*As per UGC guidelines an electronic bar code is provided to secure your paper*

*International Journal for Modern Trends in Science and Technology*
*Volume 11, Issue 10*, pages 12-18.
ISSN: 2455-3778 online
Available online at: http://www.ijmtst.com/vol11issue10.html
DOI: https://doi.org/10.5281/zenodo.17248549

# Chronic Kidney Disease Prediction

**T Anil Karuna Kumar[1] | P.Vidhyusha[2] | P.Hema Satya Sree[2] | B.Navya[2] | N.Lakshmi Sowjanya[2] | K.Bharathi[2]**

[1]*Associate Professor, Department of CSE, Vijaya Institute of Technology for Women, Enikepadu, AP, INDIA.*
[2]*Department of CSE, Vijaya Institute of Technology for Women, Enikepadu, AP, INDIA.*

**To Cite this Article**
T Anil Karuna Kumar, P.Vidhyusha, P.Hema Satya Sree, B.Navya, N.Lakshmi Sowjanya & K.Bharathi (2025). Chronic Kidney Disease Prediction. International Journal for Modern Trends in Science and Technology, 11(10), 12-18. https://doi.org/10.5281/zenodo.17248549

**Article Info**
Received: 04 September 2025; Accepted: 28 September 2025.; Published: 01 October 2025.



*KEYWORDS*

**ABSTRACT**

*Chronic Kidney Disease (CKD) is a progressive disorder marked by a decline in kidney function, leading to severe health complications such as kidney failure and cardiovascular diseases. Early detection is critical to prevent disease progression, but traditional diagnostic methods rely heavily on clinical expertise and can result in delays. Recent advancements in artificial intelligence (AI) and deep learning have facilitated the development of automated prediction models that enhance early diagnosis and intervention. This project proposes an advanced CKD prediction model using a hybrid deep learning approach that integrates Convolutional Neural Networks (CNN), Long Short-Term Memory (LSTM), and an ensemble learning framework. Instead of imaging data, the model leverages structured clinical records, including patient demographics, medical history, lab test results (e.g., serum creatinine, blood urea, and glucose levels), and other biomarkers. CNN is utilized to extract intricate feature patterns from multidimensional clinical data, while LSTM captures temporal dependencies in longitudinal patient records. The ensemble model combines these architectures to enhance predictive accuracy and robustness. Trained on publicly available CKD datasets, this model aims to outperform conventional statistical and machine learning methods by reducing false negatives and improving overall classification performance. The proposed system can be integrated into hospital workflows, telemedicine platforms, and remote patient monitoring systems to provide rapid and reliable CKD diagnosis, ultimately improving patient outcomes through timely intervention.*

## INTRODUCTION

Chronic Kidney Disease (CKD) is a significant public health issue that affects millions of individuals globally. It is characterized by a gradual loss of kidney function, which can lead to severe complications,including kidney failure, cardiovascular disease, and increased. mortality rates. The early detection and treatment of CKD are crucial for improvingpatientoutcomes.

From a machine learning perspective, the task of identifying groups of patients with similar disease progressions is an unsupervised machine learning problem in which patient progressions exhibited by their eGFR values over time are clustered. Most recent solutions, in the context of CKD, have relied on fitting a set of functions, e.g., splines, polynomials, etc., to represent the disease trajectories . However, such models require pre- specifying the form and number of such functions, which becomes a limiting factor.

At the same time, recurrent neural network architectures, such as Long Short-Term Memory networks or LSTMs, Convolutional Neural Networks or CNNs have become the workhorses for modeling sequence data and to extract intricate feature patterns from multidimensional clinical data.

*Deep Learning:*

Deep learning is a subset of machine learning that focuses on training artificial neural networks with multiple layers to recognize patterns and make predictions. Inspired by the structure of the human brain, these deep neural networks process data hierarchically, learning increasingly complex features at each layer. This allows them to perform tasks such as image recognition, speech processing, and natural language understanding with remarkable accuracy.

Deep learning models rely on large datasets and powerful computational resources to extract meaningful patterns from raw data. Unlike traditional machine learning, which often requires manual feature extraction, deep learning networks automatically learn representations through training. Techniques like backpropagation and optimization algorithms adjust network weights to minimize errors and improve performance.

*Artificial Neural Networks:*

Artificial Neural Networks (ANNs) are computational models inspired by the structure and functioning of the human brain. They consist of interconnected nodes, called neurons, organized into layers: the input layer, hidden layers, and output layer. Each neuron processes input data by applying a weighted sum followed by an activation function, allowing the network to learn and make predictions. ANNs are widely used in tasks such as image processing, speech recognition, and natural language understanding.

*Convolutional Neural Networks:*

Convolutional Neural Networks (CNNs) are deep learning models that can process both image and non-image data. While they are widely used for image recognition, they are also effective for analyzing structured data like patient records, time series, and medical datasets, such as those used in chronic kidney disease (CKD) prediction. CNNs help identify hidden patterns in data, making them useful for diagnosing diseases and assessing risks.

*Overview about Ensemble Method:*

The ensemble method is a powerful machine learning technique that combines multiple models to improve overall accuracy and reliability. Instead of depending on a single model, ensemble learning merges the predictions of several models to reduce errors, increase stability, and enhance generalization. This approach is especially useful for complex datasets, where individual models might struggle with overfitting or high variance

## LITERATURE SURVEY

Dina Saif*, Amany M. Sarhan and Nada M. Elshennawy developed a project titled "Deep- kidney: an effective deep learning framework for chronic kidney disease prediction"

The project addresses the challenge of detecting chronic kidney disease early by using this model convolutional neural network (CNN), long short- term memory (LSTM) model, and deep ensemble model. The deep ensemble model fuses three base deep learning classifiers (CNN, LSTM, and LSTM- BLSTM) using majority voting technique. To evaluate the performance of the proposed models, several experiments were conducted on two different public datasets. Among the predictive models and the reached results, the deep ensemble model is superior to all the other models.

Chronic Kidney Disease (CKD) is a growing health concern worldwide, affecting millions of people and leading to severe complications if not diagnosed early. Traditional methods for CKD diagnosis rely on laboratory tests and clinical assessments, which can be time-consuming and costly. As a result, researchers have explored machine learning (ML) and deep learning (DL) approaches to automate and enhance CKD detection. This section reviews existing research on CKD prediction and highlights the advancements made by deep learning models, particularly Convolutional Neural Networks (CNNs) and Long Short-Term Memory (LSTM) networks.

The literature indicates a clear shift from conventional ML methods to deep learning-based models for CKD detection. While earlier approaches relied on basic classifiers, the use of CNNs and LSTMs has significantly improved prediction accuracy. The proposed CNN-BLSTM ensemble model builds on these advancements by combining spatial and sequential learning, demonstrating superior performance. This study contributes to the growing field of AI in healthcare by providing an efficient, automated approach for early CKD detection, potentially leading to improved patient outcomes.

## SYSTEM DESIGN AND ANALYSIS

*Data Collection:*

The CKD dataset will be sourced from reputable medical databases. The dataset will include various patient characteristics and medical test results.

*Data Preprocessing:*

The dataset will undergo preprocessing to handle missing values, normalize numerical features, and encode categorical variables. This step is crucial for ensuring that the data is suitable for machine learning algorithms. The preprocessing code is implemented in the preprocess.py file, which includes handling missing values, label encoding, and scaling features.

*Model Development:*

The project will involve the development of LSTM and CNN models. The models will be trained on the preprocessed dataset, and hyperparameter tuning will be performed to optimize their performance. The model architecture for CNN is defined in the train_cnn.py file, while the LSTM model is defined in the pretrain.py file.

*Ensemble Modeling:*

An ensemble model will be created by combining the predictions of the LSTM and CNN models. This approach aims to enhance prediction accuracy by leveraging the strengths of both models. The ensemble prediction logic is implemented in the app.py file.

*Model Evaluation:*

The models will be evaluated using a separate test dataset. Performance metrics such as accuracy, precision, recall, and F1-score will be calculated to assess the models' effectiveness. The evaluation code is implemented in the evaluate_cnn.py file.

*Web Application Development:*

A web application will be developed using Flask, a Python web framework. The application will provide an interface for users to input patient data and receive predictions. The web application code is implemented in the app.py file, which handles user input, model predictions, and result display.

*Deployment*

The web application will be deployed on a cloud platform, ensuring accessibility for healthcare professionals. User training and documentation will be provided to facilitate the application's use.

The web application will be deployed on a cloud platform, ensuring accessibility for healthcare professionals. User training and documentation will be provided to facilitate the application's use.

## MODEL ARCHITECTURE

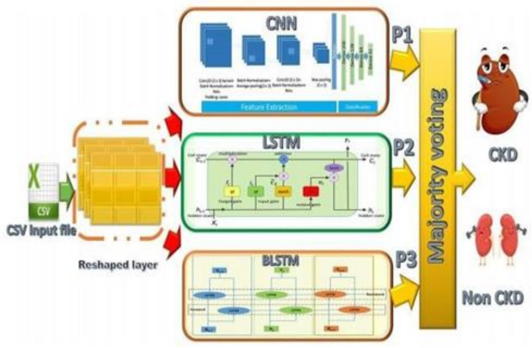


Figure 2: Architecture of CKD prediction system

The architecture of the CKD Prediction System follows a structured pipeline that begins with Data Collection & Preprocessing, where raw medical data is gathered, cleaned, and transformed for model training. This includes handling missing values, normalizing numerical data, and converting categorical attributes into a machine-readable format.

The Model Development phase, two separate models—CNN (Convolutional Neural Network) for image-based features (if applicable) and LSTM (Long Short-Term Memory) for sequential patient records—are designed. These models extract relevant patterns from the dataset.

The Training Pipeline involves training both CNN and LSTM models individually, optimizing hyperparameters, and implementing cross- validation to ensure robustness. After training, The Model Evaluation phase assesses performance using metrics like Accuracy, Precision, Recall, F1-score, ensuring that the model generalizes well to new data. The final step is the Deployment Phase, where the best-performing model (or an ensemble of CNN and LSTM) is converted into a deployable format and hosted on a cloud platform or local server for real- time CKD predictions.

The system is integrated with a user interface, allowing medical professionals to input patient data and receive instant disease predictions with detailed risk analysis reports.

## MODEL TRAINING

This section describes the training process for CNN (Convolutional Neural Network) and LSTM (Long Short-Term Memory) models to predict Chronic Kidney Disease (CKD) using patient data.

*1. CNN Model*

*Model Architecture*

The CNN model is designed to extract patterns from structured patient data. It consists of:

• Conv1D Layer: Applies 64 filters with a kernel size of 3 to detect patterns.

• MaxPooling1D Layer: Reduces feature dimensions while retaining essential information.

• Dropout Layer: Prevents overfitting by setting 20% of neurons to zero during training.

• Flatten Layer: Converts multidimensional data into a 1D format.

• Dense Layers: A hidden layer (50 units, ReLU activation) and an output layer (softmax activation) for classification.

*Training Process*

• Optimizer: Adam (efficient for deep learning models).

• Loss Function: Categorical Cross-Entropy (used for multi-class classification).

• Epochs: 20 (model learns from data in multiple iterations).

• Batch Size: 16 (samples processed before updating weights).

• Validation Data: Used to track model performance on unseen data.

*2. LSTM Model*

*Model Architecture*

LSTM is a recurrent neural network (RNN) designed to process sequential patient data. The architecture includes:

• LSTM Layers: Two layers with 50 units each; the first layer returns sequences for deeper feature extraction.

• Dropout Layer: Prevents overfitting by randomly deactivating neurons (dropout rate = 0.2).

• Dense Layers: A hidden layer (32 units, ReLU activation) and an output layer (softmax activation) for classification.

*Training Process*

• Uses the same optimizer (Adam) and loss function (Categorical Cross-Entropy) as CNN.

• Trained for 20 epochs with a batch size of 16.

• Validation data is included to ensure the model generalizes well.

*3. Model Evaluation*

Both models are tested using unseen patient data. Performance is measured using:

• Accuracy (correct predictions out of total cases).

• Precision (correct CKD predictions out of all CKD-labeled cases).

• Recall (how well the model detects CKD- positive cases).

• F1-score (a balance between precision and recall).

*4. Ensemble Model*

To improve prediction accuracy, an ensemble approach combines the outputs of both CNN and LSTM models. This method enhances reliability by leveraging the strengths of both models while reducing their weaknesses.

This training process ensures the CKD prediction system is robust, accurate, and ready for deployment.

## BLOCK DIAGRAM

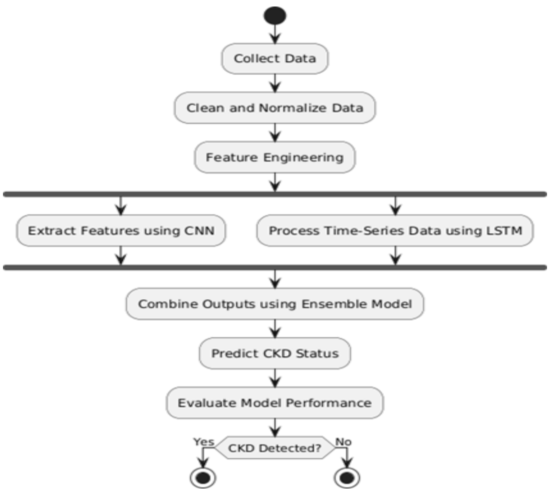


Figure 3: Block diagram of working model

The CKD Prediction System aims to assist in the early detection and diagnosis of Chronic Kidney Disease using advanced machine learning techniques, particularly deep learning models like

CNN and LSTM, combined with ensemble learning. The system follows a systematic pipeline from data collection to prediction and evaluation, as depicted in the block diagram.

The process begins with the collection of relevant data, which includes medical records, lab results, patient history, and physiological parameters known to influence kidney health. These may include blood pressure, glucose levels, age, serum creatinine, albumin levels, and more. Accurate and sufficient data collection is crucial as it lays the foundation for model training and prediction.

After gathering the data, the next step is to clean and normalize it. Real-world datasets often contain missing values, noise, and outliers. Cleaning involves handling these imperfections through methods like mean/mode imputation, outlier removal, or interpolation. Normalization scales the data to a uniform range, typically using min-max scaling or standardization, ensuring that no single feature dominates due to its scale. This step enhances the convergence of learning algorithms.

Following normalization, feature engineering is performed to transform raw data into meaningful inputs for the models. This may include deriving new features, encoding categorical variables, dimensionality reduction, or selecting the most impactful attributes. The goal is to maximize the predictive power of the data while minimizing redundancy.

The diagram then branches into two parallel deep learning paths. On one side, Convolutional Neural Networks (CNNs) are used to extract features from structured and spatial patterns in the data. Although CNNs are popular in image processing, they have also shown promise in handling tabular and structured medical data due to their ability to learn local patterns and feature hierarchies.

On the other side, Long Short-Term Memory (LSTM) networks are used to process time-series data, such as patient records over time, urine test patterns, or changes in blood pressure or serum creatinine. LSTMs are a type of Recurrent Neural Network (RNN) well-suited for capturing temporal dependencies and trends, which are crucial in monitoring the progression of CKD.

The outputs from the CNN and LSTM models are then combined using an ensemble model. Ensemble learning integrates multiple models to improve overall performance and robustness. Techniques such as stacking, voting, or weighted averaging may be used to produce a more accurate final prediction. This integration helps mitigate the individual weaknesses of each model while leveraging their strengths.

The ensemble model predicts the CKD status, classifying patients as either CKD-positive or CKD-negative based on their clinical and physiological features. This output is then passed to the model evaluation stage, where performance metrics such as accuracy, precision, recall, F1- score, and AUC-ROC are computed. This step ensures that the model is both reliable and generalizable across unseen data.

Finally, the decision point in the diagram checks whether CKD is detected. Based on the model's prediction, the system will indicate either a "Yes" or "No," which can be used by healthcare professionals as a decision support tool. While the model is not a replacement for clinical judgment, it provides a valuable second opinion and can help flag high-risk patients for further testing or early intervention.

In summary, this CKD prediction system integrates sophisticated preprocessing, deep learning models (CNN and LSTM), and ensemble learning to deliver a robust and intelligent diagnostic tool. It demonstrates the potential of AI in healthcare, particularly in early disease detection and prevention, where timely action can save lives and reduce long-term healthcare costs.

## RESULT ANALYSIS

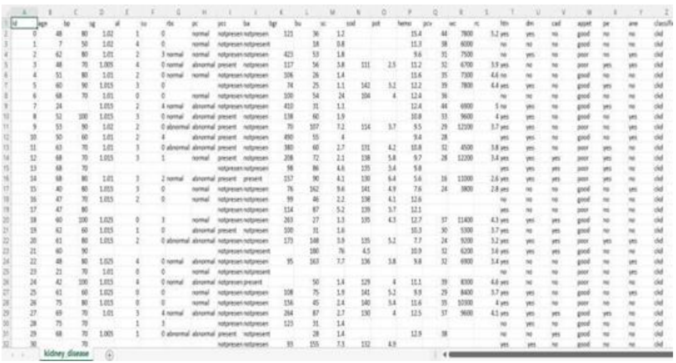

Figure 4: Sample data


Figure 5: Data input

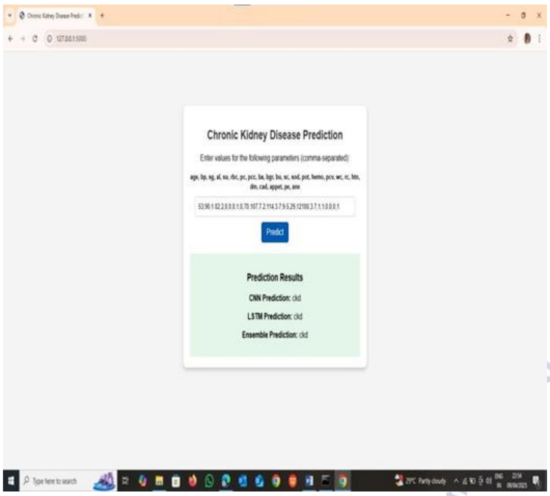


Figure 6: Output Test Results

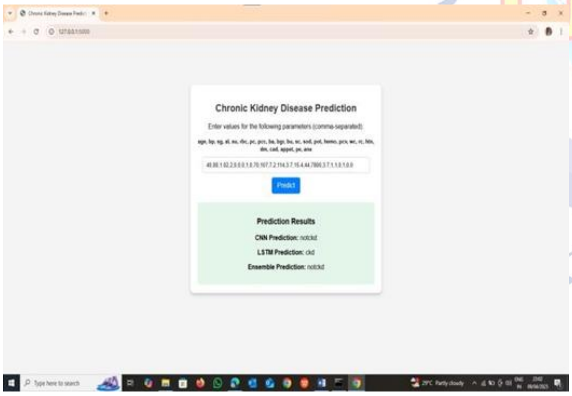


Figure 7 :Output Test Results

## CONCLUSION

The project titled "Chronic Kidney Disease (CKD) Prediction using CNN, LSTM, and Ensemble Models" aimed to develop an intelligent, data-driven system to assist in the early detection of CKD. Given the rising incidence of kidney-related ailments and the need for timely diagnosis, this project holds significant relevance in the medical field.

A robust machine learning pipeline was built, starting with comprehensive data preprocessing including handling missing values, normalization, and transformation of categorical variables. The core of the system focused on implementing three powerful deep learning approaches: Convolutional Neural Network (CNN), Long Short-Term Memory (LSTM), and a hybrid ensemble model that combines the strengths of both.

The CNN model captured spatial feature patterns in the data, while the LSTM learned sequential dependencies, improving understanding of how different symptoms affect CKD progression. The ensemble model combined their outputs to improve prediction accuracy and generalization.

Testing and evaluation using metrics such as accuracy, precision, recall, and F1-score showed that the ensemble model outperformed the individual models. The system underwent rigorous phases of testing—including unit, integration, functional, usability, and validation testing—to ensure its reliability and consistency.One of the major achievements was creating a technically sound yet user-friendly system. The prediction outputs were clear and easily interpretable, enabling use by healthcare professionals with minimal technical expertise.

In conclusion, this project successfully demonstrated how deep learning models can assist in medical diagnosis. By combining CNN, LSTM, and ensemble learning, the system provides a reliable tool for early CKD detection. Future improvements could include larger datasets, real- time integration with health records, and deployment through web or mobile applications, making it a valuable contribution to healthcare analytics.

## Conflict of interest statement

Authors declare that they do not have any conflict of interest.

## REFERENCES

[1] Barik S, et al. Analysis of prediction accuracy of diabetes using classifier and hybrid machine learning techniques. In: Intelligent and cloud com puting. Springer; 2021. p. 399–409.

[2] Bikku T. Multi-layered deep learning perceptron approach for health risk prediction. J Big Data. 2020;7(1):1–14.

[3] Alam MZ, Rahman MS, Rahman MS. A Random Forest based predictor for medical data classification using feature ranking. Inform Med Unlocked. 2019;15: 100180.

[4] Shankar V, et al. Heart disease prediction using CNN algorithm. SN Com put Sci. 2020;1(3):1–8.

[5] Mienye ID, Sun Y, Wang Z. An improved ensemble learning approach for the prediction of heart disease risk. Inform Med Unlocked. 2020;20: 100402.

[6] Almansour NA, et al. Neural network and support vector machine for the prediction of chronic kidney disease: a comparative study. Comput Biol Med. 2019;109:101–11.

[7] Kriplani H, Patel B, Roy S. Prediction of chronic kidney diseases using deep artificial neural network technique. In: Computer aided intervention and diagnostics in clinical and medical images. Springer; 2019. p. 179–87.
[8] Jongbo OA, et al. Development of an ensemble approach to chronic kidney disease diagnosis. Sci Afr. 2020;8: e00456.
[9] Ekanayake IU, Herath D. Chronic kidney disease prediction using machine learning methods. In: 2020 Moratuwa engineering research conference (MERCon). IEEE; 2020.
[10] Gudeti B, et al. A novel approach to predict chronic kidney disease using machine learning algorithms. In: 2020 4th international confer ence on electronics, communication and aerospace technology (ICECA). IEEE; 2020.
[11] Senan EM, et al. Diagnosis of chronic kidney disease using effective classification algorithms and recursive feature elimination techniques. J Healthc Eng. 2021;2021:1–10.