi.org/10.1038/s41598-025-88631-y.

[7] B. Y. Kiremit and D. Ö. Şahin, "Comparison of machine learning algorithms for predicting length of stay in chronic kidney disease patients," *Computers in Biology and Medicine*, vol. 196, Sept. 2025, Art. no. 110825, https://doi.org/10.1016/j.compbiomed.2025.110825.

[8] M. N. H. Chowdhury *et al.*, "Deep learning for early detection of chronic kidney disease stages in diabetes patients: A TabNet approach," *Artificial Intelligence in Medicine*, vol. 166, Aug. 2025, Art. no. 103153, https://doi.org/10.1016/j.artmed.2025.103153.

[9] Y. Liu, J. Chen, M. Wang, Y. Liu, J. Chen, and M. Wang, "BO–FTT: A Deep Learning Model Based on Parameter Tuning for Early Disease Prediction from a Case of Anemia in CKD," *Electronics*, vol. 14, no. 12, June 2025, https://doi.org/10.3390/electronics14122471.

[10] "Chronic Kidney Disease dataset." https://www.kaggle.com/datasets/mansoordaku/ckdisease

[11] R. C. Poonia *et al.*, "Intelligent Diagnostic Prediction and Classification Models for Detection of Kidney Disease," *Healthcare (Basel, Switzerland)*, vol. 10, no. 2, Feb. 2022, Art. no. 371, https://doi.org/10.3390/healthcare10020371.

[12] H. Polat, H. Danaei Mehr, and A. Cetin, "Diagnosis of Chronic Kidney Disease Based on Support Vector Machine by Feature Selection Methods," *Journal of Medical Systems*, vol. 41, no. 4, Feb. 2017, Art. no. 55, https://doi.org/10.1007/s10916-017-0703-x.

[13] S. K. Ghosh and A. H. Khandoker, "Investigation on explainable machine learning models to predict chronic kidney diseases," *Scientific Reports*, vol. 14, no. 1, Feb. 2024, Art. no. 3687, https://doi.org/10.1038/s41598-024-54375-4.

[14] M. G. Hegde, B. Ruthvika, R. B. Jain, P. D. Shenoy, K. R. Venugopal, and A. Canchi, "A Privacy-Preserving Federated Learning Method with Homomorphic Encryption for Chronic Kidney Disease Stage Prediction," *Engineering, Technology & Applied Science Research*, vol. 15, no. 4, pp. 26019–26026, Aug. 2025, https://doi.org/10.48084/etasr.11928.