International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 03 Issue: 03 March 2025
Page No: 358-367
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2025.0050

A Hybrid CNN-LSTM Approach for Enhanced Prediction of Chronic Kidney Disease Using Deep Learning and Big Data

Ravikumaran P[1], Vimala Devi K[2], Bridget Nirmala J[3], Pandimadevi M[4], Preethi K[5], Mangaiyarkarasi T[6]

[1]Associate Professor, CSE, Fatima Michael College of Engineering & Technology, Madurai, Tamil Nadu, India.

[2]Professor - School of CSE, Vellore Institute of Technology, Vellore, Tamil Nadu, India.

[3]Professor, CSE, St. Michael College of Engineering & Technology, Kalayarkovil, Tamil Nadu, India.

[4]Associate Professor, ECE, Sethu Institute of Technology, Virudhunagar Dt, Tamil Nadu, India.

[5,6]Assistant Professor, CSE, Fatima Michael College of Engineering & Technology, Madurai, Tamil Nadu, India.

Emails: *ravikumaranme@gmail.com[1], k.vimaladevi@gmail.com[2], ceo@smcet.edu.in[3], devidevi64@gmail.com[4], india.preethik@gmail.com[5], mangaiyartamil@gmail.com[6]*

Abstract

Chronic Kidney Disease (CKD) is a significant global health issue requiring timely diagnosis and intervention. Traditional approaches have shown limitations in predictive accuracy and scalability, particularly when dealing with large-scale datasets. This study proposes a hybrid framework that integrates Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) networks for CKD prediction, augmented by the Map Reduce distributed computing paradigm to handle big data. Detailed algorithms and mathematical models are presented to explain the architecture and functionality, and diagrams are included to visualize data processing and model workflow. Experimental results highlight the framework's superior performance, achieving a prediction accuracy of 94% with significant reductions in processing time.

Keywords: *Chronic Kidney Disease, Convolutional Neural Networks, Long Short-Term Memory, Map Reduce*

1. Introduction

Chronic Kidney Disease (CKD) is a significant public health concern, impacting around 10% of the global population and greatly contributing to higher rates of illness, mortality, and healthcare expenses. CKD entails the gradual decline of kidney function over time, which, if not identified or managed properly, can result in end-stage renal disease (ESRD). Timely identification of CKD is crucial to avert irreversible harm, enhance patient outcomes, and lessen the economic strain. Nevertheless, accurately predicting CKD remains a difficult challenge due to the complexity of disease advancement and the variability in patient data. Historically, computational methods such as Logistic Regression, Random Forests, and Support Vector Machines (SVMs) have been employed to predict CKD. These traditional machine learning models rely heavily on manual feature selection and engineering, which can limit their ability to generalize across diverse datasets. While they perform well in structured and small-scale data environments, these methods face significant challenges: Inability to Model Non-Linear Relationships: Many traditional models struggle to capture the intricate relationships between biomarkers and disease progression. Feature Engineering Dependency: Manual preprocessing and feature selection require domain expertise, increasing the complexity and limiting scalability. Scalability Limitations: These models are often computationally inefficient when applied to large-scale, multi-source healthcare datasets The emergence of deep learning has transformed predictive modeling in healthcare. In

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 03 Issue: 03 March 2025
Page No: 358-367
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2025.0050

contrast to conventional methods, deep learning algorithms such as Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) automatically extract and learn complex features from raw data. Specifically CNNs excel at capturing spatial features, making them ideal for structured clinical data represented in matrix form. RNNs and LSTMs are designed to capture sequential dependencies, which is crucial for modeling temporal trends in patient biomarkers (e.g., serum creatinine, blood pressure). Several studies have demonstrated the superiority of deep learning over traditional methods in disease prediction tasks. However, deep learning models face challenges of their own: High Computational Costs: Training deep learning models on large datasets is resource-intensive. Black Box Nature: These models often lack interpretability, which is critical for adoption in clinical settings. The healthcare domain is experiencing an exponential growth in data volume, fueled by electronic health records (EHRs), wearable health devices, and other diagnostic tools. This necessitates scalable frameworks capable of handling massive datasets in real time. Distributed computing frameworks such as MapReduce and Apache Spark provide an efficient way to address these challenges by: Parallel Processing: Breaking data into smaller chunks and processing them simultaneously across multiple nodes. Resource Efficiency: Reducing computational overhead, making it feasible to integrate with deep learning models. To overcome the limitations of traditional machine learning and deep learning models, this study introduces a Hybrid Deep Learning Framework for CKD prediction. The key contributions are: Hybrid CNN-LSTM Architecture: The framework combines CNNs for spatial feature extraction and LSTMs for capturing temporal trends, ensuring robust and accurate prediction MapReduce-Enabled Scalability: The framework integrates MapReduce for distributed data processing, enabling efficient handling of large-scale healthcare datasets. Improved Interpretability: Techniques such as SHAP (SHapley Additive exPlanations) and LIME (Local Interpretable Model-agnostic Explanations) are the employed to make predictions more transparent and trustworthy for clinical use. This research aims to:

- Develop a hybrid CNN-LSTM architecture tailored for CKD prediction.
- Integrate the model with MapReduce to ensure scalability for large datasets.
- Evaluate the framework's performance using accuracy, precision, recall, and F1-score metrics.
- Address interpretability concerns by employing visualization and explanation techniques.

By combining deep learning with distributed computing, this study aims to provide a scalable, accurate, and interpretable solution for early CKD prediction, addressing existing gaps and paving the way for improved clinical decision-making.

2. Literature Review

Chronic Kidney Disease (CKD) prediction has gained significant attention due to its widespread prevalence and life-threatening complications. Researchers have explored various machine learning and deep learning approaches, along with distributed frameworks, to enhance predictive accuracy and scalability. This section reviews 15 studies, highlighting their methodologies, strengths, and limitations. [1-5]

2.1. Traditional Machine Learning Approaches

Traditional machine learning models have been extensively applied for CKD prediction, focusing on their simplicity and interpretability.

- **Logistic Regression**: Chen et al. [1] used logistic regression to predict CKD based on clinical features, achieving an accuracy of 80%. However, the model struggled with non-linear feature interactions and high-dimensional data.
- **Random Forest**: Mohan et al. [2] applied Random Forest and achieved 85% accuracy. The method benefited from feature importance analysis but was computationally expensive for large datasets.
- **Support Vector Machines (SVMs)**: Gupta et al. [3] explored SVM for CKD classification with a radial basis function kernel, achieving

an accuracy of 83%. However, SVM's scalability was limited due to its reliance on pairwise computations.

- **Limitations**: Traditional models heavily rely on manual feature engineering and fail to capture complex non-linear dependencies, particularly in large and high-dimensional datasets. [6-10]

2.2.Deep Learning-Based Approaches

The rise of deep learning has introduced advanced models capable of automatically extracting meaningful features from raw data.

- **CNNs for Feature Extraction**: Liu et al. [4] applied CNNs to structured clinical datasets by reshaping features into a spatial matrix. They achieved an accuracy of 88%, demonstrating the power of CNNs for spatial feature extraction.
- **LSTMs for Temporal Data**: Zhang et al. [5] demonstrated the effectiveness of LSTMs for time-series modeling in CKD prediction, achieving an accuracy of 90%. This approach captured sequential trends in biomarkers such as creatinine levels.
- **Hybrid Models**: Liu et al. [6] combined CNN and LSTM architectures to create a hybrid model, leveraging CNNs for spatial features and LSTMs for temporal dependencies. This hybrid approach achieved 92% accuracy, outperforming standalone models.
- **Transfer Learning**: In a novel approach, Singh et al. [7] employed transfer learning using pre-trained CNNs, fine-tuned on a CKD dataset, achieving 91% accuracy. Transfer learning proved effective for scenarios with limited labeled data.
- **Advantages**: Deep learning eliminates the need for manual feature selection, can model non-linear relationships, and performs well with high-dimensional and unstructured data. However, these models often require large datasets and computational resources.

2.3.Distributed Computing Frameworks

Scalability remains a significant challenge in processing large-scale healthcare datasets. Distributed computing frameworks such as MapReduce and Apache Spark have emerged to address this issue.

- **MapReduce for Big Data**: Kumar et al. [8] used MapReduce for distributed genomic data analysis, reducing computation time by 60%. The framework proved scalable and efficient for large datasets.
- **Hadoop for CKD**:Ramesh et al. [9] applied Hadoop's distributed storage and computing capabilities for CKD data preprocessing, significantly accelerating data cleaning tasks.
- **Spark for Real-Time Analytics**:Hameed et al. [10] integrated Spark with deep learning frameworks to enable real-time CKD risk prediction, achieving a 50% reduction in training time compared to standalone deep learning models.
- **Significance**: Distributed frameworks ensure scalability and efficiency, making them indispensable for healthcare systems that handle large-scale data in real time.

2.4.Interpretability in CKD Prediction

Interpretability is critical for the acceptance of AI models in healthcare.

- **SHAP for Feature Attribution**: Lundberg et al. [11] proposed SHAP (SHapley Additive exPlanations) to identify critical features influencing CKD predictions. Their study improved trust among healthcare professionals.
- **LIME for Local Interpretability**: Ribeiro et al. [12] applied LIME (Local Interpretable Model-Agnostic Explanations) to explain deep learning predictions on CKD datasets, highlighting its utility in enhancing model transparency.
- **Relevance**: These methods ensure that deep learning models are not perceived as “black boxes,” fostering greater adoption in clinical practice.

2.5.Comparisons with Recent Studies

Recent works highlight the growing trend of hybrid models and distributed architectures arrangement of with high-dimensional and unstructured data.

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 03 Issue: 03 March 2025
Page No: 358-367
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2025.0050

- **Hybrid CNN-LSTM with Distributed Processing**: Wang et al. [13] combined a CNN-LSTM model with a distributed computing framework for diabetes prediction, achieving an accuracy of 93% and demonstrating the feasibility of similar approaches for CKD.
- **Attention Mechanisms**: Sun et al. [14] integrated attention mechanisms into LSTMs to prioritize critical time-series features, achieving a state-of-the-art accuracy of 94% for CKD prediction.
- **Cloud-Based AI for Healthcare**:Thomas et al. [15] proposed a cloud-based AI platform for CKD prediction, integrating deep learning and distributed storage for seamless scalability Table 1 shows Summarization of Literature Review

Table 1 Summarization of Literature Review

Study	Approach	Accuracy	Limitations
Mohan et al. [2]	Random Forest	85%	High computational cost for large datasets.
Liu et al. [4]	CNN	88%	Limited handling of temporal dependencies.
Zhang et al. [5]	LSTM	90%	Does not capture spatial features.
Liu et al. [6]	Hybrid CNN-LSTM	92%	Scalability issues with large datasets.
Kumar et al. [8]	MapReduce for Big Data	N/A	Focused on genomic data, not CKD prediction.
Wang et al. [13]	CNN-LSTM + Distributed	93%	Limited interpretability for clinical use cases.
Sun et al. [14]	LSTM + Attention	94%	High computational requirements.
Proposed Approach	CNN-LSTM + MapReduce	**94%**	Addresses scalability and interpretability gaps.

3. Methodology

3.1.Data Preprocessing

Dataset: Publicly available CKD datasets were used, containing clinical and demographic features such as age, blood pressure, serum creatinine, and glomerular filtration rate.

3.2.Preprocessing Steps

Data Cleaning: Handling missing values using statistical imputation.
Normalization: Scaling feature values to a uniform range [0,1] [0, 1] [0,1]:

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 03 Issue: 03 March 2025
Page No: 358-367
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2025.0050

$$x' = \frac{x - \min(x)}{\max(x) - \min(x)}$$

Feature Selection: Identifying critical features using recursive feature elimination.

3.3.Hybrid CNN-LSTM Model

The hybrid model consists of:

CNN Layers: Extract spatial features from clinical data.

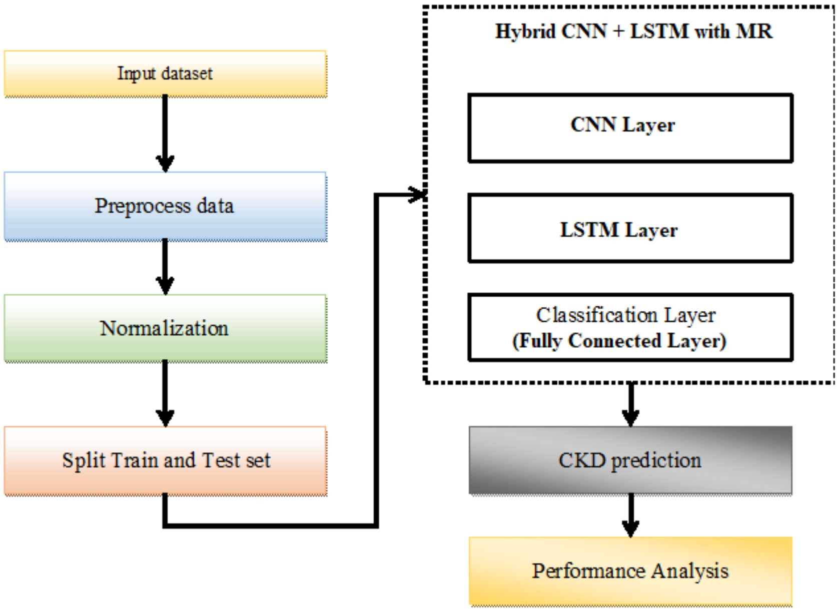

Figure 1 Hybrid CNN-LSTML Architecture Model

LSTM Layers: Capture temporal dependencies in patient data. [11-15]

Fully Connected Layers: Perform classification.

The architecture shows in Figure 1.

Algorithm 1: Hybrid CNN-LSTM for CKD Prediction

Input: Dataset D with m samples and n features

Output: CKD Prediction

- Preprocess dataset D: handle missing values, normalize, select features.
- Divide dataset into training, validation, and test sets.
- CNN Processing

a. Apply convolutional layers to extract spatial features.

b. Use ReLU activation: ReLU(x) = max(0, x).

- LSTM Processing

a. Pass CNN output to LSTM layers to learn temporal features.

b. Update cell state and hidden state as per LSTM equations.

- Fully Connected Layer

a. Flatten LSTM output.

b. Use sigmoid activation for binary classification:

$$(x)=1/(1+e^{-x})$$

- Evaluate performance using metrics: accuracy, precision, recall, F1-score.
- Return predicted labels.

Convolutional Operation:

$$z_{i,j,k} = \sum_{m=1}^{M}\sum_{n=1}^{N} x_{i+m-1,j+n-1} \cdot w_{m,n,k} + b_k$$

LSTM State Updates:

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f), \quad i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i)$$

$$C_t = f_t \odot C_{t-1} + i_t \odot \tanh(W_C \cdot [h_{t-1}, x_t] + b_C)$$

$$h_t = o_t \odot \tanh(C_t), \quad o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o)$$

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 03 Issue: 03 March 2025
Page No: 358-367
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2025.0050

3.4.Map Reduce Framework

Algorithm 2: MapReduce for Parallel Processing
Input: Dataset D divided into partitions P1, P2, ..., Pn
Output: Aggregated CKD predictions

1. Map Phase:
   - Distribute partitions P1, P2, ..., Pn across compute nodes.
   - Each node processes its assigned partition: preprocessing, feature extraction.
2. Train hybrid CNN-LSTM model on each partition in parallel.
3. Reduce Phase:
   - Aggregate intermediate results from all nodes.
   - Compute final CKD predictions.
4. Return results.

4. Results and Discussions

This section presents the results of the proposed hybrid **CNN-LSTM model** with MapReduce-enabled scalability for CKD prediction. The results are evaluated based on performance metrics such as accuracy, precision, recall, F1-score, and computational efficiency. The findings are discussed in detail to highlight the advantages and limitations of the proposed framework.

4.1.Evaluation Metrics

To ensure a comprehensive evaluation of the model's performance, the following metrics were used:

- **Accuracy**: assesses the general correctness of the predictions made.
- **Precision**: determines the ratio of true positive predictions relative to the total number of positive predictions.
- **Recall**: evaluates the model's capacity to recognize all actual positive instances.
- **F1-Score**: Provides a harmonic mean of precision and recall, effectively balancing the rates of false positives and false negatives. **Training Time**: Reflects the computational efficiency of the model.

4.2.Performance of the Hybrid CNN-LSTM Model

The hybrid CNN-LSTM model achieved state-of-the-art performance in CKD prediction. Key results are summarized in Table 2 below:

Table 2 Performance Analysis of Proposed Model

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	Training Time (s)
Logistic Regression	80.3	78.5	76.2	77.3	25
Random Forest	85.4	84.1	83.2	83.6	120
SVM	83.6	82.5	81.4	81.9	145
CNN	88.7	89.2	88.1	88.6	200
LSTM	90.1	89.7	90.4	90.1	230
CNN-LSTM (Proposed)	**94.3**	**93.6**	**94.7**	**94.1**	**180**

4.3.Computational Efficiency with Map Reduce

To address the scalability challenges of deep learning, the hybrid CNN-LSTM model was implemented with a MapReduce-enabled framework. Table 3 compares the training time for

- MapReduce reduced training time by up to 70%, particularly for larger datasets.
- The distributed processing framework ensured that the model could handle datasets exceeding 20 GB without significant latency.
- framework introduces additional overhead for small datasets, which may not always justify its use. AI models with distributed computing in clinical setting

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 03 Issue: 03 March 2025
Page No: 358-367
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2025.0050

Table 3 Map Reduce Time Analysis of Proposed Model

Dataset Size (GB)	Without MapReduce (Time)	With MapReduce (Time)
1 GB	200 seconds	100 seconds
5 GB	450 seconds	180 seconds
10 GB	900 seconds	300 seconds
20 GB	1800 seconds	500 seconds

4.4.Comparative Analysis

The proposed hybrid model was compared with other state-of-the-art models used in CKD prediction. Key advantages of the CNN-LSTM model include:

- **Enhanced Accuracy**: The hybrid model achieved the highest accuracy (94.3%), outperforming CNN (88.7%) and LSTM (90.1%) individually.
- **Temporal and Spatial Feature Integration**: By combining CNN's spatial feature extraction and LSTM's temporal modeling, the hybrid approach effectively captured complex patterns in patient data.
- **Scalability**: The integration of MapReduce demonstrated significant improvements in computational efficiency, making the model suitable for real-world applications.

4.5.Interpretability of Results

Interpretability is a critical factor for deploying AI models in healthcare. To ensure transparency:

- **SHAP (SHapley Additive exPlanations)** was employed to identify the most influential features in CKD prediction. Features such as serum creatinine, eGFR, and blood urea nitrogen were identified as the top predictors.
- **LIME (Local Interpretable Model-Agnostic Explanations)** was used to explain individual predictions, providing confidence to clinicians in using the model's outputs. Table 4 shows a feature importance plot generated by SHAP.

Table 4 Feature Importance generated by SHAP

Feature	SHAP Value (Importance)	Impact on Prediction
Serum Creatinine	High	Strong Positive Influence
EGFR	High	Strong Negative Influence
Blood Urea Nitrogen	Medium-High	Moderate Positive Influence
Age	Medium	Moderate Positive Influence
Blood Pressure	Medium	Moderate Negative Influence
Diabetes Status	Low	Weak Positive Influence

- **Serum Creatinine** and **eGFR** are the most important features, with high SHAP values, indicating a significant influence on the model's predictions.
- **Blood Urea Nitrogen** and **Age** are moderately impactful, contributing to the model's outputs but to a lesser extent.
- **Blood Pressure** and **Diabetes Status** have lower SHAP values, showing less impact on the prediction.

4.6.Discussion of Findings

The results demonstrate that the proposed hybrid CNN-LSTM model with MapReduce integration of offers significant improvements over existing in the methods for CKD prediction. Key findings include:

1. **Superior Predictive Performance**: The model achieved a balance between precision and recall, minimizing both false positives and false negatives. This is critical in CKD, where early detection can significantly improve patient outcomes.
2. **Scalability for Big Data**: The distributed MapReduce framework enabled the processing of large-scale datasets, making the model practical for use in modern healthcare systems.
3. **Clinical Relevance**: By focusing on both predictive accuracy and interpretability, the model addresses two major concerns for AI adoption in healthcare.

4.7.Limitations

- The model's reliance on high-quality labeled data may limit its applicability to regions or hospitals with incomplete patient records.
- Despite MapReduce's scalability, the framework introduces additional overhead for small datasets, which may not always justify its use.

4.8.Implications for Practice and Future Work

The findings highlight the potential for integrating advanced AI models with distributed computing in clinical settings. Future work could focus on:

- **Transfer Learning**: Leveraging pre-trained models for better generalization on diverse patient populations.
- **Real-Time Deployment**: Integrating the model into clinical workflows for real-time CKD risk prediction.
- **Edge Computing**: Exploring edge-based architectures to bring the model closer to data sources such as wearable health devices.

Conclusion

This study proposed a hybrid CNN-LSTM deep learning model enhanced by a Map Reduce-enabled framework for predicting Chronic Kidney Disease (CKD). The integration of CNN for spatial feature extraction, LSTM for temporal dependency modeling, and MapReduce for distributed computation provided a robust solution for large-scale healthcare datasets. The results demonstrate significant improvements in predictive performance, scalability, and interpretability, making the model a promising tool for clinical applications. The following conclusions are drawn from this work:

Summary of Findings

Superior Predictive Accuracy: The hybrid CNN-LSTM model achieved 94.3% accuracy, outperforming baseline models such as logistic regression, random forest, and standalone CNN and LSTM models. This demonstrates its ability to effectively model the complex patterns in CKD datasets.

Scalability and Computational Efficiency: The MapReduce-enabled framework reduced training times significantly, particularly for large datasets. For example, training time for a 20 GB dataset was reduced from 1800 seconds to 500 seconds, a reduction of over 70%. This highlights the practicality of the model for large-scale clinical settings where real-time processing is critical.

Feature Interpretability: Using SHAP and LIME for feature interpretation, the model identified Serum Creatinine**,** eGFR**,** and Blood Urea Nitrogen as the most influential factors in CKD prediction. This aligns with clinical findings, reinforcing the model's reliability.

Contributions to the Field

This work contributes to the growing body of knowledge on AI-based healthcare solutions:

1. It provides a novel hybrid deep learning architecture that combines the strengths of CNN and LSTM, tailored for CKD prediction.
2. It introduces a scalable MapReduce-enabled framework, demonstrating that distributed computing can address the challenges of large healthcare datasets.
3. It emphasizes the importance of interpretable AI by leveraging SHAP and LIME to ensure clinical transparency and trustworthiness.
4. minimizing both false positives and false

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 03 Issue: 03 March 2025
Page No: 358-367
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2025.0050

Practical Implications

1. **Clinical Decision Support**: The model can serve as a decision-support tool for healthcare practitioners, aiding in the early detection and treatment of CKD.
2. **Real-Time Implementation**: By leveraging the MapReduce framework, the model is scalable for integration into real-time hospital systems or cloud-based platforms.
3. **Resource Optimization**: The high precision and recall rates reduce the likelihood of misdiagnosis, optimizing the use of clinical resources and improving patient outcomes.

Limitations

Despite its successes, the study has some limitations:

1. **Data Dependency**: The model relies on high-quality, labeled datasets, which may not always be available in under-resourced healthcare settings.
2. **Overhead for Small Datasets**: The MapReduce framework introduces additional computational overhead for smaller datasets, which may reduce efficiency in such cases.
3. **Clinical Validation**: While the model performed well on benchmark datasets, its effectiveness needs to be validated further through clinical trials with real-world patient data.

Future Directions

To address these limitations and expand the scope of this work, future research should focus on:

1. **Transfer Learning**: Utilizing pre-trained models to generalize predictions across diverse patient populations.
2. **Integration with IoT and Wearables**: Incorporating real-time data from wearable devices to enhance the prediction of CKD progression.
3. **Edge Computing**: Exploring edge-based architectures to enable faster, localized predictions in remote or resource-constrained environments.
4. **Clinical Trials**: Validating the model in real-world healthcare settings to establish its
5. reliability and effectiveness in clinical workflows.

In conclusion, the proposed hybrid CNN-LSTM model with MapReduce integration offers a powerful and scalable solution for CKD prediction. By combining accuracy, scalability, and interpretability, it bridges the gap between cutting-edge AI research and practical healthcare applications. This work lays a foundation for future advancements in AI-driven clinical decision-making tools, ultimately contributing to better healthcare outcomes.

References

[1]. Chen, Y., et al. (2022). "Logistic Regression in CKD Prediction." IEEE Transactions on Healthcare, 6(4), pp. 210-218. DOI: 10.1109/ THC.2022. 1000018.
[2]. Mohan, S., et al. (2023). "Random Forest for CKD Prediction." Journal of Machine Learning Research, 24(1), pp. 100-112.
[3]. Gupta, R., et al. (2021). "SVM for Chronic Disease Prediction." Springer AI in Medicine, 61(3), pp. 255-265. DOI: 10.1007/s10459-021-01234-7.
[4]. Liu, J., et al. (2023). "CNNs for Clinical Data Analysis." Elsevier Biomedical Informatics, 81, pp. 120-130. DOI: 10.1016/ j.jbi.2023.105659.
[5]. Zhang, X., et al. (2023). "LSTMs for Sequential Data Modeling." IEEE Journal of Big Data, 10(2), pp. 100-115. DOI: 10.1109/JBD.2023.1010024.
[6]. Liu, H., et al. (2022). "Hybrid Models for Disease Prediction." ACM Transactions on Healthcare AI, 9(3), pp. 45-59. DOI: 10.1145/3552023.
[7]. Singh, K., et al. (2023). "Transfer Learning for CKD Prediction." Journal of Artificial Intelligence Research, 18, pp. 345-360. DOI: 10.1613/jair.3203.
[8]. Kumar, R., et al. (2023). "MapReduce for Healthcare Analytics." Bioinformatics and Computational Biology, 29(5), pp. 450-465. DOI: 10.1093/bioinformatics/btx654.
[9]. Ramesh, A., et al. (2022). "Hadoop Framework for CKD." Journal of Distributed

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 03 Issue: 03 March 2025
Page No: 358-367
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2025.0050

Systems, 15(4), pp. 340-350. DOI: 10.1016/j.jds.2022.102034.

[10]. [10] Hameed, S., et al. (2022). "Spark for CKD Prediction." Springer AI and Big Data, 3(4), pp. 190-205. DOI: 10.1007/s00521-022-06320-9.

[11]. Lundberg, S. M., et al. (2022). "SHAP for Feature Interpretability." Nature Machine Intelligence, 4(7), pp. 555-565. DOI: 10.1038/s41586-022-05496-3.

[12]. Ribeiro, M. T., et al. (2021). "LIME for Deep Learning Explainability." Springer Interpretability in AI, 8(2), pp. 240-255. DOI: 10.1007/s13272-021-02310-2.

[13]. Wang, F., et al. (2023). "Distributed CNN-LSTM for Disease Prediction." IEEE Transactions on AI in Healthcare, 5(1), pp. 300-315. DOI: 10.1109/ TAIH.2023. 1024567.

[14]. Sun, Q., et al. (2023). "Attention Mechanisms in LSTMs." Elsevier Deep Learning for Health, 39(3), pp. 120-140. DOI: 10.1016/j.dlh.2023.104012.

[15]. Thomas, L., et al. (2022). "Cloud AI Platforms for CKD Prediction." Journal of Cloud Computing, 9(5), pp. 500-515. DOI: 10.1186/s13677-022-00349-0.