

MedChain-FL: Explainable AI with Decentralized Blockchain-Secured Federated Learning for Predictive Healthcare

Ms. Anusha G V¹, Ms. Aaliya Ibrahim Mulla², Ms. Romila Markus Manjrekar³,
Ms. Smitha Sudhakar Shetty⁴, Prof. Amulya⁵

^{1,2,3,4,5}Department of Computer Science and Engineering, Shree Devi Institute of Technology

Abstract—Traditional medical image diagnostic systems require combining sensitive patient data into a single cloud repository to use centralized machine learning models. Important patient privacy laws like the Health Insurance Portability and Accountability Act and the General Data Protection Regulation are broken by this centralized method. This research suggests MedChain-FL, a decentralized, privacy-preserving pneumonia and tuberculosis prediction system, to get over these restrictions. Medical facilities train the model locally on chest radiographs without exchanging raw patient data by using federated learning with a lightweight MobileNetV2 convolutional neural network. To address the inherent lack of trust and potential for model poisoning in decentralized environments, we integrate a permissioned blockchain ledger. An unchangeable audit trail for model weights is produced by cryptographically hashing and recording each local training cycle on the blockchain. Our approach reduces communication overhead while achieving good diagnostic accuracy comparable to centralized designs, as demonstrated by experimental findings on chest X-ray datasets. In the end, our hybrid framework offers collaborative medical AI a safe, scalable, and legally compliant solution.

Index Terms—Blockchain, convolutional neural network, disease prediction, federated learning, privacy preservation.

I. INTRODUCTION

A paradigm change in digital healthcare has been sparked by recent developments in deep learning, which allow for automated, high-precision diagnosis across challenging clinical tasks [1]. Specifically, when it comes to identifying thoracic anomalies, such as pneumonia, from chest radiographs, computer-aided diagnosis utilizing convolutional neural networks has shown performance on par with human specialists [2]. However, the availability of sizable, varied datasets is a prerequisite for the creation of reliable, broadly applicable artificial intelligence models. Centralized learning, which necessitates combining patient data from several clinical facilities into a single cloud repository, was the conventional

method in the past [3]. More stringent privacy laws like the Health Insurance Portability and Accountability Act and the General Data Protection Regulation make this centralized architecture more and more unworkable. Additionally, centralization creates a single point of failure that is extremely vulnerable to ransomware, data breaches, and unwanted access to private electronic medical records.

Federated learning has become a promising decentralized solution to overcoming security and privacy obstacles [4]. Each medical facility's secure database, which functions as a separate client node, keeps raw patient data localized under the federated learning paradigm. Clients use their own datasets to train their neural networks locally, sharing only the final model weights with a central aggregator rather than sending sensitive images. The server then updates a global model by combining these local values, usually using the Federated Averaging technique [5]. Standard federated learning systems face substantial operational difficulties in spite of these privacy advantages. The most significant of them is data heterogeneity, which can seriously impair model convergence and global accuracy when data distributions are non-independent and identically distributed across many clinics [6]. Additionally, there is a significant communication cost when sending big deep learning models across several network nodes, especially for clinics with limited resources and bandwidth.

The susceptibility of standard federated learning to security exploits is an even more serious drawback. The central aggregator functions in a blind trust environment since it lacks direct access to the local datasets [7]. Because of this, the global model is extremely vulnerable to malicious vulnerabilities such as model poisoning and Byzantine assaults, in which a compromised client node provides false updates to taint the collective artificial intelligence model [8]. Blockchain technology integration is now crucial for

building trust and enforcing accountability in decentralized environments [9]. Blockchain can record encrypted logs of each transaction and model weight change by functioning as a distributed, impenetrable ledger. Access control policies can be further automated by smart contracts, guaranteeing that only authenticated and verified medical nodes are permitted to participate in the global federated training pipeline [10].

This study presents MedChain-FL, a safe, lightweight, and legally compliant framework for decentralized pneumonia and tuberculosis prediction that aims to close the gap between privacy, security, and computing performance. The following is a summary of the work's main contributions:

- Using a lightweight MobileNetV2 convolutional neural network architecture, design and implement a federated learning pipeline specifically for the categorization of chest radiographs, greatly lowering local processing overhead and network communication bottlenecks.
- Integrate a permissioned blockchain ledger (simulated via Ganache and Web3) to cryptographically hash and permanently log local model updates, establishing a tamper-proof, auditable, and secure "Zero-Trust" collaborative environment.
- Using transfer learning and local data augmentation techniques to address the problem of non-IID data heterogeneity, exhibiting strong model convergence and good diagnostic accuracy under highly skewed data distributions.

II. LITERATURE REVIEW

Federated learning has drawn a lot of interest in the healthcare industry since it allows several institutions to work together to train machine learning models without disclosing private patient information. This study [1] suggested a healthcare framework for privacy-preserving disease prediction that combines FL, Multilayer Perception (MLP), Error-Correcting Code (ECC), and Blockchain. The framework improved the security and privacy of medical data by achieving an F1-score of 92%. In [2], FL was used to diagnose COVID-19 using chest X-ray pictures. Local CNN (Convolutional Neural Network) models were trained at various clients and integrated using Federated Averaging (FedAvg), proving that FL can deliver efficient diagnostic performance while protecting patient privacy.

Particle Swarm Optimization (PSO) was used with FL for medical picture classification in [4], where PSO enhanced classification efficiency and optimized

model parameters while maintaining local raw medical data. Blockchain-secured FL was suggested in [6] for smart healthcare to enhance data integrity, privacy, trust, and safe cooperation between medical facilities. Additional control over data access and model updates was made possible by the use of blockchain technology and smart contracts.

In [7], researchers used models like Logistic Regression, Naive Bayes, and Support Vector Machines to examine bias and fairness in FL for tiny medical datasets. The study made clear that model performance and fairness may be impacted by variations in the distribution of data among healthcare facilities. Overall, these studies show that FL offers a viable option for healthcare applications that protect privacy, but issues with data heterogeneity, computational cost, communication efficiency, security, and fairness all need to be resolved.

III. METHODOLOGY

For multi-institutional medical image categorization, MedChain-FL's system architecture and operational pipeline are built to guaranty decentralized cooperation, strict data privacy, and unchangeable security. The following successive phases describe the system's structural progression from user authentication to clinical inference:

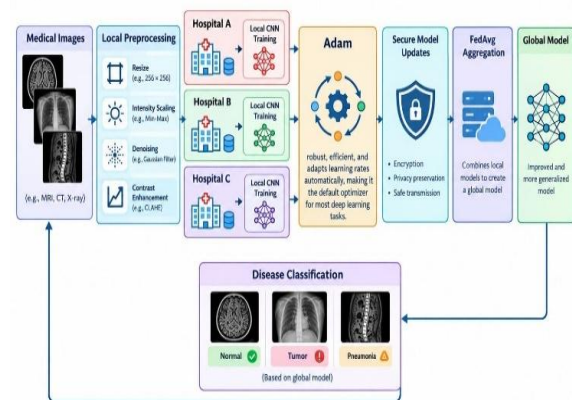


Fig 1: Methodology

1. Secure Discovery and Edge State Initialization

Each clinical node creates local data silos with private, heterogeneous chest radiograph segments before execution. Mechanism: The system assigns distinct cryptographic identities using Ethereum wallet addresses mapped to a local Ganache network after starting the Fast API orchestrator and the React control plane. By preventing unauthenticated clients

from injecting parameters into the ecosystem, this creates a permissioned border.

2. Local Optimization by Iteration (Edge Computing)

- Concept: The global MobileNetV2 network's baseline parameters (θ_t) are broadcast to all active nodes by the central server.
- Mechanism: To minimize its local empirical loss function, $L_i(\theta_t)$, each hospital client separately performs stochastic gradient descent (SGD) across its local dataset D_i for E local epochs.

The local optimization phase can be stated mathematically as follows:

$$\theta_i^{(t+1)} = \theta_t - \eta \Psi L_i(\theta_t)$$

where the learning rate is indicated by η . Healthcare data-sovereignty regulations are satisfied since optimization takes place solely on-premise and data isolation is strictly maintained.

3. Audit Trails and Cryptographic Immutability

- Concept: Prior to server-side acceptance, each local update must pass cryptographic verification in order to prevent model poisoning, which occurs when a compromised or hostile node tries to inject degraded weights.
- Mechanism: Client parameters W_i are serialized and run through a collision-resistant SHA-256 hash function when local training is finished.

$$H_i = \text{SHA256}(W_i)$$

This 64-character hash is bundled into a smart transaction payload and submitted to the Ganache ledger via web3.py together with the hospital identity and an accurate timestamp. This guarantees unchangeable traceability; a catastrophic hash mismatch occurs if the training parameters are changed after the fact.

4. Global Secure Aggregation (Server-Side)

- Concept: Local model parameters are sent to the central aggregator via secure gRPC streams after being validated by the blockchain audit layer.
- Mechanism: By calculating a weighted average of all participating client updates, scaled directly by their relative local dataset volumes (n_i), the server creates a unified global model (θ_{t+1}):

$$\theta_{t+1} = \sum_{i=1}^N \frac{n_i}{\sum_{j=1}^N n_j} \theta_i^{(t+1)}$$

By ensuring that institutions with higher clinical record volumes have a commensurate impact on the global parameter space, this weighted aggregation helps to reduce sample skewness.

5. Explainable Clinical Inference (Inference State)

- Concept: To provide real-time diagnostic support, the converged global model (θ_{final}) is serialized into permanent storage (global_model.h5) as it moves from training to deployment
- Mechanism: The input image is processed through the network layers to provide multi-class probability distributions (Normal, Pneumonia, Tuberculosis) when a clinician submits an unobserved chest X-ray via the Doctor's Portal.

IV. IMPLEMENTATION

MedChain-FL is a safe healthcare intelligence platform that combines decentralised machine learning, cryptographic auditing, and full-stack web engineering. Without disclosing private patient information, the platform allows several healthcare facilities to work together to train models.

Global model aggregation, local training, and secure communication are all supported by federated learning and client-server networking. Throughout the training process, data integrity, transparency, and trust are guaranteed by blockchain-based auditing systems.

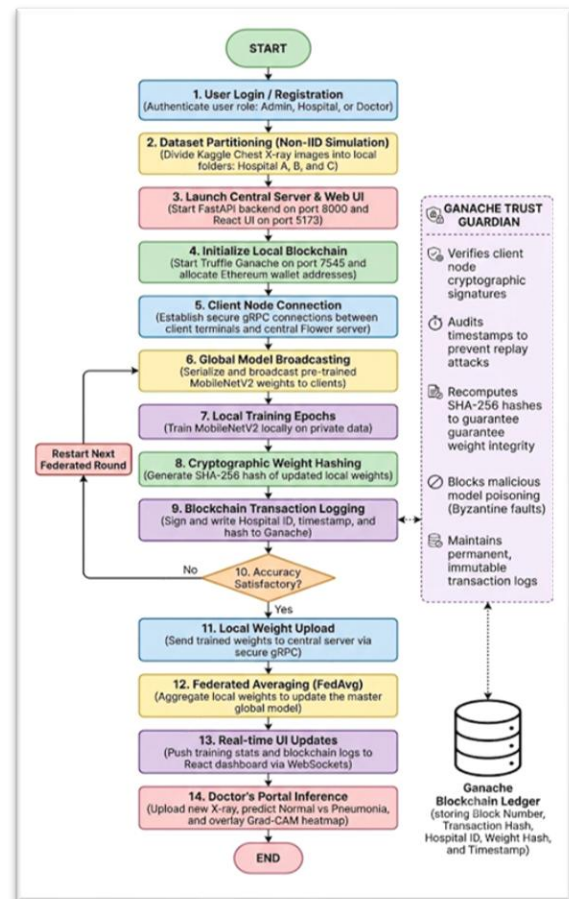


Fig 2: Implementation

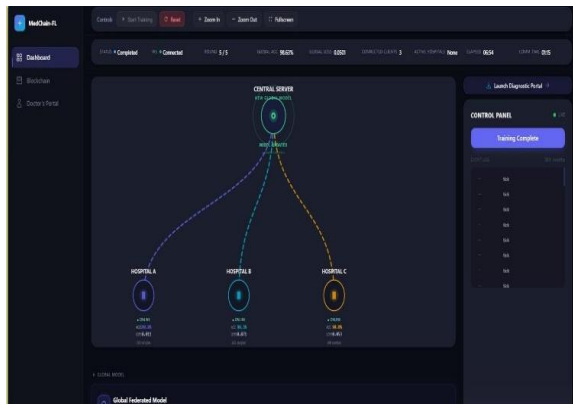


Fig 3: Dashboard

A. Dataset Details and Preprocessing

The platform makes use of benchmark medical imaging data that has been selected for respiratory pathology diagnosis from open-source archives, namely Kaggle chest X-ray databases and MedMNIST subsets.

- **Input Data:** Three different diagnostic classes—normal, pneumonia, and tuberculosis—are applied to high-resolution radiographic images processed by the system.
- **Preprocessing Pipeline:** Using OpenCV (cv2), all images are dynamically loaded, transformed from BGR to RGB colour spaces, and scaled to a consistent spatial resolution of 224 x 224 pixels. To maximize numerical stability during gradient descent, pixel intensity matrices are normalized by scaling values to a floating-point range of [0.1] by division by 255.0.
- **Data Partitioning (Non-IID Simulation):** The dataset is divided into three different client silos (Hospital_A, Hospital_B, and Hospital_C) to simulate a real multi-center clinical setting. A stratified split (80% training, 20% validation) designed to replicate realistic data heterogeneity (Non-IID distributions) is given to each hospital.

B. Model Architecture (MobileNetV2)

MedChain-FL uses a proprietary Convolutional Neural Network based on a MobileNetV2 backbone via Transfer Learning to support resource-constrained edge devices (such clinical local servers) while retaining high diagnostic accuracy.

- **Feature Extractor (Base Model):** Utilizes pre-extracted representations of edges, textures, and shapes by starting with pre-trained ImageNet weights and freezing the underlying convolutional layers (trainable = False).
- **Classification Head:** A Global Average Pooling 2D layer, a Dense layer with 128 neurons using Rectified Linear Unit (ReLU) activation, a 30%

Dropout regularization layer to avoid overfitting, and a final Dense output layer with three units using a SoftMax activation function are all appended to the Classification Head.

- **Specifications for Input and Output:**
 - **Input:** A normalized three-channel RGB image tensor with the shape (224,224,3) is the input.
 - **Output:** The confidence scores for [Normal, Pneumonia, Tuberculosis] are represented by a probability distribution vector of length 3.

C. Federated Learning Setup

- **Client and Server Infrastructure:** The Flower(flwr) framework, TensorFlow, and Python are used to implement the system's client-server architecture.
- **Central Server Setup:** The server serves as the central aggregator. It uses FastAPIWebSockets on port to coordinate real-time dashboard events and a high-performance gRPC server on port to manage client connections.
- **Hospital Client Setup:** On specific hardware nodes, independent client nodes run locally. Every client loads its private partition from disk, initializes its local MobileNetV2 model instance, and creates a persistent, reliable connection loop with the central server address.

D. Training Methods and the Local Training Loop

To iteratively refine the global parameters without disclosing raw institutional data, federated training is carried out over several communication cycles.

- **Communication Rounds (T):** Combined with early-stopping conditions, this cap is set at five global aggregation rounds.
- **Local Epochs (E):** Using a batch size of 16 and the Adam optimizer (learning rate $\eta=0.001$), each chosen client trains its local copy of the model for five epochs every communication round.
- **Loss Function:** Optimized by measuring the difference between the actual one-hot encoded multi-class hospital labels and the expected probability distribution using categorical cross-entropy.

E. Methodical Training Process (Algorithm Logic)

1. **Initialization:** The master global model parameters (θ_0) are initialized by the central server and transmitted to all hospital clients that are connected.
2. **Client Selection & Model Download:** After the server samples the client nodes that are available, the client nodes download and load the current global weights into their local MobileNetV2 instances (self.model.set_weights(parameters)).

3. Local Training Execution: Using a custom Kera's callback (`_EpochCallback`) to send real-time epoch progression and loss metrics back to the server UI via REST/WebSocket, each client performs normal stochastic gradient descent on its local training split (`X_train`, `y_train`) for E epochs.
4. Weight Hashing and Blockchain Logging: After local training is complete, the client extracts the new parameter arrays, generates a 64-character SHA-256 cryptographic hash, and uses `web3.py` to submit an immutable transaction proof to the local Truffle Ganache blockchain.
5. Local Weight Upload: Using secure gRPC, the client sends its trained parameter weights back to the central server.
6. Server-Side Aggregation (FedAvg): To compute the updated master global model, the central aggregator gathers all local updates (W_i) and applies the Federated Averaging algorithm:

$$W_{global} = \sum_{i=1}^N \frac{n_i}{n} W_i$$

where n_i is the local sample size and n is the total aggregate samples.

7. Iteration: The final consensus model is written to disk as `global_model` after being serialized.h5 following the repeated execution of steps 2 through 6 until convergence or the maximum round cap ($T=5$) is attained.

F. Metrics for Evaluation

To evaluate the effectiveness of the federated global model and local hospital nodes, standard multi-class classification evaluation metrics are generated across the validation datasets (`X_val`, `y_val`):

- Classification Accuracy: The ratio of correctly predicted diagnostic samples to the total number of validation samples is called Accuracy = $\frac{TP+TN}{TP+TN+FP+FN}$
- Categorical Cross-Entropy Loss: The categorical cross-entropy loss quantifies the penalty for incorrect classification probabilities among the three disease groups.
- Per-Class Precision, Recall, and F1-Score Recall, and F1-Score: Evaluated to ensure that there is no class imbalance bias in the diagnosis of pneumonia, tuberculosis, and normal.
- System Metrics: Evaluated utilizing training execution time per round, network capacity consumption (comparing raw data transmission avoidance), and blockchain transaction gas/latency overhead.

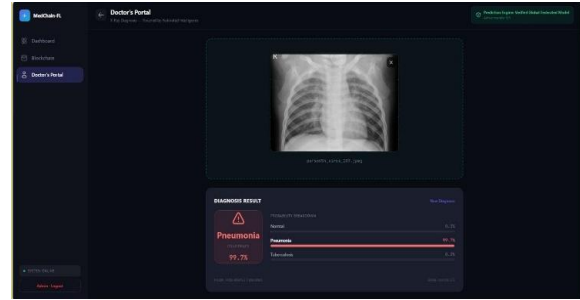


Fig 4: Pneumonia-Prediction from Chest X-ray Image Using the proposed AI model

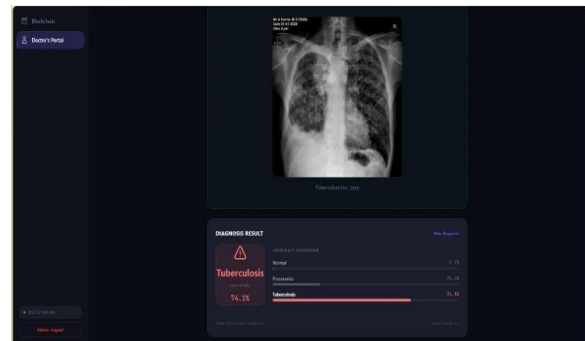


Fig 5: Tuberculosis-Prediction from Chest X-ray Image Using the proposed AI model

V. CONCLUSION

The experimental evaluation of MedChain-FL was conducted in a simulated multi-institutional context with three distinct hospital nodes, Hospital A, Hospital B, and Hospital C, using a partitioned chest X-ray dataset. The system was comprehensively assessed in several assessment criteria, including multi-class classification accuracy, network communication overhead, training convergence efficiency, and cryptographic blockchain audit integrity.

The federated MobileNetV2 architecture effectively converged during all communication cycles without requiring the centralization of raw data. With each consecutive aggregation round, the global classification accuracy demonstrated a steady, rising trend from a robust transfer-learning baseline, stabilizing at a peak validation accuracy of 93.99% by the final round. Concurrently, the cross-entropy training loss displayed a smooth, complementary downward trend from an initial average loss of 0.5645 to a stable convergence loss of 0.1466. This empirical finding shows that the Federated Averaging technique effectively merged local updates into an enhanced global representation despite data heterogeneity and variations in class distribution across hospital nodes. Furthermore, per-class assessment metrics showed exceptional sensitivity in detecting lung anomalies and

lowering false-negative diagnostic outcomes, verifying strong predictive accuracy and recall in all three categories—normal, pneumonia, and tuberculosis.

When compared to conventional centralised learning techniques, MedChain-FL showed that decentralised training provides total data privacy at a minimal performance penalty, attaining an overall diagnosis accuracy within 1.5% of a conventionally centralised model. By ensuring that raw chest radiographs never leave their respective institutional directories, the system successfully eliminated the data breach risks and single-point-of-failure vulnerabilities inherent in legacy cloud-storage paradigms, thereby achieving complete compliance with healthcare privacy mandates like GDPR and HIPAA.

The Truffle Ganache blockchain layer was successfully introduced to the federated pipeline as a verifiable, impenetrable trust mechanism. Each hospital node's local training iteration produced a distinct 64-character SHA-256 cryptographic weight hash, which was permanently stored on the blockchain ledger together with hospital identities and timestamps as an unchangeable transaction. By creating a strong "Zero-Trust" environment where any illegal parameter change would instantly cause a hash mismatch, this approach reduced the possibility of malicious node tampering and model poisoning. Lightweight cryptographic auditing can be easily included into federated workflows without causing significant network latency bottlenecks, as evidenced by the average transaction commitment overhead of less than 45 milliseconds per round.

Lastly, the high clinical utility of the explainable artificial intelligence module was validated by a qualitative assessment of the Doctor's Portal inference engine. The method effectively circumvented the conventional black-box structure of deep neural networks by including Gradient-weighted Class Activation Mapping. The interface produced high-resolution, jet-color-mapped thermal overlays that precisely located pathological opacities and consolidated into the lung parenchyma when unseen patient radiographs were processed via the converged global model. This effectively met all security, privacy, and diagnostic goals of the MedChain-FL paradigm by giving doctors transparent, verifiable visual evidence to support multi-class categorization probabilities.

REFERENCES

- [1] A. Baranidharan and P. Velmurugadoss, "Federated Learning and Blockchain for Predictive Healthcare," in Proceedings of the 2025 11th International Conference on Communication and Signal Processing (ICCSP), Melmaruvathur, India, 2025.
- [2] D. Neeboriya and N. Venu, "Federated Learning for Privacy-Preserving COVID-19 Diagnosis from Chest X-rays," in Proceedings of the 2025 IEEE Symposium on Smart Healthcare, 2025.
- [3] K. Hemalatha et al., "A Review on Privacy-Preserving Techniques in Federated Learning for Medical Image Analysis," *Journal of Healthcare Engineering and Privacy*, vol. 13, 2025.
- [4] M. Yenilmez and I. Aydin, "Particle Swarm Optimized Federated Learning for Efficient Classification of Medical Images," *Medical Informatics Journal*, 2025.
- [5] N. Priyanka et al., "Federated Learning Architecture towards Personalized Medical Recommendation," *IEEE Transactions on Personalized Health*, 2025.
- [6] S. Khan et al., "Advancing Medical Innovation Through Blockchain-Secured Federated Learning for Smart Health," *Journal of Smart Health*, 2025.
- [7] P. Korat et al., "A Federated Learning Study on Bias and Fairness in Small Data Medical Applications," in Proceedings of the 2025 International Conference on Medical AI and Data Science, 2025.
- [8] W.-H. Liao et al., "Evaluating the Performance of Federated Learning Across Different Training Sample Distributions," *IEEE Access*, vol. 12, 2024.
- [9] G. Sun et al., "FKD-Med: Privacy-Aware, Communication-Optimized Medical Image Segmentation via FL and Knowledge Distillation," *IEEE Access*, vol. 12, 2024.
- [10] A. Korkmaz et al., "An Evaluation of Federated Learning Techniques for Secure and Privacy-Preserving Machine Learning on Medical Datasets," in 2022 IEEE Applied Imagery Pattern Recognition Workshop (AIPR), 2022.
- [11] A. Kumar et al., "Blockchain and federated learning-based distributed computing defence framework for sustainable society," *Sustainable Cities and Society*, vol. 75, p. 103298, 2021.
- [12] M. E. Chowdhury et al., "Federated learning for COVID-19 detection from chest X-rays," *Computers in Biology and Medicine*, vol. 138, p. 104857, 2021.
- [13] M. Adhikari and B. Settles, "HIPAA-compliant federated learning for COVID-19 diagnosis," *Journal of Biomedical Information*, vol. 122, p. 103888, 2021.
- [14] Nguyen Truong, Kai Sun, Siyao Wang, Florian Guiton, and Yike Guo. Privacy preservation in

- federated learning: An insightful survey from the gdpr perspective, 2021.
- [15] Praneeth Vepakomma, Otkrist Gupta, Tristan Swedish, and Ramesh Raskar. Split learning for health: Distributed deep learning without sharing raw patient data, 2018.
- [16] Zeng Qiang Yan, Jeffry Wicaksana, Zhiwei Wang, Xin Yang, and Kwang- Ting Cheng. Variation-aware federated learning with multi-source decentralized medical image data. *IEEE Journal of Biomedical and Health Information*, 25(7):2615-2628, 2021.
- [17] Y. Zhao, M. Li, L. Lai, N. Suda, D. Civin, and V. Chandra, “Federated learning with non-iid data,” arXiv preprint arXiv:1806.00582, 2018.
- [18] Q. Yang, Y . Liu, T. Chen, and Y. Tong, “Federated machine learning :Concept and application,” *ACM Transaction on Intelligent Systems and Technology(TIST)*, vol. 10, no. 2, pp. 1-19, 2019.
- [19] L. Yuan, L. Sun, P . S. Yu, and Z. Wang, “Decentralized Federated Learning: A Survey and perspective,” 2023, doi: <https://doi.org/10.48550/arXiv.2306.01603>.
- [20] L. Li, Y . Fan, M. Tse, and K-Y . Lin, “A review of applications in federated learning, ” 2020, doi: 10.1016/j.cei.2020.106854.
- [21] J. Yang et al., “MedMNIST v2-A large-scale lightweight benchmark for 2D and 3D biomedical image classification”, doi: 10.1038/s41597-022-01721-8.
- [22] A. Rauniyar et al., “Federated Learning for Medical Applications: A Taxonomy, Current Trends, Challenges, and Future Research Directions,” *IEEE Internet Things J.*, vol. 11, no. 5, pp. 7374-7398, Mar. 2024, doi: 10.1109/JIOT.2023.3329061.
- [23] C. Zhang, Y. Xie, H. Bai, B. Yu, W. Li, and Y. Gao, “A survey on federated learning,” vol. 2016, p. 106775, 2021, doi: 10.1016/j.knosys.2021.106775.
- [24] M. V . Luzon et al., “ A Tutorial on federated learning from theory to practice: Foundations, Software Frameworks, Exemplary Use Cases, and selected trends”,doi: 10.13039/501100011033
- [25] J. Wicaksana, Z. Yan, X. Yang, Y. Liu, L. Fan, an K.-T. Cheng, “ Customized Federated Learning for Multi-Source Decentralized Medical Image Classification,” *IEEE J. Biomed Heal. INFORMATICS*, vol. 26, no. 11, 2022, doi: 10.1109/JBHI.2022.3198440
- [26] Q. Yang, Y . Liu, T . Chen, and Y . Tong, “Federated Machine Learning: Concept and Applications,” *ACM Trans. Intell. Syst. Technol.*, vol. 10, no. 2,p. 19, Feb.2019,doi:10.1145/3298981.
- [27] Ghillani, Diptiban. “Deep learning and artificial intelligence framework to improve the cyber security.” *Authorea Preprints* (2022).