

# A Lightweight Hybrid Model for Early Chronic Kidney Disease Prediction in Resource-Constrained Healthcare

Oluwatobi A. Ayilara-Adewale <sup>1,✉</sup>, Oluwafolake E. Ojo <sup>1</sup>, Johnson O. Olomola <sup>2</sup>, Yusuf O. Olatunde <sup>3</sup>, Kolawole Y. Obiwusi <sup>1</sup>, Damilola Ogunniyi <sup>1</sup>

1. Information Technology, Osun State University, Osogbo, NGA

2. Computer Science, Oduduwa University, Ipetumodu, NGA

3. Cybersecurity, Osun State University, Osogbo, NGA

Received: July 24, 2026 | Review began: August 18, 2026 | Review ended: September 05, 2026 | Published: October 09, 2026

© Copyright 2026

This is an open access article distributed under the terms of the Creative Commons Attribution License CC-BY 4.0., which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

## Abstract

Early diagnosis of chronic kidney disease (CKD) remains challenging because the disease may progress with few symptoms until substantial renal impairment has occurred. This study presents an optimised hybrid framework that integrates principal component analysis (PCA) with an artificial neural network (ANN) for CKD prediction using the University of California, Irvine, CKD dataset. The dataset comprised 400 patient records with 24 clinical predictors and was divided using a stratified 80:20 split into a 320-record development set and an 80-record locked independent hold-out set. Model selection and internal validation were performed using three repetitions of five-fold nested cross-validation, with three-fold inner cross-validation for preprocessing, PCA, ANN, and decision-threshold tuning. Across the 15 outer validation folds, the optimised PCA-ANN achieved a mean accuracy of 97.60%  $\pm$  2.03% (95% CI: 96.48%-98.73%) and a balanced accuracy of 97.69%. On the locked hold-out set, the final model achieved 98.75% accuracy (95% CI: 93.25%-99.78%), 98.00% recall, 98.99% F1-score, and a receiver operating characteristic-area under the curve of 1.000, with one false-negative CKD case. The proposed technique shows promising results for the early detection of CKD but needs to be validated on large-scale, more diverse clinical datasets before deployment in real-world settings

**Categories:** Advances in Biomedical data analysis, AI/ML-based decision support systems, Machine Learning (ML)

**Keywords:** artificial neural networks, chronic kidney disease, machine learning, early disease prediction, principal component analysis, resource-constrained healthcare

## Introduction

Chronic kidney disease (CKD) is one of the most significant non-communicable diseases of today's age and is a growing burden on health systems that are not adequately resourced to deal with the long-term implications of the disease. It is estimated that more than 850 million people are living with some degree of kidney disease, with the most severe impact in low- and middle-income countries, where diagnostic services and early screening programmes are often not available [1,2]. The situation is alarming: a study reported by Francis et al. [3] and Luyckx et al. [4] indicates that if the capacity for detecting CKD is not significantly enhanced, it could become a leading cause of death in the world within 20 years. In addition to mortality, the systemic burden of the disease is significant, and the individual is often affected by cardiovascular disease, hypertension, anaemia, and metabolic complications, which all negatively impact quality of life and the cost of care [5,6]. CKD fundamentally is a progressive disease, with kidney function declining gradually over time, and, if not treated at the clinical stage, eventually reaching end-stage renal disease, which is a fatal stage that requires ongoing renal replacement therapy (dialysis or organ transplantation). The significant financial and logistical costs

### How to cite this article:

Ayilara-Adewale O A, Ojo O E, Olomola J O, et al. (October 09, 2026) A Lightweight Hybrid Model for Early Chronic Kidney Disease Prediction in Resource-Constrained Healthcare. Cureus J Comput Sci 3 : es44389-026-00317-2. DOI <https://doi.org/10.7759/s44389-026-00317-2>

associated with both modalities are impractical for many healthcare systems, especially those in resource-limited settings [5,7]. This is a reality that makes the recognition of CKD at an early stage a public health imperative: CKD can be slowed by detecting it early, before it reaches a stage that may lead to life-threatening complications, and such early recognition can have a meaningful impact on patients' long-term prognosis [3,4].

However, early identification is still difficult, especially in developing regions, where health services are disjointed, the extent of screening has been low, and diagnostic capacity is limited [8,9]. In reality, many patients come to clinical attention when the disease has progressed significantly, and oedema, uncontrolled hypertension, and electrolyte disturbances are all markers of a significant, often irreversible, loss of nephrons [5]. Traditional diagnostic pathways are dependent on biomarkers such as urine albumin-to-creatinine ratios, estimated glomerular filtration rate, and serum creatinine. These markers have important limitations, however: they are not clinically dispensable, they rely on repeated serial testing, and there is a known delay between biological deterioration and the detection of abnormalities. Based on these restrictions, there is a need to supplement traditional clinical workflows with data-driven, computational intelligence-based frameworks that can better identify CKD and stratify patients earlier [10,11]. With recent developments in machine learning, new opportunities have emerged to enhance disease prediction and support clinical decision-making, including supervised, unsupervised, and reinforcement learning paradigms. Machine learning algorithms have demonstrated remarkable utility in clinical diagnostic contexts, most importantly in their capacity to uncover complex, nonlinear relationships embedded within high-dimensional medical datasets that conventional statistical models are fundamentally ill-equipped to resolve [12-14]. Machine learning has demonstrated strong predictive performance in clinical applications, particularly when the relationships among clinical variables are complex and nonlinear. Several studies have reported that machine learning models can achieve better diagnostic accuracy and support the early identification of disease when compared with some conventional analytical methods [15-17]. In addition to their predictive capability, these models can support clinical decision-making by helping practitioners identify patients at risk and make timely intervention decisions. In CKD prediction, researchers have evaluated several algorithms, including decision trees (DT), random forests (RF), support vector machines (SVM), gradient-boosting models, and deep neural networks (DNN).

While the results of the above algorithms are promising, the application of high-dimensional clinical datasets is challenging. Many variables in the clinical dataset are highly correlated or provide the same information. Incorporating these variables into the prediction model increases computational requirements and can lead to data overfitting. On the other hand, principal component analysis (PCA) transforms the original variables into a few orthogonal components based on the proportion of variance. Applying the first components ensures dimensionality reduction of the input while retaining its vital information. Previous research reveals that the application of PCA prior to the classification model reduces redundancy and increases the efficiency of the model. The usage level of the classification model varies with the dataset and classifier used [18,19]. Following dimensionality reduction, an artificial neural network (ANN) can be used to learn nonlinear relationships among the transformed inputs. This ability is useful in CKD prediction because interactions among clinical indicators may not be adequately represented by simpler linear models. In the proposed framework, PCA reduces the number of inputs supplied to the network, while the ANN performs the final classification. Their combination therefore offers a practical approach to developing a CKD prediction model with lower input complexity and the capacity to capture nonlinear patterns in clinical data. When ANN-based classifiers are combined with dimensionality reduction preprocessing, the compounding effect on CKD prediction accuracy has been well documented across recent literature [17-19], a convergence of evidence that directly motivates the hybrid framework proposed in this study.

Motivated by these advancements, this study proposes a hybrid PCA-ANN model for the early prediction of CKD. The PCA technique was used to reduce the dimensionality of the dataset by eliminating redundant clinical features, and the reduced-dimensionality data were then used to train an ANN classifier.

### Review of related works

#### *Traditional Machine Learning Approaches for the Prediction of CKD*

---

#### How to cite this article:

Ayilara-Adewale O A, Ojo O E, Olomola J O, et al. (October 09, 2026) A Lightweight Hybrid Model for Early Chronic Kidney Disease Prediction in Resource-Constrained Healthcare. *Cureus J Comput Sci* 3 : es44389-026-00317-2. DOI <https://doi.org/10.7759/s44389-026-00317-2>

Initial studies on machine learning-based CKD prediction involved only structured clinical data, with the main approaches being supervised classification algorithms. Debal and Sitote [20] applied DT, RF, and SVM classifiers to the UCI CKD (University of California, Irvine Chronic Kidney Disease) dataset, and the RF classifier achieved 96.4% accuracy, confirming that ensemble approaches outperformed individual classifiers. Similarly, Islam et al. [16] tested 12 machine learning classifiers on a clinical dataset of 400 records, and XGBoost demonstrated the highest accuracy of 98.3% for predicting the occurrence of CKD, thus strengthening the applicability of boosted ensemble methods. Likewise, Ganie et al. [21] performed experiments on five boosting models: XGBoost, CatBoost, LightGBM, AdaBoost, and Gradient Boosting, and concluded that AdaBoost demonstrated the best overall performance, with an accuracy of 98.47% and a high receiver operating characteristic-area under the curve (ROC-AUC) value. Arif et al. [18] proposed SVM and RF ensembles with enhanced preprocessing and Boruta feature selection to achieve 96.3% accuracy. Chittora et al. [15] evaluated several machine learning classifiers comprehensively on the UCI CKD dataset and concluded that SVM achieved the best accuracy (96.5%) and that systematic preprocessing plays an important role in improving classification accuracy. As such, these techniques have been used to develop complex architectures for predicting CKD.

### *Deep Learning Approaches for CKD Prediction*

Deep learning techniques have advanced, and they have been applied to create advanced architectures for CKD prediction. Singh et al. [22] proposed a DNN to predict the presence of CKD and used recursive feature elimination to determine the most important clinical predictors. The proposed DNN obtained 100% classification accuracy, which was better than that of the conventional machine learning classifiers evaluated in the study. However, the study reported that the number of clinical datasets was relatively small. With the aim of accurately diagnosing Alzheimer's disease, Alsuhibany et al. [23] proposed an ensemble deep learning-based clinical decision support system in an IoT environment that incorporated convolutional neural networks, gated recurrent units, deep belief networks, and kernel extreme learning machines, with an accuracy of 96.91%. Zhu et al. [24] applied recurrent neural networks to longitudinal electronic health record data to forecast CKD progression and showed good generalisation across a variety of patient groups. Alobaidi [25] reviewed the use of AI-driven diagnostic strategies combining novel biomarkers with deep learning, emphasising the potential of clinical and molecular information for more effective and efficient detection of CKD. Lei et al. [26] also conducted a systematic review and meta-analysis of ensemble learning models across various CKD datasets and reported that ensemble and deep learning models consistently outperformed single classifiers, highlighting the significance of model combination strategies.

### *Hybrid and Feature-Aware Approaches*

Since raw clinical data typically contain redundant and highly correlated features that may decrease model performance, numerous researchers have suggested hybrid methods that integrate feature engineering with classification methods. Khalid et al. [27] proposed a stacking-based hybrid model, which employed Gaussian Naïve Bayes, Gradient Boosting, and DT as base learners and RF as the meta-classifier. Using Pearson correlation for feature selection, their model achieved 100% accuracy on the UCI CKD dataset. Khan et al. [17] specifically examined the integration of PCA with multiple machine learning classifiers and reported an accuracy of 96.1%, showing that dimensionality reduction helped minimise overfitting and improve model generalisation - results that are closely consistent with the present study. In a clinical CKD dataset, Rahman et al. [19] compared the performance of ANN and RF models and concluded that the ANN model, along with feature preprocessing, yielded good precision and recall, with an overall accuracy of 95.8%. Building on this work, Halder et al. [28] applied a machine learning algorithm to electronic health record data to classify patients with a high degree of accuracy and also improved clinical accessibility and usability by incorporating the algorithm into a web-based clinical decision support system.

### *Explainability, Interpretability, and Clinical Deployment*

A growing number of studies have identified the need to consider interpretability and real-world deployment in addition to high predictive accuracy for clinical adoption. Delrue et al. [8] reviewed some of the earlier studies on the use of machine learning in CKD and revealed that, while there has been a significant improvement in classification performance, there are major challenges in generalising to real-world and demographically diverse populations. Likewise, Gogoi and

---

#### **How to cite this article:**

Ayilara-Adewale O A, Ojo O E, Olomola J O, et al. (October 09, 2026) A Lightweight Hybrid Model for Early Chronic Kidney Disease Prediction in Resource-Constrained Healthcare. *Cureus J Comput Sci* 3 : es44389-026-00317-2. DOI <https://doi.org/10.7759/s44389-026-00317-2>

Valan [10] conducted an analysis of machine and deep learning approaches for the diagnosis of kidney stones, highlighting the transparency and explainability of models as major research drawbacks, along with the lack of use of explainable artificial intelligence tools. The systematic review of the literature on the application of machine learning in the diagnosis and treatment of CKD by Sanmarchi et al. [29] also showed that most of the available models lacked standardised reporting, external validation, and consideration of practical deployment, making their application from research settings to clinical practice difficult.

Table 1 summarises the studies reviewed above, outlining the methods and datasets used, the accuracy obtained, and the key limitations of each study.

---

### How to cite this article:

Ayilara-Adewale O A, Ojo O E, Olomola J O, et al. (October 09, 2026) A Lightweight Hybrid Model for Early Chronic Kidney Disease Prediction in Resource-Constrained Healthcare. *Cureus J Comput Sci* 3 : es44389-026-00317-2. DOI <https://doi.org/10.7759/s44389-026-00317-2>

| Author(s)              | Dataset            | Methods                                     | Accuracy (%) | Key Limitation                                       |
|------------------------|--------------------|---------------------------------------------|--------------|------------------------------------------------------|
| Chittora et al. [15]   | UCI CKD dataset    | Multiple ML classifiers (SVM best)          | 96.5         | No dimensionality reduction                          |
| Alsuhibany et al. [23] | UCI CKD + IoT data | CNN-GRU, DBN, KELM ensemble                 | 96.91        | High computational requirement                       |
| Debal and Sitote [20]  | UCI CKD dataset    | Decision Tree, RF, SVM                      | 96.4         | No feature reduction applied                         |
| Singh et al. [22]      | UCI CKD dataset    | Deep Neural Network + RFE feature selection | 100.0        | Small benchmark dataset; external validation lacking |
| Lei et al. [26]        | Multiple datasets  | Meta-analysis of ensemble models            | N/A (review) | Limited to published datasets                        |
| Islam et al. [16]      | UCI CKD dataset    | Multiple ML classifiers (XGBoost best)      | 98.3         | No dimensionality reduction                          |
| Khalid et al. [27]     | UCI CKD dataset    | Stacking hybrid ML                          | 100.0        | Black-box, no interpretability                       |
| Ganie et al. [21]      | UCI CKD dataset    | Boosting ensemble (AdaBoost, XGBoost)       | 98.47        | No feature reduction pipeline                        |
| Sanmarchi et al. [29]  | Literature review  | Systematic review of ML in CKD              | N/A (review) | Highlights reporting gaps                            |
| Zhu et al. [24]        | EHR dataset        | Recurrent Neural Networks                   | 94.7         | Requires longitudinal EHR data                       |
| Arif et al. [18]       | UCI CKD dataset    | SVM, RF ensemble + Boruta feature selection | 96.3         | No PCA preprocessing                                 |
| Halder et al. [28]     | EHR dataset        | ML + web application                        | 95.2         | Limited generalisability                             |
| Khan et al. [17]       | UCI CKD dataset    | PCA + ML classifiers                        | 96.1         | No ANN integration                                   |
| Rahman et al. [19]     | Clinical dataset   | ANN, RF                                     | 95.8         | No dimensionality reduction                          |
| Delrue et al. [8]      | Multiple datasets  | ML review and analysis                      | N/A (review) | Highlights generalisability gaps                     |
| Alobaidi [25]          | Clinical dataset   | AI diagnostics + biomarkers                 | 96.5         | Limited to biomarker features                        |

**How to cite this article:**

Ayilara-Adewale O A, Ojo O E, Olomola J O, et al. (October 09, 2026) A Lightweight Hybrid Model for Early Chronic Kidney Disease Prediction in Resource-Constrained Healthcare. *Cureus J Comput Sci* 3 : es44389-026-00317-2. DOI <https://doi.org/10.7759/s44389-026-00317-2>

|                      |                   |                                |              |                                  |
|----------------------|-------------------|--------------------------------|--------------|----------------------------------|
| Gogoi and Valan [10] | Literature review | Analytical review of ML models | N/A (review) | Identifies interpretability gaps |
|----------------------|-------------------|--------------------------------|--------------|----------------------------------|

TABLE 1: Summary of Related Works on ML-Based CKD Prediction

AI, Artificial Intelligence; ANN, Artificial Neural Network; CKD, Chronic Kidney Disease; CNN-GRU, Convolutional Neural Network-Gated Recurrent Unit; DBN, Deep Belief Network; EHR, Electronic Health Record; KELM, Kernel Extreme Learning Machine; ML, Machine Learning; PCA, Principal Component Analysis; RF, Random Forest; RFE, Recursive Feature Elimination; SVM, Support Vector Machine; UCI, University of California, Irvine; XGBoost, Extreme Gradient Boosting

Research gaps and motivation

Although significant improvements have been documented in these studies, there are still a few significant limitations. To start with, most of the models are trained and tested only on small, homogeneous, publicly available datasets, which casts serious doubts on their capacity to generalise to diverse real-world clinical populations [8]. Second, the vast majority of the most successful models, in particular, deep learning and ensemble-based models, are black-box systems with limited clinical interpretability, which limits their real-world use [10,29]. Third, despite the suggestion of feature selection techniques in some studies, very few studies have adequately examined the impact of feature reduction and neural network optimisation on prediction accuracy and computational expense in a single framework. Last but not least, the computing resources required for many deep and ensemble models are high, and they are therefore challenging to implement in low-resource healthcare settings, especially in sub-Saharan Africa and other developing countries where the prevalence of CKD is high and diagnostic services are lacking.

Contributions

This study aimed to address these gaps by developing a machine learning model for the early prediction of CKD to enhance accuracy and computational efficiency in resource-limited healthcare settings. The major contributions of this study are as follows:

- i) A hybrid PCA-ANN framework is proposed to reduce feature redundancy and computational complexity by integrating PCA with ANN to enhance the early prediction of CKD.
- ii) A systematic, leakage-safe tuning procedure is used to optimise preprocessing, PCA, ANN architecture, regularisation, and the decision threshold. In the final development-set fit, the selected PCA configuration retained 20 principal components and 94.24% of the variance in the preprocessed feature space.
- iii) The proposed framework is evaluated using three repetitions of five-fold nested cross-validation, producing 15 outer-fold estimates for accuracy, balanced accuracy, precision, recall, F1-score, and ROC-AUC. Mean values, standard deviations, and 95% confidence intervals are reported, and the final model is evaluated once on a locked independent 20% hold-out set.
- iv) The framework is designed as a compact prediction pipeline for resource-constrained settings; however, real-world computational, clinical, and deployment benefits were not directly evaluated and require prospective validation.

Materials And Methods

The primary goal of the present study is to present an efficient machine learning framework with a dimensionality reduction approach for the early detection of CKD using a publicly available dataset. The proposed method is based on data preprocessing, the PCA dimensionality reduction technique to extract features from the dataset, and classification using ANN. The framework is developed in Python using the following libraries: Pandas, Scikit-learn, TensorFlow, and Keras.

How to cite this article:

Dataset

This study used data from the UCI Machine Learning Repository [30]. The dataset contains 400 patient records with 24 predictor variables and one binary target variable indicating CKD or non-CKD status. The predictors include demographic information, clinical markers of kidney function, and laboratory values such as age, blood pressure, serum creatinine, haemoglobin, diabetes mellitus status, and indicators of anaemia. The main attributes are summarised in Table 2.

| Attribute               | Description                          | Data Type       |
|-------------------------|--------------------------------------|-----------------|
| Age                     | Age of the patient                   | Numerical       |
| Blood Pressure          | Diastolic blood pressure (mm/Hg)     | Numerical       |
| Specific Gravity        | Urine density measurement            | Numerical       |
| Albumin                 | Albumin level in urine               | Numerical       |
| Sugar                   | Glucose concentration in urine       | Numerical       |
| Red Blood Cells         | Presence of red blood cells in urine | Categorical     |
| Pus Cell                | Presence of pus cells in urine       | Categorical     |
| Blood Glucose Random    | Random blood glucose measurement     | Numerical       |
| Blood Urea              | Blood urea concentration             | Numerical       |
| Serum Creatinine        | Kidney filtration indicator          | Numerical       |
| Sodium                  | Blood sodium concentration           | Numerical       |
| Potassium               | Blood potassium concentration        | Numerical       |
| Haemoglobin             | Haemoglobin level in blood           | Numerical       |
| Diabetes Mellitus       | Presence of diabetes                 | Categorical     |
| Coronary Artery Disease | Presence of cardiovascular disease   | Categorical     |
| Anaemia                 | Indicator of anaemia                 | Categorical     |
| Class                   | CKD/Not CKD                          | Target Variable |

TABLE 2: Description of Major Attributes in the CKD Dataset

CKD, Chronic Kidney Disease

How to cite this article:

### Data preprocessing

All preprocessing operations were fitted within the relevant training data to prevent information leakage. Candidate pipelines included median imputation with missingness indicators for numerical variables, most-frequent imputation for categorical variables, one-hot or ordinal encoding, and robust, power-transformed, or standard scaling. The final selected PCA-ANN configuration used one-hot encoding with robust scaling. The same fitted preprocessing transformation was then applied to the corresponding validation data and, after final model selection, to the locked hold-out set.

### Dimensionality reduction (PCA)

Correlated variables are often present in medical data, leading to feature redundancy and a high-dimensional feature space that can adversely affect computation time and overfitting. This problem can be addressed using PCA, in which the original variables are transformed into a set of orthogonal principal components. To prevent data leakage, PCA was fitted only within the training portion of each inner or outer cross-validation fold. The PCA configuration was treated as a tunable component rather than fixed in advance: candidate settings included variance-retention thresholds from 80% to 99% and fixed component counts from 8 to 20, with whitening also considered. After nested model selection on the development data, the final selected configuration used 20 principal components with whitening and retained 94.24% of the variance in the preprocessed development-set feature space. The fitted PCA transformation was applied unchanged to validation data and, after final model selection, to the locked hold-out set.

### System architecture

The revised system architecture separates model development from final independent testing. Figure 1 presents the overall validation and evaluation workflow. The 400-record dataset is first divided using a stratified 80:20 split into a 320-record development set and an 80-record locked hold-out set. The development set is used exclusively for model selection and internal validation through three repetitions of five-fold outer cross-validation with three-fold inner cross-validation. The preprocessing stage shown in Figure 1 represents the logical preprocessing step; all fitted preprocessing transformations, PCA operations, ANN tuning, and decision-threshold selection are learned only from the relevant training portion of each fold, as described in the Sections "Data preprocessing" and "Dimensionality reduction (PCA)", to prevent information leakage. Balanced accuracy is used as the primary tuning criterion. After the optimised configuration is selected, the final pipeline is fitted on the complete development set and evaluated once on the previously unseen hold-out set. Figure 2 provides the complementary system-level architecture, showing the data pipeline, stratified train-test separation, repeated nested cross-validation training and tuning modules, outer-fold performance evaluation, the selected PCA-ANN configuration, and the final prediction stage on the locked hold-out set.

---

#### How to cite this article:

Ayilara-Adewale O A, Ojo O E, Olomola J O, et al. (October 09, 2026) A Lightweight Hybrid Model for Early Chronic Kidney Disease Prediction in Resource-Constrained Healthcare. *Cureus J Comput Sci* 3 : es44389-026-00317-2. DOI <https://doi.org/10.7759/s44389-026-00317-2>

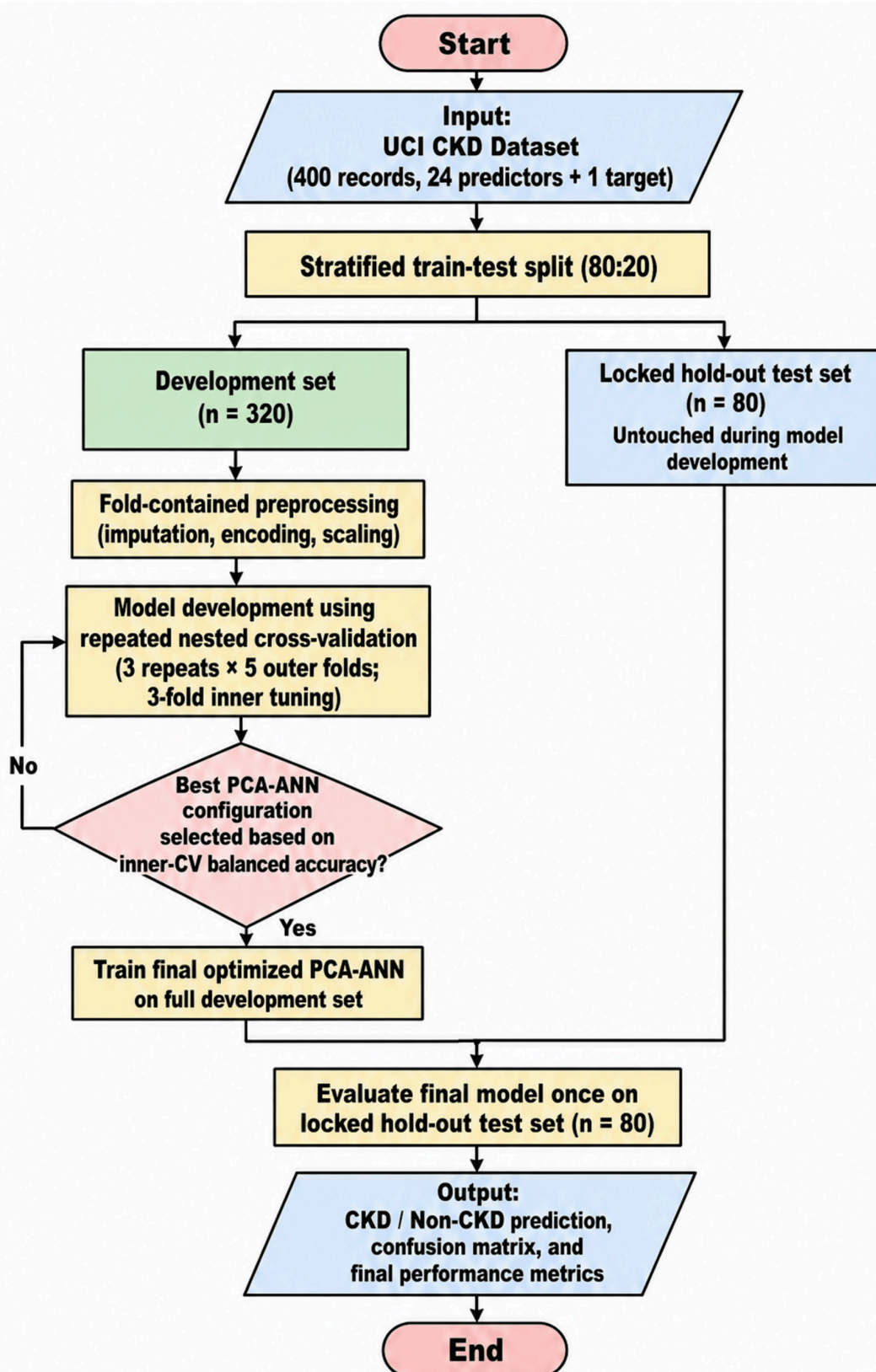


**Figure 1.** Workflow of the optimized PCA-ANN model development and independent hold-out evaluation for chronic kidney disease prediction.

#### How to cite this article:

Ayilara-Adewale O A, Ojo O E, Olomola J O, et al. (October 09, 2026) A Lightweight Hybrid Model for Early Chronic Kidney Disease Prediction in Resource-Constrained Healthcare. *Cureus J Comput Sci* 3 : es44389-026-00317-2. DOI <https://doi.org/10.7759/s44389-026-00317-2>

FIGURE 1: Workflow for the Optimised PCA-ANN Framework

ANN, Artificial Neural Network; PCA, Principal Component Analysis; UCI CKD, University of California, Irvine Chronic Kidney Disease dataset

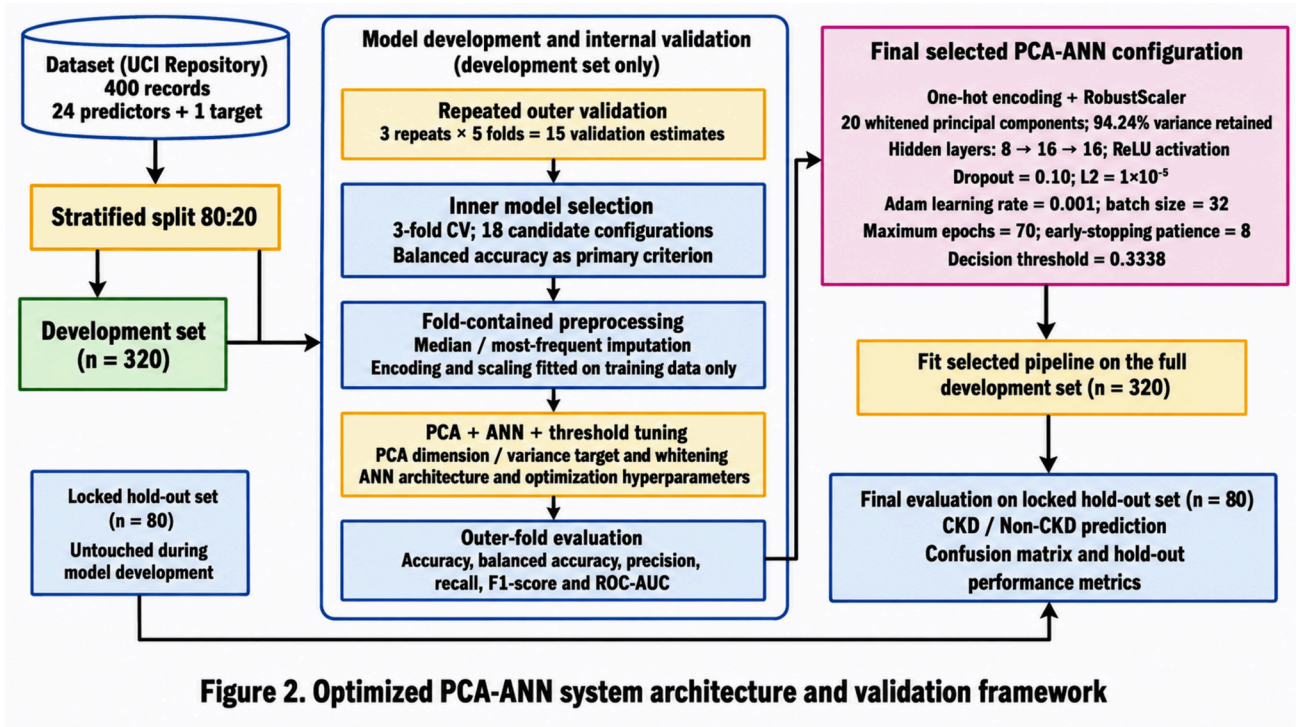


FIGURE 2: Optimised PCA-ANN System Architecture and Validation Framework

ANN, Artificial Neural Network; CKD, Chronic Kidney Disease; CV, Cross-Validation; PCA, Principal Component Analysis; ReLU, Rectified Linear Unit; ROC-AUC, Receiver Operating Characteristic–Area Under the Curve; UCI, University of California, Irvine

Figures 1-2 provide complementary views of the proposed framework: Figure 1 emphasises the end-to-end validation workflow and the strict separation of the locked hold-out set, whereas Figure 2 details the system architecture and the interactions among preprocessing, PCA transformation, nested model development, hyperparameter tuning, outer-fold evaluation, final model selection, and hold-out prediction. Algorithm 1 formalises these stages as a leakage-safe development and evaluation procedure. The UCI CKD dataset is first separated into predictors and target labels and then split into a 320-record development set and an 80-record locked hold-out set. Within the development set, three repetitions of five-fold outer cross-validation are used for performance estimation. Each outer training fold contains a three-fold inner cross-validation procedure that selects preprocessing, PCA, ANN hyperparameters, and a decision threshold using balanced accuracy as the primary criterion. The selected configuration is then refitted on the outer training fold and evaluated on the corresponding outer validation fold. After all 15 outer folds are completed, mean performance, standard deviation, and 95% confidence intervals are calculated. A final configuration is selected using only the full development set, fitted once, and then evaluated on the untouched hold-out set to generate the final confusion matrix and test metrics.

How to cite this article:

### **ANN classification**

The optimised ANN receives the PCA-transformed development data and produces a CKD probability through a sigmoid output neuron. Candidate architectures, activation functions, dropout, L2 regularisation, learning rate, batch size, preprocessing strategy, PCA configuration, and classification threshold were evaluated within the inner cross-validation loop. Model training used the Adam optimiser and binary cross-entropy loss, with a maximum of 70 epochs and early stopping to limit overfitting. The final development-set tuning selected one-hot encoding with robust scaling, 20 whitened principal components, three hidden layers containing 8, 16, and 16 neurons, ReLU activation, dropout of 0.10, L2 regularisation of  $1 \times 10^{-5}$ , a learning rate of 0.001, a batch size of 32, and a locked classification threshold of 0.3338. The final model configuration is depicted within the system architecture in Figure 2 and summarised numerically in Table 3.

---

### **How to cite this article:**

Ayilara-Adewale O A, Ojo O E, Olomola J O, et al. (October 09, 2026) A Lightweight Hybrid Model for Early Chronic Kidney Disease Prediction in Resource-Constrained Healthcare. *Cureus J Comput Sci* 3 : es44389-026-00317-2. DOI <https://doi.org/10.7759/s44389-026-00317-2>

| Hyperparameter            | Optimal Value                     |
|---------------------------|-----------------------------------|
| Preprocessing             | One-hot encoding + robust scaling |
| PCA components            | 20                                |
| PCA whitening             | Yes                               |
| Variance retained         | 94.24%                            |
| Number of hidden layers   | 3                                 |
| Neurons - Layer 1         | 8                                 |
| Neurons - Layer 2         | 16                                |
| Neurons - Layer 3         | 16                                |
| Activation (hidden)       | ReLU                              |
| Activation (output)       | Sigmoid                           |
| Learning rate             | 0.001                             |
| Optimiser                 | Adam                              |
| Batch size                | 32                                |
| Maximum epochs            | 70                                |
| Early-stopping patience   | 8                                 |
| Dropout rate              | 0.10                              |
| L2 regularisation         | $1 \times 10^{-5}$                |
| Locked decision threshold | 0.3338                            |
| Loss function             | Binary cross-entropy              |

**TABLE 3: Optimal ANN Hyperparameters**

ANN, Artificial Neural Network; ReLU, Rectified Linear Unit; PCA, Principal Component Analysis

**How to cite this article:**

Ayilara-Adewale O A, Ojo O E, Olomola J O, et al. (October 09, 2026) A Lightweight Hybrid Model for Early Chronic Kidney Disease Prediction in Resource-Constrained Healthcare. Cureus J Comput Sci 3 : es44389-026-00317-2. DOI <https://doi.org/10.7759/s44389-026-00317-2>

Evaluation metrics

Model performance was evaluated using accuracy, balanced accuracy, precision, recall, F1-score, and ROC-AUC. After the initial stratified 80:20 split, the 320-record development set was evaluated using three repetitions of five-fold nested cross-validation. The outer loop therefore produced 15 validation estimates, while three-fold inner cross-validation was used exclusively for model and threshold selection. For each repeated nested cross-validation metric, the mean, sample standard deviation, and two-sided 95% confidence interval were calculated across the 15 outer-fold estimates using the Student's t-distribution. Because outer-fold estimates from repeated cross-validation are correlated, these intervals are interpreted as descriptive uncertainty summaries rather than fully independent-sample confidence intervals. For the model comparison, matched outer-fold accuracies were analysed using a resampled t-test with variance adjustment for the overlap among cross-validation training sets; the adjustment used the outer validation-to-training ratio of  $64/256 = 0.25$ . The mean accuracy difference and its 95% confidence interval were reported. The 80-record hold-out set was not used during tuning or cross-validation and was evaluated only once after the final pipeline and decision threshold had been locked. A Wilson 95% confidence interval was additionally reported for hold-out accuracy.

Results And Discussion

Experimental results

The optimised PCA-ANN was evaluated using repeated nested cross-validation on the development set and a separate locked independent hold-out test set. Table 4 reports the mean, standard deviation, and 95% confidence interval for each performance metric across the 15 outer validation folds and compares the optimised PCA-ANN with a leakage-safe feature-selected ANN evaluated under the same  $3 \times 5$  repeated nested cross-validation design. Table 5 presents the confusion matrix obtained only from the 80-record locked hold-out set; it does not represent a cross-validation fold. The final PCA-ANN correctly classified 79 of 80 hold-out records, corresponding to an accuracy of 98.75% (Wilson 95% CI: 93.25%-99.78%).

| Metric                | Feature-Selected ANN            | Optimised PCA-ANN               |
|-----------------------|---------------------------------|---------------------------------|
| Accuracy (%)          | 96.04 ± 4.53 (93.53-98.55)      | 97.60 ± 2.03 (96.48-98.73)      |
| Balanced accuracy (%) | 96.00 ± 4.07 (93.75-98.25)      | 97.69 ± 1.97 (96.61-98.78)      |
| Precision (%)         | 97.65 ± 3.32 (95.81-99.49)      | 98.84 ± 1.61 (97.94-99.73)      |
| Recall (%)            | 96.17 ± 7.37 (92.08-100.00)     | 97.33 ± 2.75 (95.81-98.86)      |
| F1-score (%)          | 96.70 ± 3.99 (94.48-98.91)      | 98.06 ± 1.66 (97.14-98.98)      |
| ROC-AUC               | 0.9982 ± 0.0024 (0.9969-0.9995) | 0.9977 ± 0.0046 (0.9952-1.0000) |

TABLE 4: Repeated Nested Cross-Validation Performance With 95% Confidence Intervals

Values are presented as mean ± standard deviation (95% CI).  
PCA-ANN, Principal Component Analysis-Artificial Neural Network; ROC-AUC, Receiver Operating Characteristic-Area Under the Curve

How to cite this article:

Ayilara-Adewale O A, Ojo O E, Olomola J O, et al. (October 09, 2026) A Lightweight Hybrid Model for Early Chronic Kidney Disease Prediction in Resource-Constrained Healthcare. Cureus J Comput Sci 3 : es44389-026-00317-2. DOI <https://doi.org/10.7759/s44389-026-00317-2>

|                | Predicted CKD | Predicted Non-CKD |
|----------------|---------------|-------------------|
| Actual CKD     | 49            | 1                 |
| Actual Non-CKD | 0             | 30                |

**TABLE 5: Confusion Matrix for the Optimised PCA-ANN on the Locked Independent Hold-Out Test Set (n = 80)**

CKD, Chronic Kidney Disease; PCA-ANN, Principal Component Analysis-Artificial Neural Network

The hold-out confusion matrix is visualised in Figure 3. The final PCA-ANN produced 49 true positives, 30 true negatives, no false positives, and 1 false negative. Hold-out recall was 98.00%, F1-score was 98.99%, and ROC-AUC was 1.000; the perfect hold-out AUC should be interpreted cautiously because the test set contained only 80 records.

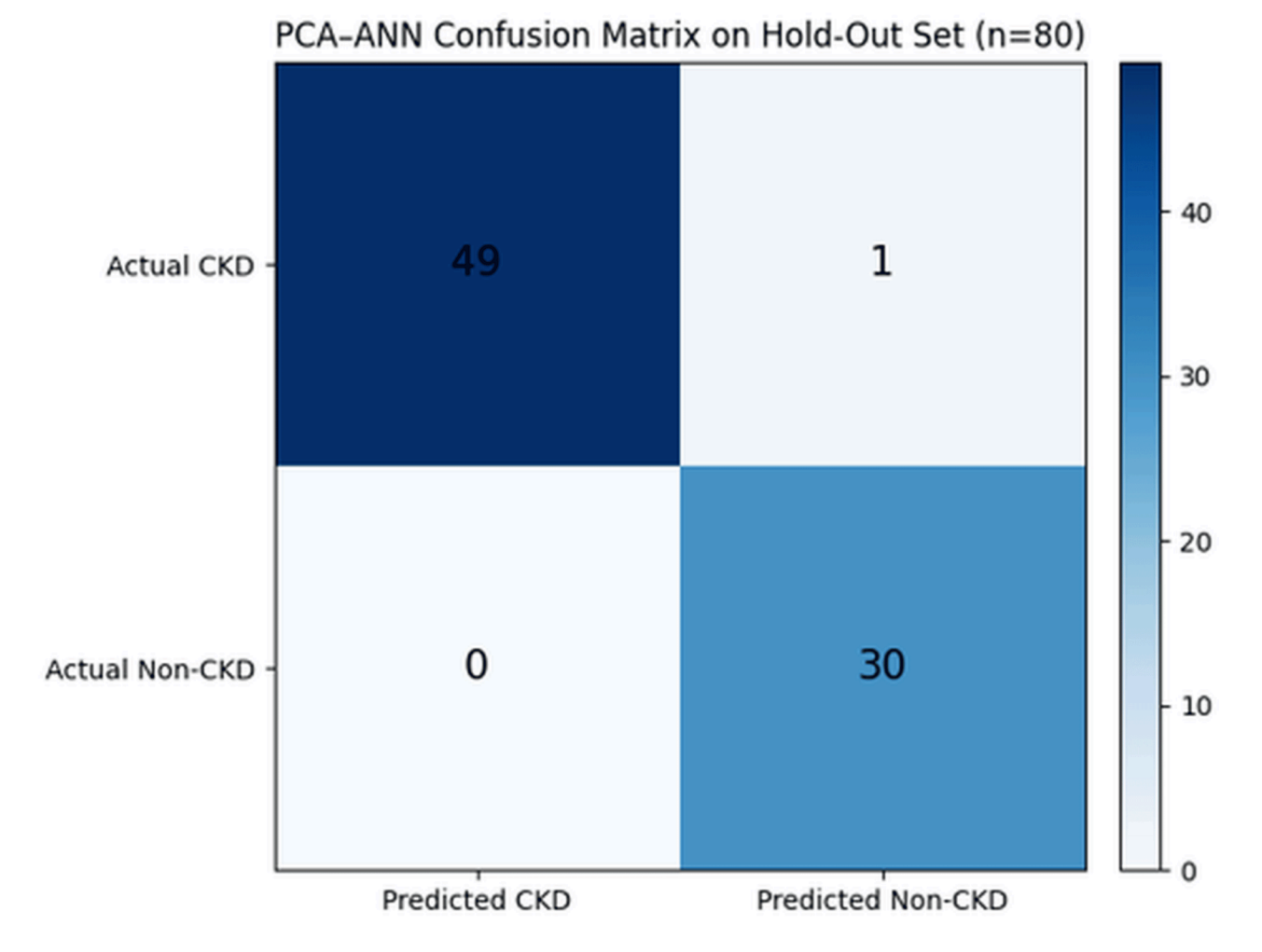


**FIGURE 3: Confusion Matrix for the Optimised PCA-ANN on the Locked Independent Hold-Out Set (n = 80)**

CKD, Chronic Kidney Disease; PCA-ANN, Principal Component Analysis-Artificial Neural Network

**How to cite this article:**

Figure 4 presents the precision-recall curve for the optimised PCA-ANN on the locked independent hold-out set. The model achieved an average precision of 1.000 across the observed probability ranking. At the locked classification threshold of 0.3338, precision was 100.00% and recall was 98.00%, corresponding to zero false-positive predictions and one false-negative CKD prediction. The perfect average precision value reflects the ranking observed in this 80-record hold-out sample and should be interpreted cautiously in the absence of external validation.

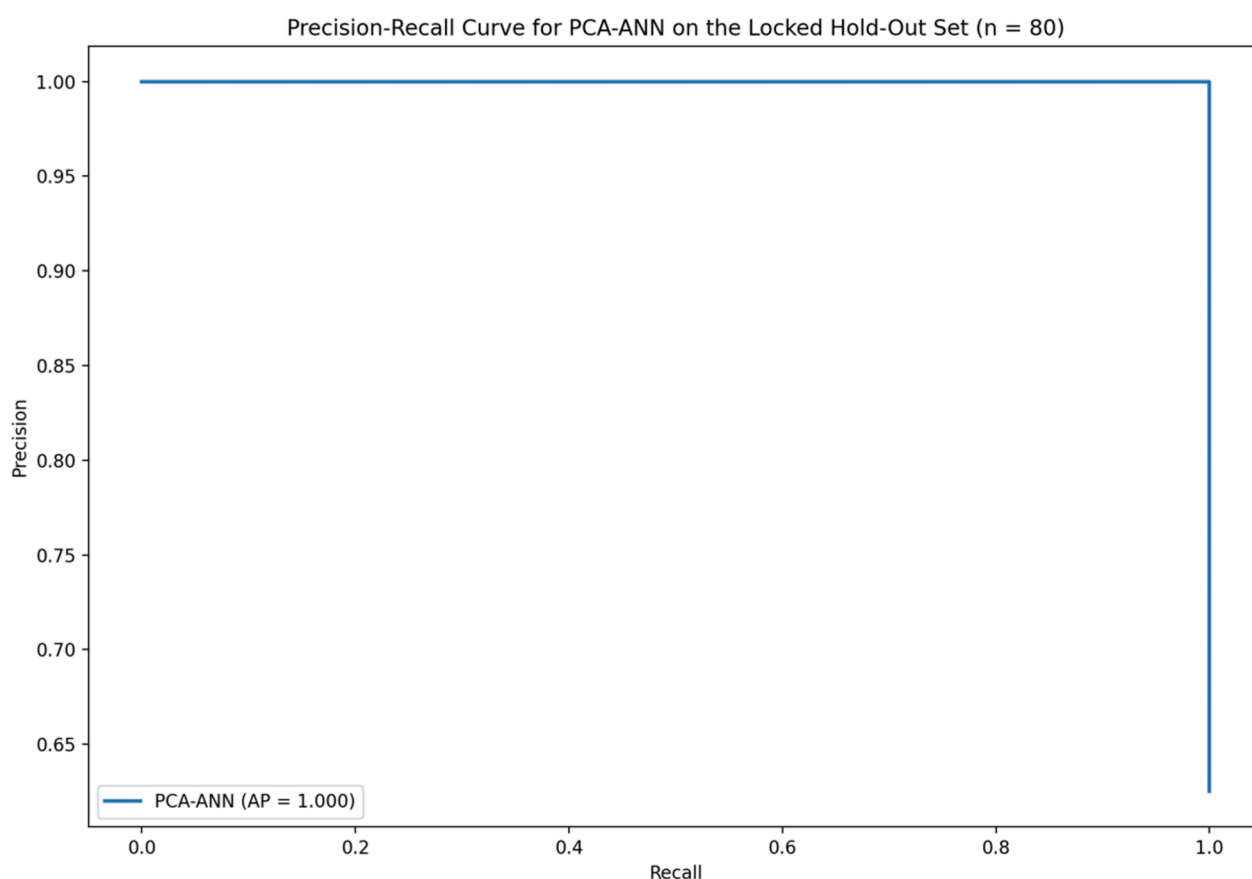


**FIGURE 4: Precision-Recall Curve for the Optimised PCA-ANN on the Locked Independent Hold-Out Set (n = 80)**

AP, Average Precision; PCA-ANN, Principal Component Analysis-Artificial Neural Network

Figure 5 shows the ROC curve for the optimised PCA-ANN on the locked independent hold-out set. The ROC-AUC was 1.000, indicating complete ranking separation between CKD and non-CKD records in this sample. This threshold-independent result is compatible with the single false-negative prediction at the locked threshold because ROC-AUC evaluates score ranking across all possible thresholds rather than performance at one decision threshold. Given the hold-out size of 80 records, the observed AUC should not be interpreted as evidence of perfect generalisation.

#### How to cite this article:

Ayilara-Adewale O A, Ojo O E, Olomola J O, et al. (October 09, 2026) A Lightweight Hybrid Model for Early Chronic Kidney Disease Prediction in Resource-Constrained Healthcare. *Cureus J Comput Sci* 3 : es44389-026-00317-2. DOI <https://doi.org/10.7759/s44389-026-00317-2>

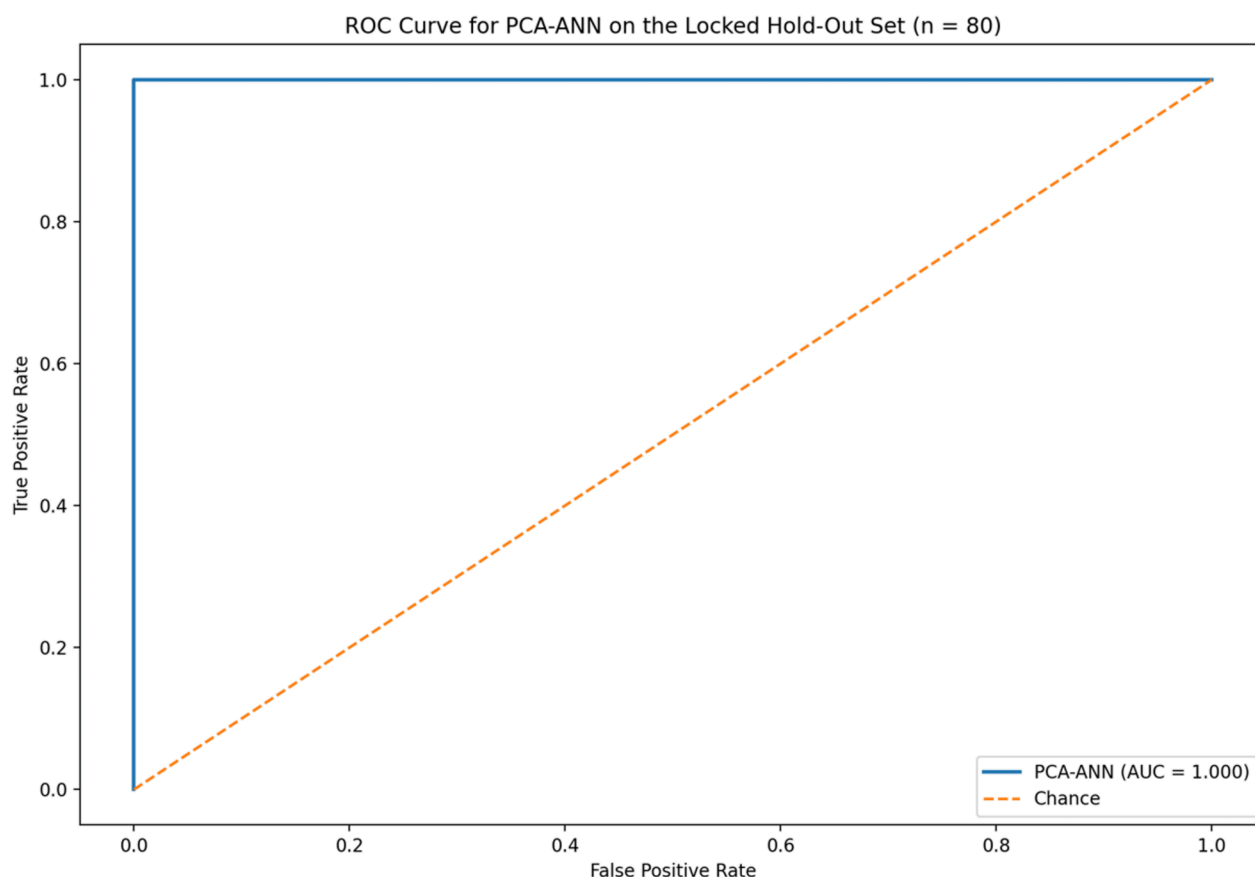


**FIGURE 5: ROC Curve for the Optimised PCA-ANN on the Locked Independent Hold-out Set**

AUC, Area Under the Curve; PCA-ANN, Principal Component Analysis-Artificial Neural Network; ROC, Receiver Operating Characteristic

Figure 6 shows the training and validation accuracy trajectories for the PCA-ANN model. Accuracy increased rapidly during the initial epochs, after which both trajectories approached their maximum values and remained closely aligned. The absence of a sustained separation between the training and validation curves indicates stable convergence during model fitting.

#### How to cite this article:

Ayilara-Adewale O A, Ojo O E, Olomola J O, et al. (October 09, 2026) A Lightweight Hybrid Model for Early Chronic Kidney Disease Prediction in Resource-Constrained Healthcare. *Cureus J Comput Sci* 3 : es44389-026-00317-2. DOI <https://doi.org/10.7759/s44389-026-00317-2>

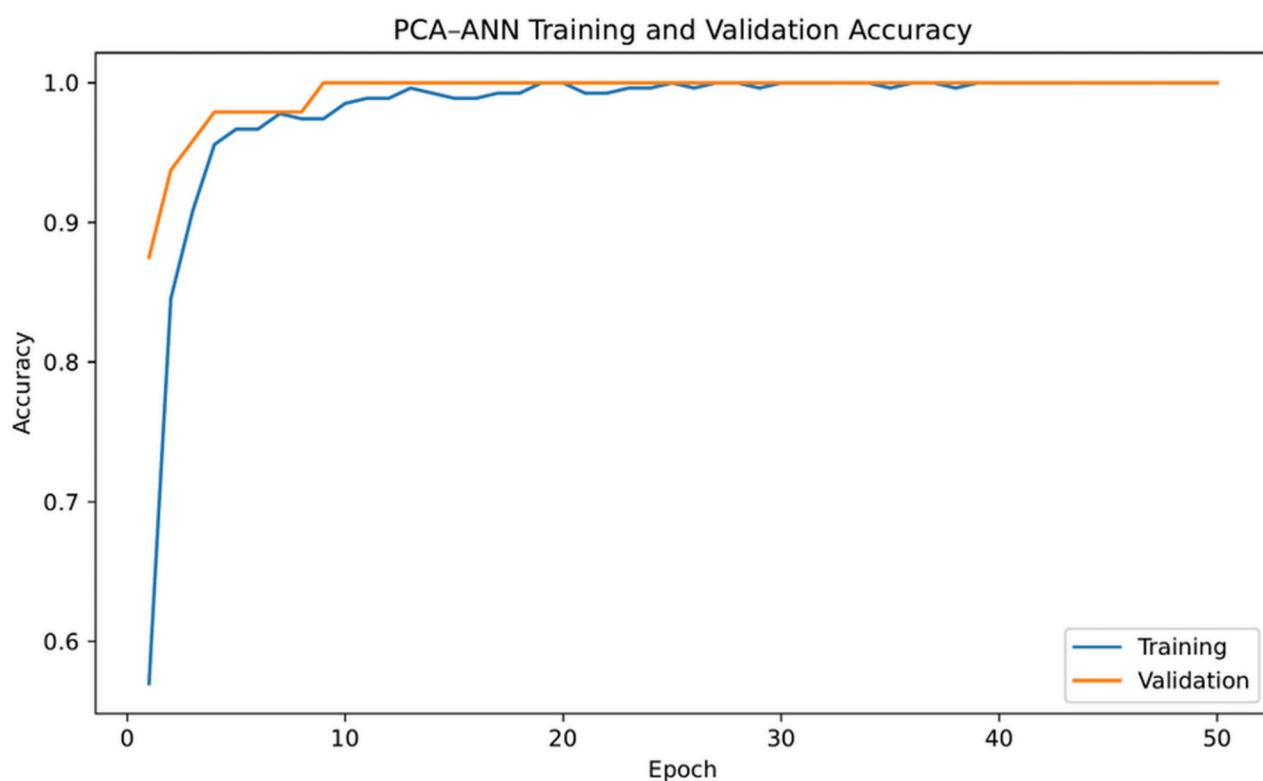


**FIGURE 6: Training and Validation Accuracy Curve for the PCA-ANN Model**

PCA-ANN, Principal Component Analysis-Artificial Neural Network

Figure 7 presents the corresponding binary cross-entropy loss trajectories. Training and validation loss decreased sharply during the early epochs and continued towards low values as training progressed. The validation-loss trajectory remained close to the training-loss trajectory without a pronounced late increase, providing a complementary indication of stable optimisation during the recorded training history.

#### How to cite this article:

Ayilara-Adewale O A, Ojo O E, Olomola J O, et al. (October 09, 2026) A Lightweight Hybrid Model for Early Chronic Kidney Disease Prediction in Resource-Constrained Healthcare. *Cureus J Comput Sci* 3 : es44389-026-00317-2. DOI <https://doi.org/10.7759/s44389-026-00317-2>

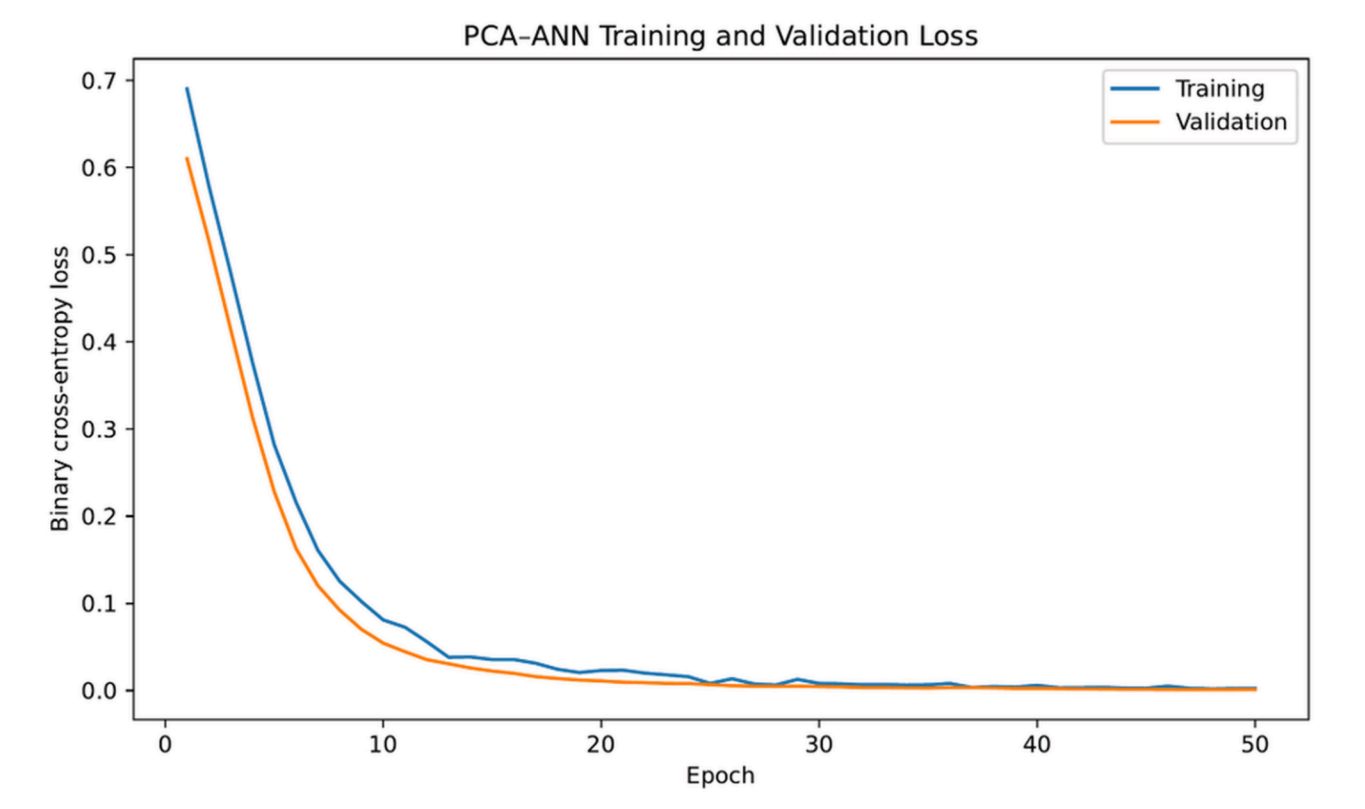


**FIGURE 7: Training and Validation Loss Curve for the PCA-ANN Model**

PCA-ANN, Principal Component Analysis-Artificial Neural Network

The independent hold-out results for the optimised PCA-ANN and the leakage-safe feature-selected ANN are summarised in Table 6. These results are reported separately from the repeated nested-cross-validation estimates because the hold-out set was evaluated only once after model selection.

**How to cite this article:**

Ayilara-Adewale O A, Ojo O E, Olomola J O, et al. (October 09, 2026) A Lightweight Hybrid Model for Early Chronic Kidney Disease Prediction in Resource-Constrained Healthcare. Cureus J Comput Sci 3 : es44389-026-00317-2. DOI <https://doi.org/10.7759/s44389-026-00317-2>

| Metric                | Feature-selected ANN        | Optimised PCA-ANN           |
|-----------------------|-----------------------------|-----------------------------|
| Accuracy (%)          | 95.00 (95% CI: 87.84-98.04) | 98.75 (95% CI: 93.25-99.78) |
| Balanced accuracy (%) | 96.00                       | 99.00                       |
| Precision (%)         | 100.00                      | 100.00                      |
| Recall (%)            | 92.00                       | 98.00                       |
| F1-score (%)          | 95.83                       | 98.99                       |
| ROC-AUC               | 0.9813                      | 1.0000                      |
| False negatives       | 4                           | 1                           |

TABLE 6: Locked Independent Hold-Out Performance Comparison

PCA-ANN, Principal Component Analysis-Artificial Neural Network; ROC-AUC, Receiver Operating Characteristic-Area Under the Curve

A paired analysis of the 15 matched outer-fold accuracy estimates was performed to quantify the uncertainty in the observed difference between the optimised PCA-ANN and the leakage-safe feature-selected ANN. The results are presented in Table 7, where the optimised PCA-ANN had an accuracy of 1.56% over the feature-selected ANN, although the 95% confidence interval is between  $-4.65\%$  and  $7.78\%$  ( $t = 0.54$ ,  $p = 0.598$ ), which means that the difference is not statistically significant at  $\alpha = 0.05$ .

| Comparison                                | Mean Accuracy Difference (percentage points) | 95% CI            | t-statistic | p-value |
|-------------------------------------------|----------------------------------------------|-------------------|-------------|---------|
| Optimised PCA-ANN vs Feature-selected ANN | +1.56                                        | $-4.65$ to $7.78$ | 0.54        | 0.598   |

TABLE 7: Paired Comparison of Repeated Nested Cross-Validation Accuracy

PCA-ANN, Principal Component Analysis-Artificial Neural Network

A comparative evaluation with selected existing studies is provided in Table 8. The optimised PCA-ANN achieved a mean repeated nested cross-validation accuracy of  $97.60\% \pm 2.03\%$  (95% CI: 96.48%-98.73%) and a mean balanced accuracy of  $97.69\% \pm 1.97\%$  (95% CI: 96.61%-98.78%). The leakage-safe feature-selected ANN achieved  $96.04\% \pm 4.53\%$  accuracy (95% CI: 93.53%-98.55%) and  $96.00\% \pm 4.07\%$  balanced accuracy (95% CI: 93.75%-98.25%). The PCA-ANN also showed higher mean precision, recall, and F1-score across the 15 outer validation folds, whereas the feature-selected ANN had a marginally higher mean ROC-AUC (0.9982 versus 0.9977). The overlapping confidence intervals indicate that these

How to cite this article:

relatively small cross-validation differences should not be interpreted as evidence of model superiority. On the locked independent hold-out set, the optimised PCA-ANN correctly classified 79 of 80 records, giving 98.75% accuracy compared with 95.00% for the feature-selected ANN. The optimised PCA-ANN achieved 98.00% recall, 98.99% F1-score, and an ROC-AUC of 1.000, while the feature-selected ANN achieved 92.00% recall, 95.83% F1-score, and an ROC-AUC of 0.9813. The number of false-negative CKD predictions decreased from four with the feature-selected ANN to one with the PCA-ANN. These hold-out results favour the PCA-ANN in this dataset, particularly for CKD sensitivity, but the small sample size and the single hold-out evaluation require cautious interpretation.

| Study                 | Method                  | Accuracy (%) |
|-----------------------|-------------------------|--------------|
| Chittora et al. [15]  | Multiple ML classifiers | 96.5         |
| Debal and Sitote [20] | Decision Tree, RF, SVM  | 96.4         |
| Khalid et al. [27]    | Hybrid ML               | 100.0        |
| Proposed Model        | PCA-ANN                 | 97.6         |

TABLE 8: Comparison With Existing Studies

ML, Machine Learning; PCA-ANN, Principal Component Analysis-Artificial Neural Network; RF, Random Forest; SVM, Support Vector Machine

Discussion

The statistical analysis did not demonstrate significant superiority of the optimised PCA-ANN over the leakage-safe feature-selected ANN. Across the 15 matched outer validation folds, the PCA-ANN showed a mean accuracy advantage of 1.56 percentage points; however, the adjusted 95% confidence interval ranged from -4.65 to 7.78 percentage points ( $t = 0.54$ ,  $p = 0.598$ ). The findings therefore indicate comparable high predictive performance, with modest observed advantages for the PCA-ANN in accuracy, balanced accuracy, recall, and F1-score. On the independent hold-out set, the PCA-ANN achieved 98.75% accuracy and missed one CKD case, compared with four missed cases for the feature-selected ANN. Its ROC-AUC of 1.000 should nevertheless be interpreted cautiously because the hold-out set contained only 80 records.

To reduce the risk of overfitting, the ANN incorporated dropout, L2 regularisation, and early stopping. All preprocessing, PCA transformation, hyperparameter optimisation, and decision-threshold selection were also confined to the appropriate training folds within the repeated nested cross-validation framework. The close alignment of the training and validation curves, together with the comparable cross-validation and hold-out performance, did not indicate a substantial generalisation gap within this dataset. However, the possibility of overfitting cannot be completely excluded because the study used a small, single-source dataset. The final PCA configuration reduced the preprocessed feature space to 20 components while retaining 94.24% of its variance. This reduced the model’s internal representation but not the number of raw clinical measurements required. Consequently, potential reductions in laboratory requirements, computational cost, and healthcare expenditure were not directly established. External and prospective validation are required to confirm the model’s generalisability, clinical utility, and suitability for resource-constrained healthcare settings.

How to cite this article:

## Limitations

This study has several limitations. First, the analysis was based on a single publicly available benchmark dataset containing 400 records from one source. Consequently, the dataset may not adequately represent the demographic, geographic, and clinical diversity encountered across different healthcare settings. Although three repetitions of nested five-fold cross-validation produced 15 outer validation estimates, these estimates remain correlated, and their confidence intervals should be interpreted as descriptive measures of uncertainty. The locked independent hold-out set also contained only 80 records; therefore, the observed accuracy of 98.75% and ROC-AUC of 1.000 may be optimistic and should not be generalised beyond the present dataset. Furthermore, the comparison with the feature-selected ANN did not establish statistically significant superiority of the PCA-ANN. No external, multicentre, prospective, explainability, or real-world deployment evaluation was conducted, and the computational benefits suggested by the lightweight architecture were not directly measured.

Future research should validate the framework using larger, externally sourced, and multicentre clinical datasets representing diverse patient populations. Such studies should assess both discrimination and calibration and determine whether model recalibration is required in new populations. Explainability methods, including SHapley Additive exPlanations analysis, permutation-based feature importance, and interpretation of PCA loadings, should also be investigated to improve clinical transparency. Finally, model latency, memory consumption, hardware requirements, workflow integration, and effects on clinical decision-making should be evaluated prospectively before the framework is considered for routine use in resource-constrained healthcare settings.

## Conclusions

This study evaluated an optimised PCA-ANN framework for CKD prediction using a leakage-safe validation design. The dataset was divided into a 320-record development set and an 80-record locked independent hold-out set. Model development used three repetitions of five-fold nested cross-validation with three-fold inner tuning. The final selected configuration used 20 principal components with whitening and retained 94.24% of the variance in the preprocessed development-set feature space. Across the 15 outer validation folds, the optimised PCA-ANN achieved  $97.60\% \pm 2.03\%$  accuracy (95% CI: 96.48%-98.73%) compared with  $96.04\% \pm 4.53\%$  for the leakage-safe feature-selected ANN. The observed mean difference was 1.56 percentage points, but the repeated cross-validation comparison was not statistically significant (95% CI: -4.65 to 7.78;  $p = 0.598$ ). On the locked hold-out set, the PCA-ANN achieved 98.75% accuracy, 98.00% recall, 98.99% F1-score, and one false-negative CKD prediction, compared with 95.00% accuracy, 92.00% recall, 95.83% F1-score, and four false negatives for the feature-selected ANN.

These results support retaining PCA-ANN as a promising prediction framework because it showed high and relatively stable internal validation performance together with favourable sensitivity on the independent hold-out sample. Nevertheless, the study does not establish statistically significant superiority or perfect generalisation. Validation using larger, external, multicentre clinical datasets is required before the model can be recommended for routine clinical use, particularly in resource-constrained healthcare settings.

## Appendices

Algorithm 1: Hybrid PCA-ANN CKD Prediction Framework

- 1: Load UCI CKD dataset  $D$  and separate predictors  $X$  from target  $y$
- 2: Stratified split: development set = 80%; locked hold-out set = 20%
- 3: // Repeated nested cross-validation on development set
- 4: for repeat = 1 to 3 do
- 5:   Create 5 stratified outer folds

---

### How to cite this article:

Ayilara-Adewale O A, Ojo O E, Olomola J O, et al. (October 09, 2026) A Lightweight Hybrid Model for Early Chronic Kidney Disease Prediction in Resource-Constrained Healthcare. *Cureus J Comput Sci* 3 : es44389-026-00317-2. DOI <https://doi.org/10.7759/s44389-026-00317-2>

```
6:   for each outer fold do
7:     Use outer-training data only for model selection
8:     Run 3-fold inner CV across 18 candidate configurations
9:     Fit preprocessing and PCA within each inner training fold
10:    Train ANN candidates and obtain inner out-of-fold probabilities
11:    Select decision threshold and configuration by balanced accuracy
12:    Refit selected preprocessing, PCA, and ANN on outer-training data
13:    Evaluate on outer-validation data and record all metrics
14:  end for
15: end for
16: Compute mean, SD, and 95% CI across 15 outer-fold estimates
17: // Final model and independent hold-out evaluation
18: Tune the final configuration using only the full development set
19: Fit final preprocessing and PCA on the full development set
20: Fit final ANN and lock the selected classification threshold
21: Apply the fitted pipeline once to the previously unseen hold-out set
22: Compute confusion matrix, accuracy, balanced accuracy, precision, recall, F1-score, and ROC-AUC
23: Return the trained pipeline and performance results
```

## Additional Information

### Author Contributions

All authors have reviewed the final version to be published and agreed to be accountable for all aspects of the work.

**Concept and design:** Oluwatobi A. Ayilara-Adewale, Oluwafolake E. Ojo, Johnson O. Olomola, Yusuf O. Olatunde, Damilola Ogunniyi

**Acquisition, analysis, or interpretation of data:** Oluwatobi A. Ayilara-Adewale, Oluwafolake E. Ojo, Johnson O. Olomola, Yusuf O. Olatunde, Kolawole Y. Obiwusi

**Drafting of the manuscript:** Oluwatobi A. Ayilara-Adewale, Oluwafolake E. Ojo, Johnson O. Olomola, Yusuf O. Olatunde, Kolawole Y. Obiwusi, Damilola Ogunniyi

**Critical review of the manuscript for important intellectual content:** Oluwatobi A. Ayilara-Adewale, Oluwafolake E. Ojo, Yusuf O. Olatunde, Kolawole Y. Obiwusi, Damilola Ogunniyi

### Disclosures

**Human subjects:** All authors have confirmed that this study did not involve human participants or tissue.

**Animal subjects:** All authors have confirmed that this study did not involve animal subjects or tissue.

**Conflicts of interest:** In compliance with [the ICMJE uniform disclosure form](#), all authors declare the following:

---

### How to cite this article:

Ayilara-Adewale O A, Ojo O E, Olomola J O, et al. (October 09, 2026) A Lightweight Hybrid Model for Early Chronic Kidney Disease Prediction in Resource-Constrained Healthcare. *Cureus J Comput Sci* 3 : es44389-026-00317-2. DOI <https://doi.org/10.7759/s44389-026-00317-2>

- **Payment/services info:** All authors have declared that no financial support was received from any organization for the submitted work.
- **Financial relationships:** All authors have declared that they have no financial relationships at present or within the previous three years with any organizations that might have an interest in the submitted work.
- **Other relationships:** All authors have declared that there are no other relationships or activities that could appear to have influenced the submitted work.

## Data Availability Statements

The datasets (and/or code) supporting this study are available from the corresponding author upon reasonable request.

## Acknowledgements

Declaration of Large Language Models Including Artificial Intelligence Chatbots: During the preparation of this work, the authors used ChatGPT for language editing, formatting support, and manuscript organization. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the article.

## References

1. GBD Chronic Kidney Disease Collaboration: [Global, regional, and national burden of chronic kidney disease, 1990-2017: a systematic analysis for the Global Burden of Disease Study 2017](#). *Lancet*. 2020, 395:709-733. [10.1016/s0140-6736\(20\)30045-3](#)
2. Kovesdy CP: [Epidemiology of chronic kidney disease: an update 2022](#). *Kidney International Supplements*. 2022, 12:7-11. [10.1016/j.kisu.2021.11.003](#)
3. Francis A, Harhay MN, Ong ACM, et al.: [Chronic kidney disease and the global public health agenda: an international consensus](#). *Nature Reviews Nephrology*. 2024, 20:473-485. [10.1038/s41581-024-00820-6](#)
4. Luyckx VA, Tuttle KR, Abdellatif D, et al.: [Mind the gap in kidney care: translating what we know into what we do](#). *American Journal of Hypertension*. 2024, 37:640-649. [10.1093/ajh/hpae056](#)
5. Selby NM, Taal MW: [What every clinician needs to know about chronic kidney disease: Detection, classification and epidemiology](#). *Diabetes, Obesity and Metabolism*. 2024, 26:3-12. [10.1111/dom.15683](#)
6. Kalantar-Zadeh K, Jafar TH, Nitsch D, Neuen BL, Perkovic V: [Chronic kidney disease](#). *Lancet*. 2021, 398:786-802. [10.1016/s0140-6736\(21\)00519-5](#)
7. Kim D, Perkovic V, Kotwal S: [Barriers to care: new medications and CKD](#). *Kidney International Reports*. 2024, 9:504-507. [10.1016/j.ekir.2023.12.012](#)
8. Delrue C, De Bruyne S, Speckaert MM: [Application of machine learning in chronic kidney disease: current status and future prospects](#). *Biomedicine*. 2024, 12:568. [10.3390/biomedicine12030568](#)
9. GBD 2023 Chronic Kidney Disease Collaborators: [Global, regional, and national burden of chronic kidney disease in adults, 1990-2023, and its attributable risk factors: a systematic analysis for the Global Burden of Disease Study 2023](#). *Lancet*. 2025, 406:2461-2482. [10.1016/s0140-6736\(25\)01853-7](#)
10. Gogoi P, Valan JA: [Machine learning approaches for predicting and diagnosing chronic kidney disease: current trends, challenges, solutions, and future directions](#). *International Urology and Nephrology*. 2025, 57:1245-1268. [10.1007/s11255-024-04281-5](#)
11. Inker LA, Eneanya ND, Coresh J, et al.: [New creatinine- and cystatin C-based equations to estimate GFR without race](#). *The New England Journal of Medicine*. 2021, 385:1737-1749. [10.1056/nejmoa2102953](#)
12. Sarker IH: [Machine learning: algorithms, real-world applications and research directions](#). *SN Computer Science*. 2021, 2:160. [10.1007/s42979-021-00592-x](#)
13. Habehh H, Gohel S: [Machine learning in healthcare](#). *Current Genomics*. 2021, 22:291-300. [10.2174/1389202922666210705124359](#)
14. Javaid M, Haleem A, Singh RP, Suman R, Rab S: [Significance of machine learning in healthcare: features, pillars and applications](#). *International Journal of Intelligent Networks*. 2022, 3:58-73. [10.1016/j.ijin.2022.05.002](#)

---

## How to cite this article:

Ayilara-Adewale O A, Ojo O E, Olomola J O, et al. (October 09, 2026) A Lightweight Hybrid Model for Early Chronic Kidney Disease Prediction in Resource-Constrained Healthcare. *Cureus J Comput Sci* 3 : es44389-026-00317-2. DOI <https://doi.org/10.7759/s44389-026-00317-2>

15. Chittora P, Chaurasia S, Chakrabarti P, et al.: [Prediction of chronic kidney disease - a machine learning perspective](#). IEEE Access. 2021, 9:17312-17334. [10.1109/access.2021.3053763](#)
16. Islam MA, Majumder MZH, Hussein MA: [Chronic kidney disease prediction based on machine learning algorithms](#). Journal of Pathology Informatics. 2023, 14:100189. [10.1016/j.jpi.2023.100189](#)
17. Khan N, Raza MA, Mirjat NH, Balouch N, Abbas G, Yousef A, Touti E: [Unveiling the predictive power: a comprehensive study of machine learning model for anticipating chronic kidney disease](#). Frontiers in Artificial Intelligence. 2024, 6:1339988. [10.3389/frai.2023.1339988](#)
18. Arif MS, Mukheimer A, Asif D: [Enhancing the early detection of chronic kidney disease: a robust machine learning model](#). Big Data and Cognitive Computing. 2023, 7:144. [10.3390/bdcc7030144](#)
19. Rahman MM, Al-Amin M, Hossain J: [Machine learning models for chronic kidney disease diagnosis and prediction](#). Biomedical Signal Processing and Control. 2024, 87:105368. [10.1016/j.bspc.2023.105368](#)
20. Debal B, Sitote M: [Chronic kidney disease prediction using machine learning techniques](#). Journal of Big Data. 2022, 9:109. [10.1186/s40537-022-00657-5](#)
21. Ganie SM, Dutta Pramanik PK, Mallik S, Zhao Z: [Chronic kidney disease prediction using boosting techniques based on clinical parameters](#). PLoS ONE. 2023, 18:e0295234. [10.1371/journal.pone.0295234](#)
22. Singh V, Asari VK, Rajasekaran R: [A deep neural network for early detection and prediction of chronic kidney disease](#). Diagnostics. 2022, 12:116. [10.3390/diagnostics12010116](#)
23. Alsuhibany SA, Abdel-Khalek S, Algarni A, Fayomi A, Gupta D, Kumar V, Mansour RF: [Ensemble of deep learning based clinical decision support system for chronic kidney disease diagnosis in medical internet of things environment](#). Computational Intelligence and Neuroscience. 2021, 2021:4931450. [10.1155/2021/4931450](#)
24. Zhu Y, Bi D, Saunders M, Ji Y: [Prediction of chronic kidney disease progression using recurrent neural network and electronic health records](#). Scientific Reports. 2023, 13:22091. [10.1038/s41598-023-49271-2](#)
25. Alobaidi S: [Emerging biomarkers and advanced diagnostics in chronic kidney disease: early detection through multi-omics and AI](#). Diagnostics. 2025, 15:1225. [10.3390/diagnostics15101225](#)
26. Lei N, Zhang X, Wei M, et al.: [Machine learning algorithms' accuracy in predicting kidney disease progression: a systematic review and meta-analysis](#). BMC Medical Informatics and Decision Making. 2022, 22:205. [10.1186/s12911-022-01951-1](#)
27. Khalid H, Khan A, Khan MZ, Mehmood G, Qureshi MS: [Machine learning hybrid model for the prediction of chronic kidney disease](#). Computational Intelligence and Neuroscience. 2023, 2023:9266889. [10.1155/2023/9266889](#)
28. Halder RK, Uddin MN, Uddin MA, et al.: [ML-CKDP: Machine learning-based chronic kidney disease prediction with smart web application](#). Journal of Pathology Informatics. 2024, 15:100371. [10.1016/j.jpi.2024.100371](#)
29. Sanmarchi F, Fanconi C, Golinelli D, Gori D, Hernandez-Boussard T, Capodici A: [Predict, diagnose, and treat chronic kidney disease with machine learning: a systematic literature review](#). Journal of Nephrology. 2023, 36:1101-1117. [10.1007/s40620-023-01573-4](#)
30. Dua D, Graff C: [UC Irvine Machine Learning Repository](#). University of California, Irvine. (2021). Accessed: July 24, 2026: <https://archive.ics.uci.edu/ml>.

---

#### How to cite this article:

Ayilara-Adewale O A, Ojo O E, Olomola J O, et al. (October 09, 2026) A Lightweight Hybrid Model for Early Chronic Kidney Disease Prediction in Resource-Constrained Healthcare. Cureus J Comput Sci 3 : es44389-026-00317-2. DOI <https://doi.org/10.7759/s44389-026-00317-2>