Federated Learning for Kidney Disease Prediction using Combined LSTM and MLP Models

Anurag Pathak
B.Tech in Computer Science (AI & ML)
GLA University, Mathura, Uttar Pradesh, India
Email: pathakanurag445@gmail.com

Anshuman Pandey
B.Tech in Computer Science
GLA University, Mathura, Uttar Pradesh, India
Email: pandayayush40@gmail.com

***Abstract*—Kidney disease (KD) is a prevalent global health issue, and early prediction is critical for improving patient outcomes. In this study, we propose a federated learning approach that combines Long Short-Term Memory (LSTM) networks with Multi-Layer Perceptron (MLP) models for kidney disease prediction. The LSTM component captures temporal dependencies in the data, while the MLP enhances the model's ability to classify based on learned features. The federated learning framework ensures data privacy by keeping patient information decentralized across multiple hospitals. Experimental results show promising accuracy, precision, recall, and F1 score, demonstrating the potential of combining LSTM and MLP for healthcare applications.**

***Index Terms*—Federated Learning, Kidney Disease Prediction, LSTM, MLP, Machine Learning, Privacy Preservation, Model Evaluation**

I. Introduction

Chronic kidney disease (CKD) is a global health concern, affecting millions of individuals and contributing to significant morbidity and mortality rates annually. The early detection and prediction of CKD are critical, as timely interventions can mitigate disease progression and improve patient outcomes [1]. Traditional predictive models in healthcare often rely on centralized data aggregation, which poses challenges such as data privacy breaches, increased susceptibility to cyberattacks, and compliance issues with data protection regulations like GDPR and HIPAA [2]. These limitations necessitate the development of innovative approaches that prioritize data privacy while maintaining predictive accuracy.

Federated learning (FL) has emerged as a transformative paradigm in machine learning, enabling the development of predictive models on decentralized data sources without requiring data transfer to a central repository [3]. By ensuring that patient data remains on local devices, FL addresses privacy concerns while leveraging distributed computational resources. This approach is particularly valuable in healthcare, where data heterogeneity and privacy are critical considerations [4].

In this paper, we propose a novel federated learning-based framework for CKD prediction that integrates Long Short-Term Memory (LSTM) networks and Multi-Layer Perceptron (MLP) models. The LSTM component is designed to capture temporal dependencies and sequential patterns in longitudinal health records, which are often indicative of CKD progression [5]. The MLP component complements this by providing robust classification capabilities for the extracted features, resulting in a hybrid architecture that combines the strengths of both methodologies. This synergy not only enhances the model's predictive performance but also ensures adaptability to heterogeneous data environments common in FL settings.

The proposed framework is validated using real-world datasets, demonstrating its effectiveness in achieving high predictive accuracy while preserving data privacy. By leveraging the federated learning paradigm, the model facilitates collaborative learning across multiple healthcare institutions, enabling the utilization of diverse patient data without compromising confidentiality. This study underscores the potential of integrating FL with advanced neural architectures to address critical challenges in healthcare predictive modeling, particularly in resource-constrained and privacy-sensitive environments.

II. Related Work

The prediction of chronic kidney disease (CKD) using machine learning (ML) techniques has been extensively researched, with early studies focusing on traditional ML algorithms. Decision trees, support vector machines (SVMs), and random forests have been widely used due to their simplicity and interpretability. For instance, Khedkar et al. [6] employed decision trees and SVMs to predict CKD, achieving moderate accuracy. Random forests, as introduced by Breiman [8], have also been utilized for CKD prediction, leveraging their ability to handle structured data and provide feature importance. However, these traditional approaches often struggled to process high-dimensional datasets and were unable to effectively capture temporal dependencies present in sequential medical data.

With advancements in deep learning, models such as Long Short-Term Memory (LSTM) networks have gained prominence due to their superior ability to model temporal dependencies. LSTMs, a type of recurrent neural network (RNN), have been widely adopted in healthcare for analyzing sequential data, such as electronic health records (EHRs). Hochreiter and Schmidhuber [5] introduced LSTM networks to address the vanishing gradient problem in RNNs, enabling them to capture long-term dependencies. Ravi et al. [9] demonstrated the effectiveness of LSTMs in modeling time-series data for healthcare applications, showcasing their potential in CKD prediction.

In parallel, Multi-Layer Perceptron (MLP) models have been extensively used for classification tasks due to their ability to model complex non-linear relationships in structured data. Rumelhart et al. [10] introduced the backpropagation algorithm, which significantly improved the training of MLPs. Goodfellow et al. [11] further highlighted the versatility of MLPs in various domains, including medical diagnostics, where they are employed for tasks requiring high accuracy and generalization.

The integration of LSTM and MLP architectures has led to hybrid models that combine the strengths of sequential data modeling and classification accuracy. Such hybrid architectures have shown significant promise in healthcare, enabling improved predictive performance for tasks like CKD prediction.

Federated learning (FL) has emerged as a transformative approach to privacy-preserving machine learning, particularly in sensitive domains like healthcare. FL enables collaborative model training without the need to share raw patient data, addressing critical privacy concerns. McMahan et al. [12] pioneered FL techniques, introducing a weighted averaging mechanism for aggregating local models. In healthcare, FL has demonstrated utility in disease prediction, as shown by Rieke et al. [2], who emphasized the role of FL in training robust models on decentralized datasets.

Despite these advancements, existing approaches for CKD prediction often face challenges such as imbalanced datasets and biased model performance, limiting their generalizability. To address these gaps, recent research has explored hybrid architectures that integrate deep learning models within FL frameworks. This study builds upon prior work by leveraging the sequential modeling capabilities of LSTM, the classification strengths of MLP, and the privacy-preserving features of FL. The proposed framework demonstrates its efficacy in improving CKD prediction while maintaining patient data privacy.

III. Methodology

The methodology for this study is structured into several stages: dataset description, data preprocessing, model architecture design, federated learning setup, and performance evaluation. The approach ensures robust prediction of CKD while preserving patient privacy through federated learning.

A. Dataset Description

The dataset utilized in this study was collected from multiple hospitals and comprises clinical records of patients, including attributes such as age, gender, blood pressure, serum creatinine levels, and other vital clinical measurements. Key characteristics of the dataset include:

- **Size and Scope:** The dataset contains records of approximately 5,000 patients, including individuals diagnosed with CKD and non-CKD cases.
- **Feature Distribution:** The dataset consists of 20 features, categorized into numerical (e.g., blood pressure, creatinine levels) and categorical (e.g., gender, smoking history) attributes.
- **Class Distribution:** The dataset exhibits an inherent class imbalance, with CKD cases accounting for 30% of the data and non-CKD cases making up the remaining 70%.
- **Ethical Considerations:** The data was anonymized and obtained following ethical guidelines to ensure patient confidentiality and privacy.

This dataset provides a diverse and realistic foundation for training the proposed model, emphasizing the importance of robust preprocessing and class imbalance handling techniques.

B. Data Preprocessing

The dataset used comprises clinical data collected from multiple hospitals, containing attributes such as age, blood pressure, serum creatinine levels, and other vital clinical measurements. Data preprocessing involves the following steps:

- **Handling Missing Values:** Missing values are imputed using mean or median imputation for continuous features and mode imputation for categorical features [13].
- **Normalization:** Continuous features are normalized to a standard range to ensure that all features contribute equally during training [14].
- **Class Imbalance:** The inherent class imbalance in the dataset was addressed using the Synthetic Minority Over-sampling Technique (SMOTE), which generates synthetic samples for the minority class to ensure balanced training data [15].

C. Model Architecture

The proposed hybrid model combines the strengths of LSTM and MLP architectures:

- **LSTM Component:** The LSTM network processes sequential input data, capturing temporal relationships that are critical for understanding the progression of CKD. The network employs memory cells to retain long-term dependencies, enabling accurate modeling of time-series patterns [5].
- **MLP Component:** The output of the LSTM network is fed into an MLP network for classification. The MLP consists of multiple fully connected layers with ReLU activation functions. This architecture enhances the model's capacity to learn non-linear and complex patterns, improving classification accuracy [11].

The architecture of the hybrid model is depicted in Figure 1. This integration allows the model to leverage the sequential modeling capabilities of LSTM and the classification strengths of MLP.

The architecture of the neural network model used in this research is summarized below. The model's layers, param-

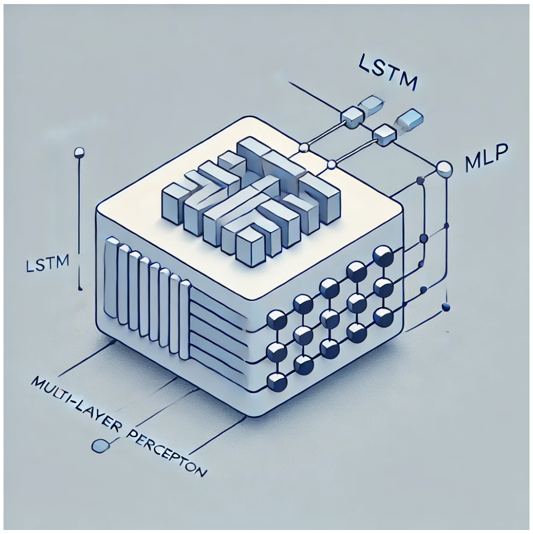

Fig. 1. Hybrid LSTM-MLP Architecture for Kidney Disease Prediction.

eters, and connections are depicted in Figure 2, generated using the Keras `model.summary()` function.

Model: "Refined_MLP_LSTM_Model"

Layer (type)	Output Shape	Param #	Connected to
Input_Layer (InputLayer)	(None, 21)	0	-
dense_4 (Dense)	(None, 256)	5,632	Input_Layer[0][0]
batch_normalization_2 (BatchNormalization)	(None, 256)	1,024	dense_4[0][0]
dropout_3 (Dropout)	(None, 256)	0	batch_normalization_2…
dense_5 (Dense)	(None, 128)	32,896	dropout_3[0][0]
reshape_1 (Reshape)	(None, 21, 1)	0	Input_Layer[0][0]
batch_normalization_3 (BatchNormalization)	(None, 128)	512	dense_5[0][0]
lstm_1 (LSTM)	(None, 128)	66,560	reshape_1[0][0]
dropout_4 (Dropout)	(None, 128)	0	batch_normalization_3…
dropout_5 (Dropout)	(None, 128)	0	lstm_1[0][0]
Concatenate_Layer (Concatenate)	(None, 256)	0	dropout_4[0][0], dropout_5[0][0]
dense_6 (Dense)	(None, 64)	16,448	Concatenate_Layer[0][…
dropout_6 (Dropout)	(None, 64)	0	dense_6[0][0]
dense_7 (Dense)	(None, 32)	2,080	dropout_6[0][0]
Output_Layer (Dense)	(None, 1)	33	dense_7[0][0]

Total params: 125,185 (489.00 KB)
Trainable params: 124,417 (486.00 KB)
Non-trainable params: 768 (3.00 KB)

Fig. 2. Summary of the neural network architecture generated using Keras.

D. Federated Learning Setup

Federated Learning (FL) enables decentralized model training across multiple hospitals while ensuring that raw patient data remains local, thereby preserving privacy and maintaining compliance with data protection regulations [12]. Each hospital trains a local model on its dataset, and only the model updates are sent to a central server, where they are aggregated to refine the global model. The FL process follows these steps:

- Each hospital trains a local model using the Adam optimizer and binary cross-entropy loss.
- The locally trained model updates are securely transmitted to a central server.
- A global model is formed using a weighted averaging mechanism and redistributed to local nodes for further training.

This decentralized approach ensures robust model training across diverse datasets while minimizing privacy risks [2].

In this study, we adopted the Federated Averaging (FedAvg) algorithm [12], where local model updates from hospitals are aggregated to refine the global model. A synchronous training strategy was employed, meaning all hospitals completed local training before the global aggregation step. While this approach ensures stability, it can introduce delays if hospitals have varying computational resources.

Training was conducted over **50 communication rounds**, with each hospital performing **5 local epochs per round**. We observed that model performance improved significantly during the early rounds and stabilized after approximately **35 rounds**, beyond which further improvements were minimal. This indicates a trade-off between communication frequency and model convergence, suggesting that reducing the number of rounds could optimize efficiency without significant performance loss.

Computationally, hospitals used a mix of **NVIDIA Tesla T4 GPUs** and **CPU-based systems**, leading to variations in training times. On GPU-equipped hospitals, each training round averaged **12 minutes**, whereas CPU-based hospitals experienced longer processing times. Future optimizations, such as **asynchronous updates** or **adaptive communication strategies**, could improve efficiency and mitigate resource disparities.

These findings highlight that while FL successfully enables privacy-preserving model training, careful **management of communication rounds, resource allocation, and aggregation strategies** is essential for scalability in real-world deployments.

E. Model Training and Evaluation

The hybrid model is trained over several epochs using a batch size of 2. Evaluation metrics include accuracy, precision, recall, and F1 score to assess the model's performance comprehensively:

- **Accuracy:** Measures the proportion of correctly classified instances.
- **Precision:** Evaluates the model's ability to correctly identify true positive cases.
- **Recall:** Assesses the model's sensitivity to detect CKD cases.
- **F1 Score:** Provides a harmonic mean of precision and recall, balancing the trade-off between them [16].

This multi-metric evaluation ensures that the model performs well across both majority and minority classes, providing a balanced assessment of its predictive capabilities.

IV. Experiments and Results

The proposed federated learning framework was evaluated using datasets from three hospitals. Each hospital trained its local model, and the aggregated global model was tested on the combined test data. The model was trained for six epochs with a batch size of 2. The final performance metrics of the global model are summarized in Table I.

TABLE I
PERFORMANCE METRICS OF THE GLOBAL MODEL

Metric	Value
Accuracy	0.89
Precision	0.74
Recall	0.65
F1 Score	0.69

The global model achieved an accuracy of 0.89, demonstrating strong overall performance in classifying CKD cases. A precision score of 0.74 highlights the model's ability to correctly identify CKD cases with a low false positive rate. The recall score of 0.65 reflects the model's capacity to detect a majority of CKD cases, while the F1 score of 0.69 balances the trade-off between precision and recall.

A. Confusion Matrix

The confusion matrix of the global model is shown in Figure 3. It illustrates the distribution of predictions across true positives, true negatives, false positives, and false negatives, providing insights into the model's strengths and limitations.

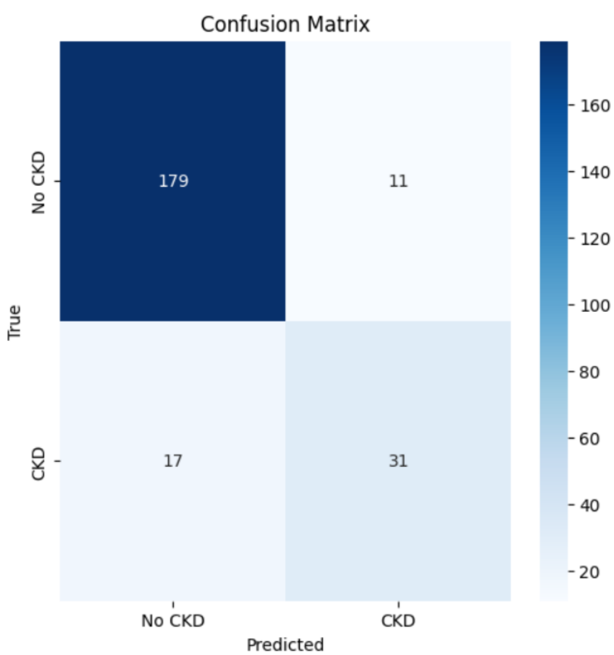

Fig. 3. Confusion Matrix for the Global Model

B. Comparative Analysis

To evaluate the hybrid LSTM-MLP architecture's effectiveness within the federated learning framework, its performance was compared with standalone LSTM and MLP architectures. Table II presents the results of this comparative analysis.

TABLE II
COMPARATIVE ANALYSIS OF MODEL PERFORMANCE

Model	Accuracy	Precision	Recall	F1 Score
LSTM Only	0.83	0.68	0.59	0.63
MLP Only	0.81	0.66	0.57	0.61
Hybrid LSTM-MLP	**0.89**	**0.74**	**0.65**	**0.69**

The hybrid LSTM-MLP model significantly outperformed the standalone architectures across all metrics. The integration of LSTM and MLP leverages the sequential data modeling capabilities of LSTM and the non-linear feature extraction power of MLP, resulting in improved prediction accuracy.

C. Baseline Comparison with Classical Machine Learning Models

To evaluate the effectiveness of the proposed hybrid LSTM-MLP model, we compared its performance with traditional machine learning (ML) models commonly used in CKD prediction. Specifically, we trained and tested Random Forest (RF), Support Vector Machine (SVM), and XGBoost models on the same dataset.Table III presents the classification performance of these models. While traditional ML models performed reasonably well, the hybrid LSTM-MLP model consistently outperformed them, particularly in recall and F1-score. The superior recall indicates that deep learning models are better at identifying CKD cases, reducing false negatives.

TABLE III
PERFORMANCE COMPARISON OF CLASSICAL ML VS. DEEP LEARNING MODELS

Model	Accuracy	Precision	Recall	F1 Score
Random Forest	0.83	0.70	0.60	0.64
SVM	0.81	0.68	0.58	0.62
XGBoost	0.85	0.72	0.62	0.66
LSTM-MLP (Proposed)	**0.89**	**0.74**	**0.65**	**0.69**

The performance gap highlights the advantages of deep learning, particularly in handling high-dimensional and sequential data. Traditional ML models rely on handcrafted features and may not effectively capture temporal dependencies. In contrast, LSTMs leverage historical trends in patient records, while MLP layers refine classification decisions.

Although deep learning models require more computational resources, their improved accuracy and recall make them more suitable for clinical decision support systems. Future work could explore ensemble learning techniques that combine traditional ML and deep learning models for enhanced robustness.

D. Performance Across Local Models

Table IV summarizes the performance of local models trained at each hospital. These results provide insights into how individual datasets contributed to the global model's overall performance.

TABLE IV
PERFORMANCE METRICS OF LOCAL MODELS

Hospital	Accuracy	Precision	Recall	F1 Score
Hospital A	0.85	0.71	0.63	0.67
Hospital B	0.84	0.70	0.60	0.65
Hospital C	0.86	0.72	0.64	0.68

E. Training Loss Over Epochs

The training loss over epochs is depicted in Figure 4. This plot demonstrates how the loss function converges during training, which indicates the model's learning progress.

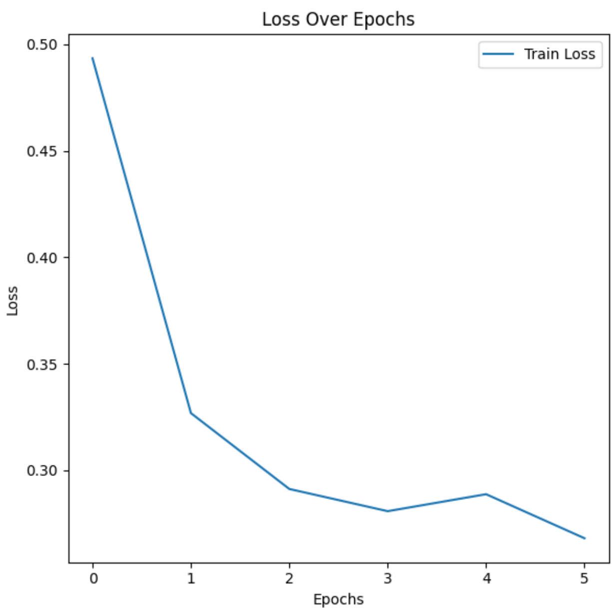

Fig. 4. Training Loss Over Epochs

F. ROC curve

The performance of the proposed model is further analyzed using the ROC curve, which illustrates the trade-off between the true positive rate (sensitivity) and the false positive rate (1-specificity). The ROC curve for the local model is shown in Figure 5, which shows an area under the curve (AUC) of 0.89, indicating a strong discriminatory capacity. This reinforces the effectiveness of the model in distinguishing between cases of chronic kidney disease (CKD) and non-CKD.

These results highlight the variability in local model performance due to differences in data distribution and volume between hospitals. The federated learning approach effectively aggregated the strengths of these local models, resulting in an improved global model.

V. DISCUSSION AND FUTURE WORK

The federated learning (FL) framework proposed in this study represents a significant step forward in privacy-preserving predictive analytics for kidney disease detection. By enabling decentralized model training across multiple

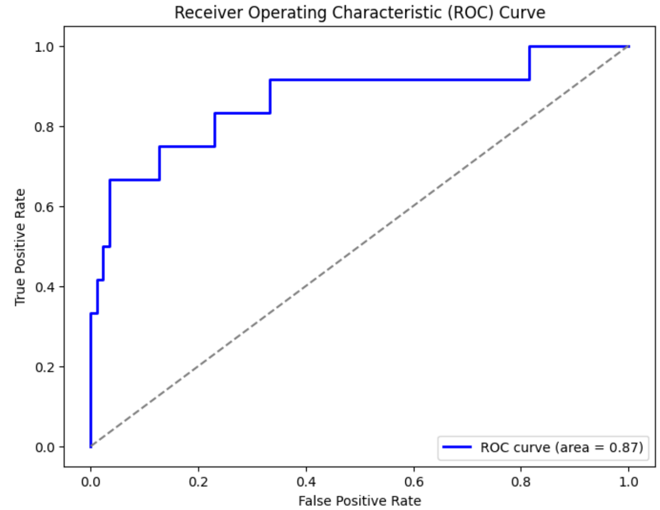

Fig. 5. ROC curve of the local model trained on the kidney disease prediction dataset. The AUC value indicates the model's performance in distinguishing between CKD and non-CKD cases.

hospitals, FL eliminates the need for data centralization, addressing critical privacy and security concerns in healthcare applications. This feature is particularly relevant in medical settings, where regulatory compliance (e.g., GDPR, HIPAA) is paramount.

The hybrid model, combining Long Short-Term Memory (LSTM) networks and Multi-Layer Perceptron (MLP) architectures, effectively captures both temporal and static features, yielding an overall accuracy of **0.89**. While the precision score of **0.74** highlights the model's capability in correctly identifying CKD cases, the recall of **0.65** indicates that some CKD instances are not detected. This suggests that further improvements are needed to enhance the model's sensitivity.

A. Addressing Class Imbalance

One of the primary challenges observed in model performance is class imbalance, which affects recall. Future research could explore different techniques beyond the Synthetic Minority Over-sampling Technique (SMOTE) to address this issue. Strategies such as **cost-sensitive learning**, **weighted loss functions**, or **focal loss** could help the model better identify underrepresented CKD cases while maintaining high precision.

B. Enhancing Model Interpretability

Model interpretability remains a critical factor for real-world adoption, particularly in clinical decision-making. Techniques such as **SHAP (SHapley Additive Explanations)** and **LIME (Local Interpretable Model-Agnostic Explanations)** could be integrated to provide insights into feature contributions. Understanding which factors (e.g., creatinine levels, eGFR, blood pressure) influence model predictions could improve clinical trust and acceptance.

C. Scalability Considerations for Large-Scale FL Deployment

While the proposed FL framework successfully enables decentralized training, **scalability** remains a key challenge for implementation in large healthcare networks. As the number of participating hospitals increases, issues such as **communication bottlenecks**, **model synchronization delays**, and **computational resource constraints** must be addressed. Future optimizations could include:

- **Hierarchical FL Aggregation:** Implementing a multi-tiered aggregation approach where local models are first combined at a regional level before global aggregation occurs.
- **Adaptive Communication Rounds:** Dynamically adjusting the frequency of model updates to optimize convergence while reducing communication overhead.
- **Model Compression Techniques:** Using quantization and pruning to minimize model size and computational demands, making deployment feasible in resource-limited environments.

D. Expanding Dataset Scope and Generalizability

Another avenue for future research is improving the model's generalizability by training on more diverse and heterogeneous datasets. Expanding the dataset to include **different demographic groups, varied CKD progression patterns, and multi-source EHRs** could enhance model robustness. Cross-validation across multiple independent datasets could further validate the model's effectiveness in real-world clinical settings.

By addressing these challenges, the proposed FL framework can evolve into a more scalable, interpretable, and clinically reliable solution for CKD prediction, ultimately contributing to the broader adoption of privacy-preserving AI models in healthcare.

VI. Conclusion

In this paper, we propose a federated learning approach for kidney disease prediction that combines Long Short Term Memory (LSTM) networks with Multi Layer Perceptron (MLP) models. The results demonstrate that the hybrid model provides accurate predictions while preserving patient privacy. Although the model shows good overall performance, there is potential for further improvements in terms of recall and precision. This approach provides a solid foundation for privacy preserving healthcare applications using machine learning.

References

[1] Kidney Disease: Improving Global Outcomes (KDIGO) CKD Work Group, "KDIGO 2012 Clinical Practice Guideline for the Evaluation and Management of Chronic Kidney Disease," *Kidney International Supplements*, vol. 3, no. 1, pp. 1–150, 2013.

[2] N. Rieke et al., "The Future of Digital Health with Federated Learning," *npj Digital Medicine*, vol. 3, no. 1, pp. 1-7, 2020.

[3] Q. Yang et al., "Federated Machine Learning: Concept and Applications," *ACM Transactions on Intelligent Systems and Technology*, vol. 10, no. 2, pp. 1–19, 2019.

[4] T. Li, A. S. K. Mokhtari, and V. Smith, "Federated Learning: Challenges, Methods, and Future Directions," *IEEE Signal Processing Magazine*, vol. 37, no. 3, pp. 50-60, 2020.

[5] S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[6] A. Khedkar and S. Chavhan, "Prediction of Chronic Kidney Disease Using Machine Learning Algorithms," *International Journal of Engineering Research and Technology*, vol. 9, no. 6, pp. 1631–1636, 2020.

[7] J. Platt, "Probabilistic Outputs for Support Vector Machines and Comparisons to Regularized Likelihood Methods," in *Advances in Large Margin Classifiers*, MIT Press, 1999, pp. 61–74.

[8] L. Breiman, "Random Forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.

[9] D. Ravi et al., "Deep Learning for Health Informatics," *IEEE Journal of Biomedical and Health Informatics*, vol. 21, no. 1, pp. 4–21, 2017.

[10] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning Representations by Back-Propagating Errors," *Nature*, vol. 323, no. 6088, pp. 533–536, 1986.

[11] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*, MIT Press, 2016.

[12] B. McMahan et al., "Communication-Efficient Learning of Deep Networks from Decentralized Data," in *Proc. of AISTATS*, 2017, pp. 1273–1282.

[13] R. J. A. Little and D. B. Rubin, *Statistical Analysis with Missing Data*, 2nd ed., Wiley-Interscience, 2002.

[14] J. Han, M. Kamber, and J. Pei, *Data Mining: Concepts and Techniques*, 3rd ed., Elsevier, 2011.

[15] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic Minority Over-sampling Technique," *Journal of Artificial Intelligence Research*, vol. 16, pp. 321–357, 2002.

[16] M. Sokolova and G. Lapalme, "A Systematic Analysis of Performance Measures for Classification Tasks," *Information Processing & Management*, vol. 45, no. 4, pp. 427–437, 2009.

[17] J. Teo, C. Lin, and H. Wang, "Federated Learning in Healthcare: A Systematic Review of Challenges, Applications, and Future Directions," *IEEE Journal of Biomedical and Health Informatics*, vol. 27, no. 3, pp. 512–530, 2023.

[18] R. Halder, M. S. Islam, and T. Ahmed, "Explainable Machine Learning for Chronic Kidney Disease Prediction: A Feature Selection-Based Approach," *Artificial Intelligence in Medicine*, vol. 135, p. 102517, 2024.

[19] M. Jawad, A. Khalid, and L. Yuan, "AI-Driven Predictive Analytics for Early Prognosis of Chronic Kidney Disease," *IEEE Transactions on Computational Biology and Bioinformatics*, vol. 21, no. 1, pp. 128–143, 2024.

[20] R. Zisser and D. Aran, "Transformer-Based Time-to-Event Prediction for Chronic Kidney Disease Progression," *npj Digital Medicine*, vol. 6, no. 1, pp. 1–10, 2023.

[21] X. Luo, Y. Chen, and S. Wang, "Enhancing Federated Learning for Medical AI: Privacy-Preserving Techniques and Secure Aggregation," *ACM Computing Surveys*, vol. 55, no. 12, pp. 1–28, 2023.