RESEARCH ARTICLE

# Secure and Decentralized Heart Sound Analysis using Federated Learning and Blockchain Technology

**Shivani Agarwal**[1], **Meenakshi Kandpal**[2], **Vishan Kumar Gupta**[3,*], **Shubhani Aggarwal**[4] **and Paras Jain**[5]

[1] Ajay Kumar Garg Engineering College, Ghaziabad, Uttar Pradesh, India
[2] Odisha University of Technology and Research, Bhubaneswar, Odisha, India
[3] Amity School of Engineering & Technology, Amity University Punjab, Mohali, Punjab, India
[4] University of Petroleum and Energy Studies, Dehradun, Uttarakhand, India
[5] VIT Bhopal University, Kothrikalan, Sehore, Madhya Pradesh, India

## Abstract

**Early diagnosis of cardiac abnormalities depends on accurate classification of heart sounds, but centralized training methods run the danger of violating patient privacy. We thus propose a privacy-preserving and reliable heart sound abnormality detection system combining Blockchain Technology with Federated Learning (FL). Training is spread among seven clients, each simulating an independent data source, using a preprocessed dataset from the PhysioNet Challenge 2016 to enable distributed learning without sharing raw data. CNN-LSTM model using FedAvg achieved the best performance: 94% accuracy, 0.90 precision, 0.96 recall, and an AUC of 0.98 among five deep learning architectures evaluated with FedAvg and FedProx strategies. Along with metadata including client ID and round number, SHA-256 hashes of local and global model weights were recorded on a local Ethereum blockchain following every communication round to guarantee model integrity. The hash of the final model is revalidated against the blockchain to confirm authenticity prior to deployment. It then guarantees safe, distributed, clinically valuable AI-based diagnostics by real-time classification of heart sounds as normal or abnormal.**

**Keywords**: federated learning, blockchain-based model validation, heart sound classification, privacy-preserving artificial intelligence, CNN-LSTM architecture, SHA-256 model integrity, decentralized healthcare diagnostics.

## 1 Introduction

Particularly in helping with early diagnosis and decision support, the use of artificial intelligence (AI) in healthcare has expanded fast recently. One such field is heart sound analysis, which is very important in spotting possible cardiac problems before they become more severe. Big amounts of heart sound data are being gathered as wearable monitoring devices and digital stethoscopes become more common [1]. Sharing this information with centralized systems, however, begs for major questions about patient privacy, data abuse, and regulatory compliance.

**Submitted:** 05 October 2025
**Accepted:** 30 October 2025
**Published:** 16 November 2025

**Vol.** 2, **No.** 1, 2026.
10.62762/TMI.2025.567350

***Corresponding author:**
Vishan Kumar Gupta
meenakshikandpal14@gmail.com

**Citation**
Agarwal, S., Kandpal, M., Gupta, V. K., Aggarwal, S., & Jain, P. (2025). Secure and Decentralized Heart Sound Analysis using Federated Learning and Blockchain Technology. *ICCK Transactions on Machine Intelligence*, 2(1), 1–11.

In sensitive fields like healthcare, traditional machine learning approaches—which demand centralizing data for training—are progressively seen as unworkable [2].

By allowing the training of machine learning models across several hospitals or devices without ever transferring patient data, federated learning (FL) presents a hopeful substitute [3]. This method keeps data safely on-site and lets every institution help the model to be improved. FL poses fresh difficulties despite its privacy-preserving character: it is hard to verify whether model updates are real or whether they have been altered during the training process [4]. Ensuring the trustworthiness and integrity of models is just as crucial in medical uses where decisions can impact life as in ensuring their accuracy.

We present a new system combining Blockchain Technology with Federated Learning to generate a transparent, tamper-proof training environment [5], so addressing this problem. Every contributing device or hospital teaches a local model on its own heart sound data. Following training, every model update produces a distinct fingerprint—known as a cryptographic hash—which is generated and recorded on a blockchain ledger together with metadata including client trained the model and at which round. This guarantees that the whole training background is verifiable and cannot be changed, so strengthens the dependability and security of the model for clinical application.

Using a heart sound dataset obtained from the PhysioNet Challenge 2016, preprocessed and distributed among seven clients to replicate real-world hospital environments in our system. We tested CNNs and LSTMs among other deep learning models meant to process audio signals. Among these, the CNN+LSTM model trained with FedAvg showed the most consistent and accurate outcomes [6]. Like a cardiologist listening for abnormalities during a physical examination, this architecture catches the minute sound patterns of the heart as well as their sequence of occurrence.

Every variant of the model—local as well as global—is tracked on a local Ethereum blockchain using smart contracts, so guaranteeing security. The fingerprint of the model is checked against the blockchain to confirm that no manipulation has happened before it is applied for real-time diagnosis. Once confirmed, the model can evaluate fresh heart sound inputs and generate instantaneous predictions: 0 for normal and 1 for aberrant [7, 8]. This configuration supports distributed, real-time, privacy-conscious cardiac screening without violating model integrity or data confidentiality.

All things considered, our work fills a significant demand for reliable, safe, and interpretable artificial intelligence in the medical domain. Combining Federated Learning with Blockchain lets healthcare providers work together to create strong diagnostic models without compromising patient privacy or data control. Our system not only increases the accuracy of heart sound classification but also generates confidence in technology, something crucial for practical acceptance in clinical environments [9].

### 1.1 Objective

The goal of this study is to create a privacy-preserving and secure heart-sound abnormality identification system based on Federated Learning and Blockchain Technology. With this approach, decentralized training of models becomes achievable on different clients without the sharing of raw medical data, thus preserving patient privacy. In order to enhance trust and integrity in the system, blockchain technology is used to save cryptographic hashes of model updates, thus making the process verifiable and tamper-free. This system overcomes major challenges such as data leakage and model tampering to provide accurate, secure, and reliable AI-based diagnostics in the healthcare industry.

### 1.2 Contributions

Our work mostly makes the following important contributions:

- Combining Federated Learning and Blockchain Technology, it suggests a privacy-preserving and reliable heart sound abnormality detection system.
- Using the FedAvg approach over seven distributed clients, it develops and implements a CNN+LSTM model, obtaining good performance in real-time heart sound classification.
- To guarantee model integrity, traceability, and tamper-evidence all through the training process, it combines blockchain to document SHA-256 hashes of local and worldwide model updates.
- Using the verified global model, it provides real-time diagnosis and accurate predictions of

either normal or aberrant heart sounds without compromising sensitive patient data.

### 1.3 Organization

The remainder of the article is organized in this manner. "Related Works" gives a summary of earlier research on heart sound classification, federated learning applications in healthcare, and blockchain-based model verification. The "Proposed Framework" describes the system's overall architecture, with a focus on how federated learning and blockchain technology can be combined to offer private and secure training. "System Architecture" describes the technical components, including training workflow, model design, data preparation, and blockchain integration. "Implementation" explains how the framework is implemented, with a focus on the CNN+LSTM model, federated training across seven clients, and smart contract-based logging. "Results and Discussion" discusses the system's performance through the use of quantitative metrics and the results' interpretability. The section "Challenges and Future Work" discusses existing constraints and suggests potential avenues for further advancement. Lastly, "Conclusion" provides a summary of the study's key findings and contributions.

## 2 Related Work

Especially with developments in machine learning (ML), deep learning (DL), and privacy-preserving systems like Federated Learning (FL), heart sound abnormality detection has become a fundamental domain within biomedical signal processing. Over a wide range of methods, including signal segmentation, unsupervised clustering, and anomaly detection across noisy and multi-institutional datasets, the examined literature. Based on abrupt changes in heart sound signals, Tatulli et al. [10] presented an unsupervised segmentation method. Reaching top-tier F1 scores across PhysioNet, CirCor, and PASCAL datasets, the method—which requires little parameter tuning—validated its resilience against inter-database variability.

Using 12-lead ECG data, Jimenez et al. [11] investigated federated learning for arrhythmia classification. Their work focused on training across several data silos without sharing raw patient data, so preserving privacy. Particularly under both IID and non-IID environments, the performance of the FL-based model was found to be rather similar to that of centralized models. Integrating wavelet reconstruction, convolutional autoencoders, and one-class SVM, Zeng et al. [12] proposed WCOS, a hybrid anomaly detection system). Designed to solve noise interference and sample imbalance, WCOS outperformed classical semi-supervised models in AUC standard deviation under noisy environments.

Combining horizontal and vertical FL techniques catered for multi-institutional heart sound databases helped Qiu et al. [13] progress this domain. Their framework shows that cooperative training can be both safe and diagnostically effective by matching feature spaces and safeguarded model interpretability. Using wavelet packets, Karan et al. [14] presented a fresh Hilbert-domain characterization. Using packet-level instantaneous frequency and energy deviations alongside ECOC with SVM/KNN classifiers, their classification framework achieves nearly perfect UAR metrics on two standard PCG datasets.

Following time-frequency and statistical feature extraction then PSO and SFFS-based feature selection, Sadeghi et al. [15] By means of improved feature engineering and SMote-based balancing, their system—evaluated on the PhysioNet 2016 dataset—achieved 98.03% accuracy, exceeding previous benchmarks. Applying data augmentation and ML/DL hybrid classifiers, Abbas et al. [16] underlined resilience against noisy heart sound signals. The multilayer perceptron model achieved a high 95.65% accuracy on noisy subsets from the PASCAL challenge dataset by means of their feature ensemble combining MFCCs and spectrograms.

These analyses show overall the advantages of combining modern artificial intelligence methods with traditional signal processing [17]. Table 1 lists the approaches, keywords, and contributions of every work referenced. These cited papers together provide the technological basis and practical relevance of federated detection systems and automated auscultation. They offer important new perspectives and benchmarks that directly guide the design and execution of our anomaly detection in heart sound recording systems.

## 3 Proposed Framework

The model outlined in the diagram (Figure 1) employs a federated learning system to determine whether heart sounds are normal or abnormal, using a hybrid CNN-LSTM architecture and integrating blockchain for improved security and transparency.

The approach begins with an unprocessed data

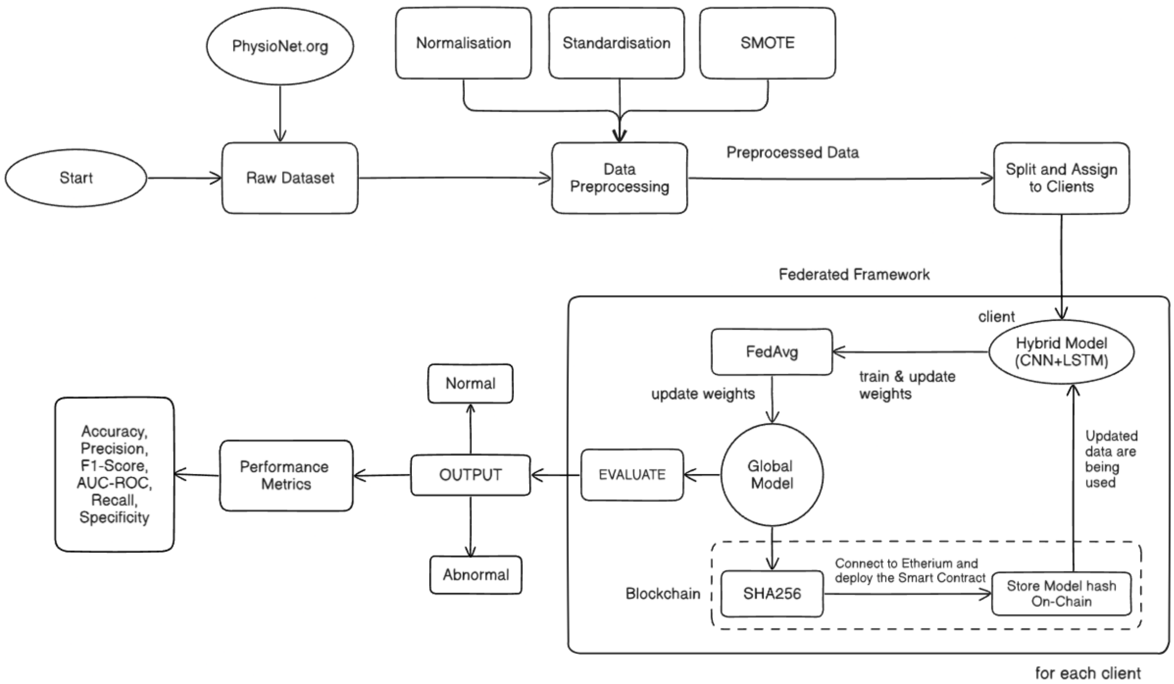


**Figure 1.** Diagram of the proposed framework.

set downloaded from PhysioNet.org, which is preprocessed with normalization, standardization, and SMOTE operations to ensure data consistency and class balance prior to distribution to different clients. Each client trains a separate CNN-LSTM model, where the CNN block captures spatial patterns and the LSTM block captures temporal sequences from the heart sound data, and the localized updates are combined into a global model as per the FedAvg algorithm. The obtained global model classifies the heart sounds as "Normal" or "Abnormal," and its performance is measured using metrics like Accuracy, Precision, F1-Score, AUC-ROC, Recall, and Specificity. For security purposes, the hash of the model is fetched based on SHA256, allocated to the Ethereum blockchain through a smart contract, and stored on-chain for each client. The approach thus ensures accurate heart sound classification while using federated learning for privacy of data and blockchain technology for secure and transparent tracking of model updates, thus rendering it suitable for healthcare applications.

### 3.1 System Architecture Information

Integrating Federated Learning (FL) with Blockchain-based auditability, this part shows the system architecture for our privacy-preserving and trustworthy Heart Sound Abnormality Detection framework. The system is designed to operate in a clinically realistic, distributed environment whereby raw patient data stays local to every healthcare node and insights from each help to create a shared worldwide intelligence — the global model.

#### *3.1.1 Federated Learning Framework and Client Simulation*

Our architecture is based on a Federated Learning (FL) paradigm, meant to replicate a distributed healthcare environment in which sensitive patient data stays strictly local yet helps to build a globally optimal diagnosis model. Medical situations where institutional autonomy and data confidentiality are major constraints are especially suited for this paradigm. Following preprocessing, the whole dataset was split up among seven simulated healthcare clients, each of which stood for an independent institution—such as clinics or hospitals. Most importantly, neither the central server nor any client exchanged any raw data. Rather, every client trained a local model using its own private dataset acting as an autonomous learning node [18].

To investigate the adaptability of federated optimization algorithms across temporal and spatial learning environments, the local models at each client were instantiated with different deep learning architectures including CNN-LSTM — hybrid structures, Long Short-Term Memory (LSTM), and CNN [19].

It is crucial to underline, though, that these architectures define the local computational units inside the larger FL pipeline only. Federated algorithms dominated the learning process itself, most famously:

- FedAvg, sometimes known as federated averaging, averages local weights across clients following each round.
- FedProx (Federated Proximal) is a variation on FedAvg with a proximal term meant to reduce variation across client data distributions.

Every client engaged in twelve communication rounds, locally training for twenty years in every round. Model parameters, not data, were delivered to the central server for aggregation into a global model following each round.CNN-LSTM architecture trained using FedAvg produced the most strong and generalizable performance across several clinical metrics among all architecture–algorithm combinations [20].

All things considered, the foundation of the system is the federated structure rather than the model architecture. Under a cooperative FL framework, every model serves as a local learner guaranteeing data privacy, inter-institutional cooperation, and safe convergence toward a single diagnostic intelligence [21].

### 3.2 Global Model and Aggregation Protocol

Our system exploits a shared global model as the knowledge hub instead of a centralized vector store or external memory module. Local model weights from all seven customers are sent to the federated server following every training round, where they are aggregated using FedAvg (and in some cases FedProx) to generate an updated global model [22].

**Federated Averaging (FedAvg)**: FedAvg combines the weights of local models by computing a weighted average:

$$w^{t+1} = \sum_{k=1}^{K} \frac{n_k}{n} w_k^t \tag{1}$$

where $w^{t+1}$ represents the updated global weights after round $t$, $w_k^t$ denotes the local weights from client $k$ at round $t$, $n_k$ is the number of samples at client $k$, and $n = \sum_{k=1}^{K} n_k$ is the total number of samples across all $K$ clients.

This approach guarantees that clients having more data have correspondingly more impact on the global model.

**Federated Proximal (FedProx)**: FedProx introduces a proximal term to account for heterogeneous data distributions:

Each client minimizes the following objective:

$$\min_w \left\{ f_k(w) + \frac{\mu}{2} \|w - w^t\|^2 \right\} \tag{2}$$

where $f_k(w)$ is the local loss function at client $k$, $w^t$ represents the global weights from the previous round, $\mu$ is the regularization coefficient, and the term $\|w - w^t\|^2$ penalizes large deviations from the global model.

This term ensures stability in convergence when clients have non-IID data. Without ever access to patient-level raw signals, this global model develops over time accumulating generalized representations from distributed datasets. The revised weights are returned to the clients, so maintaining the loop of privacy-preserving collaborative learning.

### 3.3 Blockchain Integration for Model Provenance

We built a blockchain-based provenance mechanism into our system to handle the hazards of model tampering, data manipulation, and unverifiable training contributions.

After each local training round ends:

1. Every client computes a SHA-256 cryptographic hash of its trained model weights:

   $$h = \text{SHA-256}(\text{Serialize}(w)) \tag{3}$$

   where $\text{Serialize}(w)$ denotes the byte-string representation of model weights, $h$ is the fixed-length (256-bit) hash output, and SHA-256 ensures that even the slightest change in model weights yields a completely different hash.

2. The following metadata is then immutably stored on the blockchain via a custom Solidity smart contract:
   - Client ID
   - Training Round Number
   - UTC Timestamp
   - Local Model Hash hhh

3. After aggregation, the global model is also hashed and recorded on-chain:

   $$h_g = \text{SHA-256}(\text{Serialize}(w'^*)) \tag{4}$$

This process creates a tamper-proof audit trail, allowing anyone to verify:

- Which client participated in which round
- What version of the model was deployed
- Whether a given model instance matches its blockchain-stored fingerprint

Blockchain anchoring together with federated optimization guarantees that our heart sound classification system is not only accurate but also auditable and reliable.

This generates a tamper-proof ledger including provenance for every model version. A stakeholder can verify the integrity of the last model used and track which client made which contribution at any future point.

### 3.4 Verification and Inference Pipeline

The last model is hash-based verified before deployment: its current hash is computed anew and matched against the hash kept on-chain. It is used for inference only if the hash matches, so verifying the absence of tampering with the model. New heart sound inputs are preprocessed identically for real-time diagnosis and then fed to the verified global CNN-LSTM model to generate a classification output:

- 0 → Normal
- 1 → Abnormal

Respecting strict data privacy rules, these outputs gain from a collective intelligence developed from several scattered institutions.

## 4 Implementation

### 4.1 Dataset Used

Real-world phonocardiogram (PCG) recordings annotated for either normal or abnormal heart function come from the initial dataset available on https://physionet.org/content/challenge-2016/1.0.0/.The dataset calls for 5012 rows and 26 columns. The federated training environment could access the safely kept dataset on a cloud drive.

### 4.2 Dataset Preprocessing

The PhysioNet/Computing in Cardiology Challenge 2016 provided the dataset used in this project; labeled phonocardiogram (PCG) recordings annotated as either normal or abnormal. First the dataset was randomly mixed to guarantee objective model training and remove any natural ordering trends. Originally denoted as -1 for normal and 1 for aberrant, the label encoding was remapped to a binary format of 0 and 1, respectively, to conform with standard binary classification criteria.

Feature scaling then was done with z-score normalizing the StandardScalar module. This procedure guaranteed that every feature in all samples had a mean of zero and a standard deviation of one, so encouraging more stable and effective convergence during training. The feature matrix was rebuilt from a two-dimensional structure into a three-dimensional tensor with the shape (samples, timesteps, features), since the models used in this project—especially CNN and LSTM—need sequential data inputs. This change was essential to guarantee fit with 1D convolutional and recurrent neural network layers, so allowing the system to record spatial and temporal patterns in cardiac sound signals.

### 4.3 Implementation of Federated Learning

To simulate a privacy-preserving healthcare environment, the preprocessed dataset was divided among seven clients, each representing a decentralized medical institution. Federated training was implemented using the workflow illustrated in Figure 2:

- Every client started its model—e.g., LSTM, CNN, CNN+BILSTM, CNN+LSTM—and trained it on a local data partition.
- Clients trained their local models for twenty Epochs each round. There were twelve rounds of communication all overall.
- Data for forthcoming rounds will come from the revised weights instead of depending just on their local data in communicational rounds.
- Client models sent just their weights to a central server following every round. To generate a fresh global model, these weights were aggregated applying the Federated Averaging (FedAvg) technique.

### 4.4 Implementation of Blockchain Technology

Integrated into the federated learning architecture was a blockchain-based logging mechanism to guarantee model integrity and offer an auditable training history. Every customer calculates a SHA-256 hash of its serialized model weights following every training round. This hash is the model's distinct fingerprint

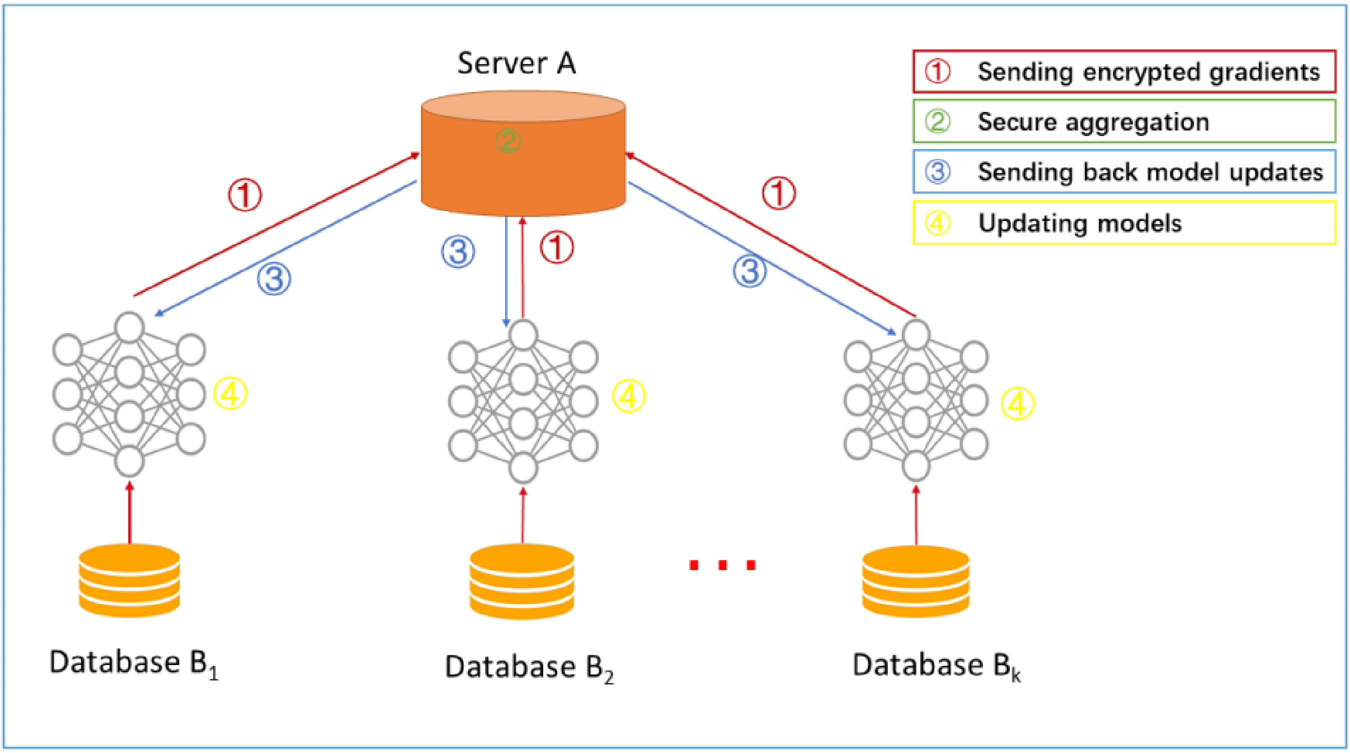


**Figure 2.** Overview of federated learning architecture.

at that round.Every customer logs the following metadata on-chain using a Solidity smart contract housed on a local Ethereum blockchain:

- Client ID
- Round Number
- UTC Timestamp
- Model SHA-256 of local weights

Once FedAvg aggregation forms the global model, it is also hashed and has a fingerprint stored on the blockchain together with pertinent metadata. This process guarantees traceability, verifiable, tamper-evident local and global model versions, so supporting reliable deployment in sensitive uses including healthcare artificial intelligence.

### 4.5 Evaluation

Following all the rounds of communication, the final global model was saved and then evaluated on a held-out test set that was not used in training. The model was expected to make binary predictions, where:

- 0 = Normal cardiac auscultation sound
- 1 = Abnormal cardiac auscultation

The decision of classification was made based on the sigmoid activation function of the last dense layer of the CNN+LSTM model that generates a score probability between 0 and 1.

### 4.6 Metrics

To comprehensively assess the performance of the model, the following evaluation metrics were employed:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{5}$$

Accuracy represents the overall proportion of correctly classified samples.

$$\text{Precision} = \frac{TP}{TP + FP} \tag{6}$$

Precision measures the proportion of true positive predictions among all predicted positives.

$$\text{Recall} = \frac{TP}{TP + FN} \tag{7}$$

Recall (Sensitivity) indicates the proportion of true positive cases correctly identified by the model.

$$\text{F1 Score} = \frac{2 \cdot TP}{2 \cdot TP + FP + FN} \tag{8}$$

F1 Score is the harmonic mean of precision and recall, balancing false positives and false negatives.

$$\text{Specificity} = \frac{TN}{TN + FP} \quad (9)$$

Specificity measures the proportion of true negative predictions among all actual negatives. AUC (Area Under the ROC Curve) measures the model's ability to distinguish between normal and abnormal heart sounds. A higher AUC indicates better discrimination capability. Log Loss evaluates the uncertainty of predictions by penalizing overconfident incorrect classifications. Lower log loss values indicate better-calibrated probability estimates.

These metrics collectively provide a holistic understanding of the model's performance, particularly in a clinical context where both sensitivity (recall) and specificity are critical for reliable diagnosis.

## 5 Results and Discussion

### 5.1 Results for FedAvg Aggregation

Apart from contrasting neural network designs, we investigated the impact of several federated optimization strategies on general model performance. Although the fundamental model architecture (e.g., CNN-LSTM) stayed constant across experiments, the choice of optimization strategy—namely FedAvg, FedProx, and FedOpt—was varied to grasp how each handles distributed training under heterogeneous data conditions. FedAvg obtained the best overall results across all performance criteria, as the table below shows; FedProx gave somewhat better regularization in some cases. Although adaptive in nature, FedOpt underperformed relative to the other two, underscoring that its gradient-tuning benefits may not generalize well to sensitive, non-IID medical data. This analogy emphasizes the need to choose a federated approach fit for the clinical accuracy requirements and data distribution.

**Table 1.** Comparison across different aggregation techniques of federated learning.

| Strategy | Acc | Precision | Recall | F1-Score |
|---|---|---|---|---|
| FedAvg | 0.941 | 0.91 | 0.95 | 0.93 |
| FedProx | 0.912 | 0.87 | 0.91 | 0.89 |
| FedOpt | 0.885 | 0.83 | 0.89 | 0.85 |

Table 1 shows a comparative analysis of three popular federated optimization algorithms—FedAvg, FedProx, and FedOpt—against chosen performance metrics using deep learning for heart sound anomaly detection. The results demonstrate that FedAvg is the most efficient overall, with an accuracy of 94.1% and an F1-score of 0.93, demonstrating superior generalizability over decentralized datasets. While FedProx shows slightly reduced accuracy, it shows superior precision and recall, thanks to its resistance to client variability. In contrast, FedOpt shows slightly inferior performance on all metrics measured, suggesting that its adaptive optimization parameters are not as effective in this specific medical application. These results support the fact that FedAvg, combined with CNN-LSTM architecture, is the most reliable choice for this federated learning task.

### 5.2 Model Architecture Comparison under FedAvg

Several deep learning models were investigated under the same FedAvg strategy in order to assess how neural architecture affects federated learning performance. Every model was independently trained over distributed clients under the same training parameters. Hybrid models such as CNN-LSTM and CNN-BiLSTM outperformed stand-alone CNN, LSTM, and BiLSTM architectures as compiled in Table 2 and illustrated in Figure 3. With 94% accuracy and an F1-score of 0.93, CNN-LSTM performed the best among them in capturing both spatial and temporal patterns in heart sound signals, as clearly demonstrated in Figure 3. This emphasizes in federated medical artificial intelligence systems the need of model architecture.

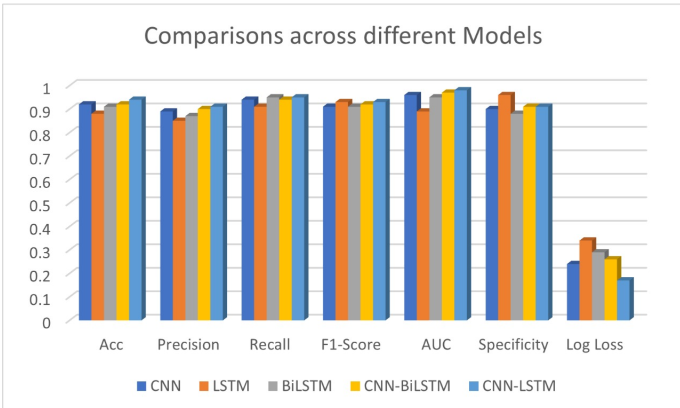


**Figure 3.** Comparisons among the different Models.

In Table 2 and Figure 3, five federated learning models—namely CNN, LSTM, BiLSTM, CNN-BiLSTM, and CNN-LSTM—are tested on seven metrics: accuracy (acc), precision, recall, F1-score, AUC, specificity, and log loss. Remarkably, the CNN-LSTM

**Table 2.** Comparison across different deep learning models using federated learning

| Models | Acc | Precision | Recall | F1-Score | AUC | Specificity | Log Loss |
|---|---|---|---|---|---|---|---|
| CNN | 0.92 | 0.89 | 0.94 | 0.91 | 0.96 | 0.90 | 0.24 |
| LSTM | 0.88 | 0.85 | 0.91 | 0.93 | 0.89 | 0.96 | 0.34 |
| BiLSTM | 0.91 | 0.87 | 0.95 | 0.91 | 0.95 | 0.88 | 0.29 |
| CNN-BiLSTM | 0.92 | 0.90 | 0.94 | 0.92 | 0.97 | 0.91 | 0.26 |
| CNN-LSTM | 0.94 | 0.91 | 0.95 | 0.93 | 0.98 | 0.91 | 0.17 |

model has the lowest log loss of 0.17, thereby proving its supremacy by attaining the highest values in accuracy (0.94), precision (0.91), recall (0.95), F1-score (0.93), and AUC (0.98), as visually confirmed in Figure 3. Therefore, the CNN-LSTM model is the best performer in this study in general since it has the lowest error and improved predictive power.

## 6 Challenges and Future Works

Even though the suggested system meets its core purposes, there is room for future development and expansion. There are several areas of future work that could improve its scalability, usage, and performance. Some suggestions are:

- **Evaluating the generalizability of the framework across several datasets**: Although the results from the PhysioNet 2016 dataset show superior performance for the system, its results on real-world and heterogeneous clinical datasets still need to be confirmed. The future studies can compare the model on larger, multi-center heart sound datasets with higher clinical and demographic heterogeneity. The system's stability in real-world clinical practice and the assessment of its generalizability would be determined by this comparison.

- **Exploring system scalability in real federated system**:The present setup mimics federated clients; yet real-world deployment of the system in actual distributed healthcare settings, e.g., hospitals or rural clinic settings—may be hindered by practical challenges like non-homogeneous network latency, hardware constraints, and data skewness. Future research could explore applying the framework in real-world settings to evaluate performance under varying run conditions.

- **Improving resource efficiency and communication effectiveness**: Federated learning is bandwidth and computationally expensive since it takes several rounds of communication. Applying model compression, client sampling, and asynchronous update, experimentation with different techniques can render the system more deployable on low-resource edge devices, such as digital stethoscopes or mobile units.

- **Growing blockchain adoption and effectiveness**: Blockchain adds further computational and storage overhead, even as it offers proof of tamper-evident training. Future work may consider more scalable alternatives, including permissioned or Layer-2 blockchains, or layering several Layer-2 solutions to restrict on-chain congestion with guarantees of verifiability.

- **Enhancing model clarity and clinical openness**: Doctors need to be able to trust AI predictions if their application in medicine is to be a success. Particularly in cases of borderline or uncertain heart sound cases, the incorporation of XAI tools into the system will allow practitioners to see and understand model decisions. By overcoming such obstacles, the system can be a more scalable, clinically valid, and secure solution for cardiac screening based on artificial intelligence, and hence greatly promote privacy-sensitive healthcare innovation in real-world applications.

## 7 Conclusion

In this work, we have combined Federated Learning with Blockchain Technology to create a safe and privacy-preserving system for heart sound anomaly detection. The system guarantees that sensitive medical data stays local and protected by allowing distributed model training among several clients. Strong accuracy in correctly classifying heart sounds as normal or aberrant was shown by a CNN+LSTM model trained on FedAvg. Following every communication round, cryptographic hashes of local and worldwide model updates were entered on a blockchain to preserve the integrity and openness of the training process. This method forbids illegal changes and lets post-training verification possible. The last confirmed model guarantees clinical relevance, trust, and responsibility as well as real-time predictions. All things considered, this framework provides a scalable, interpretable, and safe answer for cardiac screening driven by artificial intelligence in actual medical settings.

## Data Availability Statement

Data will be made available on request.

## Funding

This work was supported without any funding.

## Conflicts of Interest

The authors declare no conflicts of interest.

## Ethical Approval and Consent to Participate

Not applicable.

## References


[1] Yaqoob, M. M., Nazir, M., Yousafzai, A., & Khan, M. A. (2022). Modified artificial bee colony-based feature optimized federated learning for heart disease diagnosis in healthcare. *Applied Sciences, 12*(23), Article 12080. [CrossRef]

[2] Bashir, A. K., Victor, N., Bhattacharya, S., Huynh-The, T., Chengoden, R., Yenduri, G., ... & Liyanage, M. (2023). Federated learning for the healthcare metaverse: Concepts, applications, challenges, and future directions. *IEEE Internet of Things Journal, 10*(24), 21873-21891. [CrossRef]

[3] Alasmari, S., AlGhamdi, R., Tejani, G. G., Sharma, S. K., & Mousavirad, S. J. (2025). Federated learning-based multimodal approach for early detection and personalized care in cardiac disease. *Frontiers in Physiology, 16*, 1563185. [CrossRef]

[4] Khan, R., Taj, S., Ma, X., Noor, A., Zhu, H., Khan, J., ... & Khan, S. U. (2024). Advanced federated ensemble internet of learning approach for cloud based medical healthcare monitoring system. *Scientific Reports, 14*(1), 26068. [CrossRef]

[5] Gupta, M., Kumar, M., & Gupta, Y. (2024). A blockchain-empowered federated learning-based framework for data privacy in lung disease detection system. *Computers in Human Behavior, 158*, 108302. [CrossRef]

[6] Raza, A., Tran, K. P., Koehl, L., & Li, S. (2022). Designing ECG monitoring healthcare system with federated transfer learning and explainable AI. *Knowledge-Based Systems, 236*, 107763. [CrossRef]

[7] Saidi, R., Rahmany, I., Dhahri, S., & Moulahi, T. (2024). A privacy-enhanced framework for chest disease classification using federated learning and blockchain. *IEEE Access*. [CrossRef]

[8] Qayyum, A., Ahmad, K., Ahsan, M. A., Al-Fuqaha, A., & Qadir, J. (2022). Collaborative federated learning for healthcare: Multi-modal covid-19 diagnosis at the edge. *IEEE Open Journal of the Computer Society, 3*, 172-184. [CrossRef]

[9] Sun, L., Wu, J., Xu, Y., & Zhang, Y. (2023). A federated learning and blockchain framework for physiological signal classification based on continual learning. *Information Sciences, 630*, 586-598. [CrossRef]

[10] Tatulli, E., Souriau, R., & Fontecave-Jallon, J. (2025). Unsupervised segmentation of heart sounds from abrupt changes detection. *Computers in Biology and Medicine, 186*, Article 109712. [CrossRef]

[11] Jimenez Gutierrez, D. M., Hassan, H. M., Landi, L., Vitaletti, A., & Chatzigiannakis, I. (2023). Application of federated learning techniques for arrhythmia classification using 12-lead ECG signals. In *International Symposium on Algorithmic Aspects of Cloud Computing* (pp. 38–65). Springer Nature Switzerland. [CrossRef]

[12] Zeng, P., Kang, S., Fan, F., & Liu, J. (2025). Enhanced heart sound anomaly detection via WCOS: a semi-supervised framework integrating wavelet, autoencoder and SVM. *Frontiers in Neuroinformatics, 19*, 1530047. [CrossRef]

[13] Qiu, W., Quan, C., Zhu, L., Yu, Y., Wang, Z., Ma, Y., ... & Schuller, B. W. (2024). Heart sound abnormality detection from multi-institutional collaboration: Introducing a federated learning framework. *IEEE Transactions on Biomedical Engineering, 71*(10), 2802-2813. [CrossRef]

[14] Karan, B., & Padhi, T. (2024). Hilbert domain characterizations of wavelet packets for automated heart sound abnormality detection. *Biomedical Signal Processing and Control, 90*, Article 105793. [CrossRef]

[15] Nia, P. S., & Hesar, H. D. (2024). Abnormal heart sound detection using time-frequency analysis and machine learning techniques. *Biomedical Signal Processing and Control, 90*, 105899. [CrossRef]

[16] Abbas, S., Ojo, S., Al Hejaili, A., Sampedro, G. A., Almadhor, A., Zaidi, M. M., & Kryvinska, N. (2024). Artificial intelligence framework for heart disease classification from audio signals. *Scientific reports, 14*(1), 3123. [CrossRef]

[17] Li, X., Huang, K., Yang, W., Wang, S., & Zhang, Z. (2020). On the convergence of fedavg on non-iid data. In *International Conference on Learning Representations (ICLR)*. [CrossRef]

[18] Fallah, A., Mokhtari, A., & Ozdaglar, A. (2020). Personalized federated learning with theoretical guarantees: A model-agnostic meta-learning approach. In (NeurIPS). [CrossRef]

[19] Liu, Y., Kang, Y., Xing, C., Chen, T., & Yang, Q. (2020). A secure federated transfer learning framework. *IEEE*

*Intelligent Systems*, *35*(4), 70-82. [CrossRef]

[20] Xu, R., Baracaldo, N., & Ludwig, H. (2021). Federated learning for healthcare: A comprehensive review. ACM Computing Surveys, 54(10s), 1-36. [CrossRef]

[21] Chen, H., Yuan, Z.,& Wang, Y. (2023). Explainable AI for cardiac sound classification: Interpreting deep learning decisions in telehealth applications. IEEE Transactions on Biomedical Engineering, 70(5), 1567-1578. [CrossRef]

[22] Wang, T., Lu, G., & Zhang, K. (2024). Robust heart sound detection in noisy environments using self-supervised learning and data augmentation. *Biomedical Signal Processing and Control*, *87*, 105512. [CrossRef]

**Shivani Agarwal** is currently working as an Associate Professor in Department of Information Technology, Ajay Kumar Garg Engineering College, Ghaziabad. She Completed her Ph.D & M.Tech in Computer Science and Engineering from Uttarakhand Technical University, Dehradun. B.Tech in IT from UPTU, Lucknow. Dr. Shivani Agarwal has more than 18 years experience in Teaching and Research. Her areas of interest are Soft Computing, Machine Learning and Bioinformatics. She has Published many Research papers in different reputed International Journals and Conferences. (Email: agarwalshivani@akgec.ac.in, kasishivani@gmail.com)

**Meenakshi Kandpal** is an Assistant Professor in the Computer Science and Engineering Department at OUTR, Bhubaneswar, India. She earned her M.Tech degree from Uttarakhand Technical University, Dehradun, and completed her Ph.D. in Computer Science and Engineering from KIIT Deemed to be University, Bhubaneswar. With over 15 years of experience in teaching and research, Dr. Meenakshi is a seasoned academic and an innovative researcher. Her expertise spans a wide array of courses, including Programming in C, Database Management Systems, Artificial Intelligence, Computer Networks, IoT, Data Mining, Machine Learning, and Real-Time Systems. Dr. Meenakshi's research interest lies at the intersection of cutting-edge technologies, focusing on Blockchain, Machine Learning, Cloud Computing, Geo-spatial Data, and Serverless Computing. Meenakshi has made significant contributions to the field through numerous publications, including book chapters, conference papers, and articles in prestigious international journals and patents. Her dedication to advancing technology and her passion for education continue to inspire her students and peers alike. (Email: meenakshikandpal14@gmail.com)

**Vishan Kumar Gupta** has received his Ph.D. degree in Computer Science and Engineering from the Thapar Institute of Engineering and Technology, Patiala, Punjab, India. He has received his M.Tech degree in Information Technology from ABV-Indian Institute of Information Technology and Management, Gwalior, MP, India, and his Bachelor of Engineering in CSE from Rajiv Gandhi Technical University, Bhopal, MP, India, Currently, he is working as an Associate Professor in Amity University Punjab, Mohali, India. His areas of research are Drug Toxicity, Data Mining, Machine learning, and Deep Learning. (Email: vishangupta@gmail.com)

**Shubhani Aggarwal** received her Ph.D. in CSE from Thapar Institute of Engineering and Technology (Deemed to be University), Patiala, Punjab, India and was a postdoctoral research fellow in École de Technologie Supérieure, Montréal, Québec, Canada. She is working as an Assistant Professor-III in SOCS, UPES, Dehradun, India. Her research interests are in the areas of Blockchain, cryptography, Internet of Drones, and information security. She has published a Book titled "The Blockchain Technology for Secure and Smart Applications across Industry Verticals" in Elsevier Publication. She has also published more than 50 research papers in top-cited journals such as IEEE Transactions on Vehicular Technology, IEEE Transactions on Industrial Informatics, IEEE Transactions on Intelligent Transportation Systems, IEEE Internet of Things, Elsevier Journal Networks of Computer and Applications, IEEE Access, Computers and Security, Mobile Networks and Applications, Computer Communications and many more. She has published a book based on Blockchain Technology in Advances in Computers, Elsevier. (Email: shubhaniaggarwal529@gmail.com)

**Paras Jain** has received his Ph.D. degree in Computer Science and Engineering from Jaypee University of Engineering & Technology, Guna, MP, India. He has received his Master of Engineering degree in Computer Engineering from Shri Govindram Seksaria Institute of Technology and Science, Indore, MP, India and his Bachelor of Engineering in CSE from Rajiv Gandhi Technical University, Bhopal, MP, India, Currently, he is working as an Associate Professor Grade-2 in School of Computing Science Engineering and Artificial Intelligence, VIT Bhopal University, Kothrikalan, MP, India. His areas of research are Image Processing, Data Mining, Machine learning, and Deep Learning. (Email: paras.jain@vitbhopal.ac.in)