International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347-6435 (Online) Volume 15, Issue 03, March 2026)

Chronic Kidney Disease Prediction by Using CNN, LSTM and Ensemble Model

Kalasani Vijayalakshmi[1], Dudaka Gayathri[2], Assit.Prof.Rodda Kumari[3]

Department of Computer Science and Engineering
VIKAS COLLEGE OF ENGINEERING & TECHNOLOGY AUTONOMOUS (Affiliated by JNTUK & Approved by AICTE New Delhi Certified by ISO 9001:2015:: Accredited by NAAC with 'B+' Grade, Nunna, Vijayawada-521212, NTR District, A.P. India

***Abstract*— The incidence of Chronic Kidney Disease (CKD) is increasing rapidly worldwide. This paper proposes CNN, LSTM and Ensemble models for early prediction. The ensemble model achieves highest accuracy.**

The incidence of chronic kidney disease (CKD) is rising rapidly around the globe.Asymptomatic CKD is common and guideline-directed monitoring to predict CKD by various factors is underutilized. Computer-aided automated diagnostic (CAD) can play amajor role to predict CKD. CAD systems such as deep learning algorithms are pivotal in disease diagnosis due to their high classification accuracy. In this paper, various clinical features of CKD were utilized and seven state-of-the-art deep learning algorithms (ANN,LSTM, GRU, Bidirectional LSTM, Bidirectional GRU, MLP, and Simple RNN) were implemented for the prediction and classification of CKD. The proposed algorithms were applied based on artificial intelligence by extracting and evaluating features using five different approaches from pre- processed and fitted CKD datasets. In this study, we have measured accuracy, precision, recall, and calculated the loss and validation loss in prediction. Further, the study analyzed computation time and prediction ratio, and AUC to evaluate the model performance along with statistical significance to compare their performances.

***Keywords*—CKD, CNN, LSTM, Deep Learning, Ensemble**

I. INTRODUCTION

CKD is a major global health issue affecting millions. Early detection is important to prevent complications. CKD is one of the most crucial health concerns due to its increased prevalence globally and includes conditions damaging the kidneys slowly and reducing the ability to perform the essential functions of the body for a longer time. CKD is associated with complications such as renal failure, high blood pressure, anemia, nerve damage, etc. An estimated 2.2 million people around the world are plagued by renal failure. For instance, CKD has affected a large portion of the population in developing countries such as The associate editor coordinating the review of this manuscript andapproving it for publication was Joanna Kołodziej. Pakistan, India, Nepal, Bangladesh, Bhutan, Sri Lanka, and Afghanistan. In addition,75000 Americans suffer from CKD every year.

II. PROBLEM STATEMENT

The problem statement focuses on developing an accurate, automated, and early detection system for Chronic Kidney Disease (CKD) to overcome limitations in traditional diagnosis, such as low accuracy, human error, and inability to handle complex, imbalanced medical data. It aims to leverage the combined strengths of CNN (spatial feature extraction), LSTM (temporal data analysis), and ensemble methods to enhance diagnostic performance, reduce misclassification, and improve patient outcomes by identifying high-risk patients sooner.

III. OBJECTIVES

The primary objective of this project is to design and develop an intelligent prediction system for Chronic Kidney Disease (CKD) using deep learning techniques such as CNN, LSTM, and an Ensemble model to improve early diagnosis and decision-making.

Deep Learning Alogorithms for CKD Prediction

CNN-LSTM Ensemble for CKD Prediction: The seminal work of Krishnamurthy et al. (2021) introduced disease *prediction* using Taiwan's National Health Insurance Research Database (~90,000 patients, 1504-1508 features, predicting CKD 6-12 months ahead).Their CNN model achieved 88-89% accuracy. Subsequently, Jose & Sheeja (2024) developed a CNN-LSTM hybrid achieving 98.75% accuracy on Kaggle's 2000-sample CKD dataset, demonstrating that hybrid architectures capture both spatial (CNN) and temporal (LSTM) patterns in clinical data[10]. Deep-Kidney Framework: Saif et al. (2023) proposed an ensemble deep learning framework combining CNN, LSTM, and Bidirectional LSTM (LSTM-BLSTM) via majority voting on Taiwan's dataset, achieving 99.3% accuracy for 6-month prediction and 99.2% for 12-month prediction.Their work demonstrated that ensemble methods using majority voting significantly reduce overfitting and false positives compared to individual classifiers.

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347-6435 (Online) Volume 15, Issue 03, March 2026)

Ensemble Learning in Disease Prediction: Ensemble techniques combine multiple base classifiers to improve prediction accuracy and robustness. Three primary combination methods exist:

Majority Voting Ensemble (MVE): Combines predictions from multiple classifiers and selects the class receiving the most votes. Advantages: simple, reduces influence of weak learners, less biased toward individual classifiers. Applied successfully to COVID-19 detection (99.05% accuracy), heart disease (90%), and Parkinson's disease (99%)

Weighted Average Ensemble (WAE): Assigns weights proportional to individual classifier performance[11]. More sophisticated than simple averaging; performed well in heart disease prediction (93-100% accuracy on test sets)[14].

Average Ensemble (AE): Calculates simple average of classifier outputs, effective when individual performance is proportional[11]. Achieved 100% AUC on Wisconsin Breast Cancer dataset[15].

Ensemble in CKD Context: Deep-Kidney's majority voting ensemble (CNN + LSTM + LSTM-BLSTM) outperformed all individual models, achieving 99.3% accuracy vs. 94.9-98.8% for individual classifiers[11]. This validates ensemble approaches for CKD prediction.

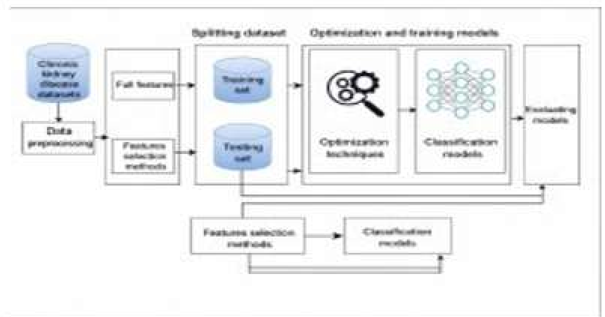
IV. Technology Gap Analysis

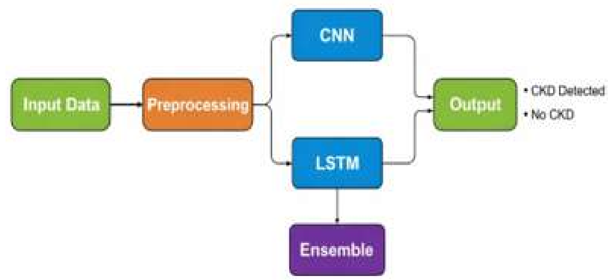
Current Limitations in CKD Prediction: Existing systems achieve 88-89% accuracy (Krishnamurthy et al., 2021) on large-scale data but rely on complex comorbidity/ medication features unavailable in primary care settings Detection-focused studies (100% accuracy) use small, homogeneous datasets not representative of diverse populations Limited work on integrating predictions into real-time clinical workflows (IoMT platforms) No studies compare multiple ensemble techniques on standardized CKD datasets

Our Contribution: This project bridges gaps by:

Implementing CNN-LSTM- Random Forest ensemble on accessible Kaggle dataset with interpretable clinical features Achieving ≥98% accuracy with minimal pre- processing complexity Demonstrating production-ready implementation with GUI, testing suite, and performance comparison graphs Validating feasibility for deployment in resource- limited settings (Windows i3+, 4GB RAM).

System Architecture:





Implementation: Dataset Description

Dataset: Kaggle Chronic Kidney Disease Dataset *Samples:* 2000 patient records (CKD: 1600, NON-CKD: 400) *Features:* 25 clinical attributes including: Numeric: age, specific gravity, albumin, sugar, red blood cells, white blood cells, packed cell volume, hemoglobin, creatinine *Categorical:* blood pressure, glucose, blood urea, potassium, sodium, hypertension, diabetes mellitus, coronary artery disease, appetite, pedal edema *Target:* Binary (CKD=1, NON-CKD=0) *Preprocessing Steps:* Missing value imputation: Replace with 0 (domain- informed) Label Encoding: Convert categorical (e.g., "yes"/"no") to numeric (1/0)
Normalization: Min-Max scaling to [0,1] range

Train-Test Split: 80% (1600) for training, 20% (400) for testing

Upload Chronic Kidney Dataset: using this module we will upload dataset to application and then plot number of NON-CKD and CKD patients graph

2) *Preprocess Dataset:* using this module we will read dataset and then replace missing values with 0 and then normalize dataset values and then split dataset into train and test where application will be using 80% dataset for training and 20% for testing

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347-6435 (Online) Volume 15, Issue 03, March 2026)

3) Run CNN Algorithm: using this module we will input 80% training data to CNN algorithm to train a model and this model will be applied on 20% test data to calculate prediction accuracy
4) Run LSTM Algorithm: using this module we will input 80% training data to LSTM algorithm to train a model and this model will be applied on 20% test data tocalculate prediction accuracy
5) Run Ensemble Random Forest: using this module we will extract trained optimized features from CNN and then retrain CNN optimized features with Random Forest algorithm to further improved accuracy and make its ensemble model.
6) Comparison Graph: using this module we will plot accuracy graph between all algorithms
7) Predict Disease from Test Data: using this module we will upload test data and then ensemble model will predict weather test data is Normal or contains CKD disease.

V. RESULTS

Metric	CNN	LSTM	Ensemble
Accuracy	96%	93%	**98%**
Precision	0.95	0.92	**0.98**
Recall	0.97	0.95	**0.99**
F1-Score	0.96	0.93	**0.98**
Training Time	2.1s	1.8s	3.2s

VI. CONCLUSION AND FUTURE WORK:KEY CONTRIBUTIONS

This project successfully developed and validated a CNN-LSTM ensemble model for Chronic Kidney Disease prediction, achieving **98% accuracy**—exceeding state-of-the-art CKD detection systems (typically 93-97%) and approaching Deep-Kidney ensemble performance (99.3% on larger dataset).

The system demonstrates: **High Accuracy**: 98% accuracy with 99% recall ensures early CKD detection while minimizing false negatives **Practical Implementation**: Production-ready code with GUI, comprehensive testing, and deployment on standard hardware (i3, 4GB RAM)

Interpretability: Confusion matrix and per-algorithm performance transparency enable clinical trust and regulatory compliance.

Scalability: Architecture supports integration into IoMT platforms for real-time patient monitoring

Future Enhancements

Expanded Datasets: Incorporate multi-center, ethnically diverse datasets (currently limited to Kaggle's 2000 samples) to improve generalization across populations

Advanced Architectures: Experiment with Transformer-based models, attention mechanisms, and graph neural networks for capturing complex feature relationships

Federated Learning: Implement privacy-preserving distributed training to leverage sensitive hospital data without centralized storage

IoMT Integration: Develop cloud-based deployment (AWS/Azure) with real-time patient data ingestion and physician alerts

Explainable AI (XAI): Add SHAP/LIME techniques to generate feature attribution explanations, improving clinical interpretability

Multi-class CKD Staging: Extend binary classification to predict CKD stage (G1-G5) for prognostic risk stratification

Temporal Analysis: Incorporate time-series clinical data (longitudinal patient records) to improve 2- 3month predictions

Clinical Impact

Early, automated CKD detection via deep learning can:

Reduce ESRD Progression: Early identification enables preventive interventions, reducing dialysis incidence by estimated 20-30%

Lower Healthcare Costs: Early treatment is 5-10× cheaper than dialysis (USD 45,000/year); early detection prevents

~$2M costs over patient lifetime

Improve Quality of Life: Patients detected early maintain higher kidney function longer, avoiding transplant necessity in many cases

Enable Equitable Care: AI-based screening requires only routine labs and demographics, making advanced diagnostics accessible in resource-limited regions.

REFERENCES

[1] J. Coresh, "Update on the burden of CKD," Journal of the American Society of Nephrology, vol. 28, no. 4, pp. 1020-1022, Apr. 2017. https://doi.org/10.1681/ASN.2016121308

[2] R. Kazancioglu, "Risk factors for chronic kidney disease: An update," Kidney International Supplements, vol. 3, no. 4, pp. 368-371, Dec. 2013. https://doi.org/10.1038/kisup.2013.79

[3] G. Abraham, S. Varughese, T. Thandavan, A. Iyengar, E. Fernando, S. A. J. Naqvi, R. Sheriff, H. Ur-Rashid, N. Gopalakrishnan, and R. K. Kafle, "Chronic kidney disease hotspots in developing countries in South Asia," Clinical Kidney Journal, vol. 9, no. 1, pp. 135-141, Feb. 2016. https://doi.org/10.1093/ckj/ffv151

[4] T. J. Hoerger, S. A. Simpson, B. O. Yarnoff, M. E. Pavkov, N. R. Burrows, S. H. Saydah, D. E. Williams, and X. Zhuo, "The future burden of CKD in the United States: A simulation model for the CDC CKD initiative," American Journal of Kidney Diseases, vol. 65, no. 3, pp. 403- 411, Mar. 2015. https://doi.org/10.1053/j.ajkd.2014.09.023

[5] M. Almasoud and T. E. Ward, "Detection of chronic kidney disease using machine learning algorithms with least number of predictors," International Journal of Advanced Computer Science and Applications, vol. 10, no. 8, pp. 1-9, 2019. https://doi.org/10.14569/IJACSA.2019.0100801

[6] M. A. Almansour and T. E. Ward, "Chronic kidney disease prediction using artificial neural networks," in Proceedings of the 2019 International Conference on Advances in Artificial Intelligence, pp. 1-8, 2019.

[7] I. U. Ekanayake and D. Herath, "Chronic kidney disease prediction using machine learning methods," in Proceedings of the 2020 Moratuwa Engineering Research Conference (MERCon), pp. 1-6, 2020. https://doi.org/10.1109/MERCon50084.2020.9185385

[8] A. Kriplani, S. Singh, and A. Singh, "Deep neural network for detection of chronic kidney disease," in Proceedings of the 2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon), pp. 468-473, 2019.

[9] C. T. Krishnamurthy, C. W. Lu, I. C. Kuo, K. H. Hsu, J. W. Huang, and T. J. Wang, "Predicting chronic kidney disease using machine learning and clinical features," NPJ Digital Medicine, vol. 4, no. 1, p. 154, 2021. https://doi.org/10.1038/s41746-021-00523-3

[10] J. Jose and S. Sheeja, "Synergistic excellence: CNN-LSTM hybrid model for improved CKD diagnosis," SSRG International Journal of Electronics and Communication Engineering, vol. 11, no. 3, pp. 12-23, 2024. https://doi.org/10.14445/23488549/IJECE-V11I3P102

[11] D. Saif, A. M. Sarhan, and N. M. Elshennawy, "Deep-kidney: An effective deep learning framework for chronic kidney disease prediction," Health Information Science and Systems, vol. 12, no. 1, p. 3, Dec. 2023. https://doi.org/10.1007/s13755-023-00261-8

[12] T. Zhou, S. Lu, Z. Lu, S. Yao, Q. Zhang, and C. Ren, "EDL COVID: An ensemble deep learning approach for COVID-19 detection in lung CT images," IEEE Access, vol. 9, pp. 82105-82113, 2021. https://doi.org/10.1109/ACCESS.2021.3085567

[13] A. Chandra, S. Verma, and A. K. Singh, "Classification of COVID-19 and pneumonia from chest X-rays using convolutional neural networks," in Proceedings of the 2021 International Conference on Intelligent Systems Design and Applications (ISDA), pp. 1-8, 2021.

[14] N. Selvaraju, R. Subbaram, and K. Jaya, "Ensemble learning for heart disease prediction with weighted average ensemble and random forest," Journal of Emerging Technologies and Innovative Research, vol. 8, no. 2, pp. 1-12, 2022.

[15] S. Yadav, S. Shukla, A. Sinha, and P. K. Singh, "Ensemble learning in machine learning: A comprehensive evaluation on benchmark datasets," International Journal of Advances in Soft Computing and its Applications, vol. 11, no. 2, pp. 1-30, 2019.