Attention-Driven Hierarchical Federated Learning for Privacy-Preserving Edge AI in Heterogeneous IoT Networks

Pournima Pande[1], Bukya Mohan Babu[2], Poonam Bhargav[3], T L Deepika Roy[4], Elangovan Muniyandy[5], Prof. Ts. Dr. Yousef A.Baker El-Ebiary[6], Dr. V Diana Earshia[7]

Department of Applied Chemistry, Yeshwantrao Chavan College of Engineering, Nagpur, India[1]
Department of CSE (Data Science), CMR Technical Campus, Hyderabad, Telangana, India[2]
Lecturer, Department of Computer Science-College of Engineering and Computer Science, Jazan University, Saudi Arabia[3]
Department of Computer Science & Engineering, Koneru Lakshmaiah Education Foundation, Green Fields, Vaddeswaram, A.P. – 522302, India[4]
Department of Biosciences-Saveetha School of Engineering. Saveetha Institute of Medical and Technical Sciences, Chennai - 602 105, India[5]
Applied Science Research Center, Applied Science Private University, Amman, Jordan[5]
Faculty of Informatics and Computing, UniSZA University, Malaysia[6]
Assistant Professor/ECE, Vel Tech Rangarajan Dr.Sagunthala R&D Institute of Science and Technology, Chennai, India[7]

***Abstract*—ECG arrhythmia detection is very important in identification and management of patients with cardiac disorders. Centralized machine learning models are privacy invasive, and distributed ones poorly deal with the data heterogeneity of the devices. These challenges are responded to by presenting the edge AI an attention-driven hierarchical federated learning framework with 1-Dimensional Convolutional Neural Network (1D-CNN) - Long Short-Term Memory (LSTM) -Attention to classify arrhythmia in ECG recordings. This model includes the spatial characteristics of ECG signals and the temporal characteristics of attention maps, identifying the significant areas of the inputs and providing high interpretability and accuracy of the model. Thus, federated learning is applied to perform model training in a decentralized process through the Privacy-Preserving while the raw data remains on the edge devices. For assessment, this study utilized St. Petersburg INCART 12-lead Arrhythmia Database and Wearable Health Monitoring has given an overall classification accuracy of 96.5% with an average of AUC-ROC of 0.98 with five classes as Normal (N), Supraventricular (S), Ventricular (V), Fusion (F), and Unclassified (Q). The proposed model was created using the Python programming language with the TensorFlow framework deep learning and tested using Raspberry Pi devices to mimic edge settings. Overall, this study proves that it is possible to classify using IoT Device ECG arrhythmia reliably and securely on devices with limited resources, which will enable real-time cardiac monitoring.**

***Keywords*—Edge AI; federated learning; wearable health monitoring; arrhythmia; privacy-preserving; IoT device; 1DCNN-LSTM**

I. INTRODUCTION

The advancement of IoT over the recent years has contributed to an unprecedented inflow of interconnected, smart devices in the network periphery creating masses of data that need processing[1]. The applications running on such devices include multiple lifesaving operations, such as healthcare, transport systems, and households[2],[3]. Most central processing and analyzing this information would be less effective because bandwidth utilization[4], delays, and privacy concerns tend to be critical obstacles. One of the basic issues in edge AI deployment is balancing user privacy with model performance [5]. Centralized learning paradigms, which depend on aggregating raw data in a single server, are heavily privacy-invasive and many cases, infeasible because of bandwidth and latency constraints [6]. To overcome the challenges by allowing collaborative model training on distributed devices without local data sharing, federated learning (FL) method is utilized [7]. Nevertheless, traditional FL methods are not optimally suited for heterogeneous IoT networks. It tends to experience non-IID (non-independent and identically distributed) data distributions, heterogeneity in device capabilities, and uneven participation owing to constrained resources[8].These drawbacks severely compromise model convergence and accuracy.

The approaches in healthcare and IoT, models like the Observational Medical Outcomes Partnership Common Data Model (OMOP-CDM) have offered a single platform for organizing and analyzing patients' data [9]. Similarly, federated learning models need standardized frameworks and adaptive coordination protocols that can facilitate effective model training on diverse IoT devices [10]. But current FL structures are generally flat and do not take advantage of hierarchical relations between edge devices, thus creating scalability and performance issues. While federated learning has promise, existing models are not adaptable and efficient in the most heterogeneous settings. Most study makes uniform device participation assumptions and do not consider the dynamic nature of IoT networks. A little investigation into attention-based models with the ability to discriminate between significant and redundant updates when aggregating the model, [11]. Additionally, prevalent hierarchical FL setups are either extremely simplistic or do not optimize on contextual attributes

like data authenticity, local domain relevance, and device stability. This deficiency motivates the establishment of an intelligent federated infrastructure that ensures data privacy and accommodates the network's underlying statistical and structural variability[12].

Although federated learning has potential, existing models are not adaptable and efficient in highly heterogeneous settings. Most study considers uniform device participation and does not consider the dynamic nature of IoT networks. Limited work has been done on attention-based models that can distinguish between significant and redundant updates during model aggregation [13]. Additionally, the current hierarchical FL architectures are too simplistic or do not tap into contextual insights like data consistency, task-level importance, and device credibility within the domain [14]. Hence, the federated framework not only addresses the privacy of data but also adjusts itself according to structural and statistical variability in the network. To overcome the above challenges, this study presents a new framework: Attention-Driven Hierarchical Federated Learning (AHFL). The new architecture leverages the advantages of attention mechanisms and hierarchical learning to support efficient and privacy-respecting edge AI over heterogeneous IoT networks. Through the experimental organization of devices into hierarchical groups based on resource abilities and data types, the model facilitates localized learning and then selective aggregation. Attention mechanisms are applied at several levels to detect and prioritize the most important model updates for global learning. This not only minimizes redundant communication but also allows the global model to leverage high-quality contributions, thus enhancing generalization performance.

The key contributions of the study are listed below:

- The study introduces a hybrid model that is a 1D-CNN and LSTM, where 1D-CNN model extracts the ECG data automatically. A new hierarchical FL framework designed for edge AI in heterogeneous IoT networks.
- It controls a large amount of ECG time-series data. Incorporation of attention mechanisms to weigh client inputs adaptively.
- Improved privacy protection without compromising model accuracy. Experimentation of the proposed model on different non-IID IoT scenarios shows better convergence and scalability.
- This study demonstrates that the hybrid 1DCNN-LSTM gives a good performance by using the wearable IoT device.

The rest of the section focuses on: Section II describes the related work of an attention edge AI in IoT networks. Section III reviews the problem statement. Section IV explains the methods used in this study. Section V examines the results, and Section VI details conclusion and future works.

II. Related Works

Raza et al. [15] proved that identifying arrhythmias using ECG and enhance federated learning while combined with explainable AI. The study keeps the data secure without accessing any dataset. The CNN and XAI methods are used in the classification and feature extraction process. Communication needs a huge amount, and “black box” detection is the challenge in the existing model. The dataset used in this study is MIT-BIH Arrhythmia Database. Furthermore, it received 94.5% accuracy using ECG data and 98.9% using ECG clean data.

Wang et al. [16] states the Human Activity Recognition (HAR) based on wearable sensor data, which is crucial for health monitoring, medical diagnosis, and motion analysis applications. The main objective of this study is to overcome the shortcomings of current deep learning-based HAR models, especially problems concerning incomplete and inefficient feature extraction that may result in erroneous activity recognition. To accomplish this, the authors introduce a new deep multi-feature extraction framework known as DMEFAM, which supports feature learning via attention mechanisms. In [17], the authors integrates two specialized layers: a Temporal Attention Feature Extraction Layer (TAFEL) that merges Bi-GRU and self-attention (SA) for sequential data, and a Channel and Spatial Attention Feature Extraction Layer (CSAFEL) that employs CBAM and ResNet-18 for spatial and channel-level feature highlighting. This framework enables the model to better distinguish and weight significant features within sensor data. A weakness catered by this is that previous models used low feature use rates, something the suggested attention-based framework resolves. High levels of recognition are demonstrated in results, at 97.9% on WISDM data, 96.0% on UCI-HAR, and 99.2% on a collected dataset, DAAD. These data sets were utilized to test the efficacy and usability of the proposed framework.

H. Zhang et al. [18] aims to enhance smart object recommendation in Internet of Things (IoT) and Social Internet of Things (SIoT) settings. The objective is to precisely choose the right smart objects from a huge collection based on a new deep learning model known as BLA (BERT and Bi-LSTM with Attention). The model combines BERT for semantic feature extraction and Bi-LSTM with self-attention and global-attention mechanisms for extracting contextual and relevance-based representations. Self-attention determines significant features without human intervention, whereas global-attention corresponds object data with user requirements. The model further uses thing–thing relationships for improved recommendation quality. Although not mentioned, problems faced could be computation requirements and responsiveness to diverse data. Findings indicate that BLA is superior to baseline models in terms of effectiveness.

This study targets Human Activity Recognition (HAR) with mobile sensor data like accelerometer and gyroscope signals. Akter et al. [19] primary aim is to improve HAR performance through a deep learning-based model with CNN. The method integrates features from several Convolutional stages and includes a Convolutional Block Attention Module (CBAM) to enhance feature refinement and extraction. Rather than relying on hand-crafted features, the model takes raw signal spectrograms as inputs, enabling automatic and effective high-level feature learning. The novelty is in the combination of multi-stage features and attention mechanisms to enhance model accuracy. While particular constraints aren't mentioned,

managing model complexity and generalization across datasets might be possible challenges. The datasets employed here are KU-HAR, UCI-HAR, and WISDM. The model resulted in good performance with 96.86% accuracies in KU-HAR, 93.48% in UCI-HAR, and 93.89% in WISDM. The outcomes portray better performance in comparison to past methods.

Dirgová Luptáková, Kubovčík, and Pospíchal [20] investigates Human Activity Recognition (HAR) based on smartphone sensors such as accelerometers and gyroscopes for healthcare, sports, and human–robot interaction applications. To enhance real-time activity classification by modifying the transformer model, which was initially created for NLP and vision tasks, to time-series analysis of motion signals. The approach takes advantage of the self-attention of the transformer to capture dependencies in the time series and provides a robust countermeasure against CNN and LSTM-based methods. This allows for better recognition of intricate activity patterns. This study assesses the implemented methods on the biggest publicly released smartphone motion sensor dataset, which is not specified by name. Some of the potential problems include computational complexity of transformers and the requirement of large training data. The proposed model attains a remarkable average activity recognition accuracy of 99.2%, which is well above traditional machine learning techniques with an accuracy of 89.67%. These findings prove the high potential of transformer models in HAR tasks[21].

This study focuses on HAR Industry such as IoT, e-health, and smart homes with 5.0 application. Al-Qaness et al. [22] aims to create a robust HAR system based on a new deep learning model known as Multi-ResAtt. This model utilizes multilevel residual networks and attention mechanisms to learn relevant features from inertial measurement unit (IMU) data. For efficient time-series analysis, the model include a recurrent neural network with attention which is tested on three public datasets, namely Opportunity, UniMiB-SHAR, and PAMAP2. Although there are no clear limitations discussed, real-time complexity and sensor heterogeneity could be potential issues. Multi-ResAtt outperformed current deep learning-based HAR approaches. It shows its strong performance in detecting complicated human actions.

III. Problem Statement

The continuous development of wearable IoT devices in ECG monitoring poses new chance of early detection of abnormal arrhythmias but also major emerging issues linked to data privacy, heterogeneity, and model accuracy. The traditional method of centralized learning schemes are non-private as it transmits raw data, and basic FL works such as FedAvg do not address the non-independent and identically distributed data or different qualities of data available to clients [23]. In the healthcare field, not all device data is equal for analysis as some are random, skewed, and less meaningful [24]. This results in distorting the models of the world and poor generalization. As such, this study presents an attention-driven hierarchical federated learning model that employs a CNN-LSTM model. In the proposed solution, the adversarial interference is mitigated by using attention mechanisms before LSTM for focusing on relevant ECG segments and during aggregation to guarantee clients' privacy while maintaining high accuracy and the framework's robustness for ECG classification in the different IoT environments.

IV. Methods for Privacy-Preserving Edge AI in Heterogeneous IoT Networks

The main objective of this study is based on the well-established federated learning method. It introduces an attention-driven hierarchical federated learning (HFL) mechanism, which is suitable for the St. Petersburg INCART 12-lead Arrhythmia Database to provide privacy-preserving arrhythmia classification. This has been achieved with the consideration of the main issues related to privacy, scalability, and heterogeneity experienced in the Internet of Things (IoT) healthcare systems. Based on the above design goals, a three-tier system architecture is adopted for the system, which includes edge devices, intermediate aggregators, and a central server. Each edge device uses a deep learning model of CNN-LSTM architecture for the ECG signal, where the CNN captures spatial patterns while the LSTM captures temporal patterns. An attention mechanism is also applied to pay more attention to the discriminative segments in signal for improving the interpretability and classification accuracy. This data never goes through the cloud and hence, does not violate any privacy policies. Information in local models is regularly transmitted to intermediate fog nodes, where initial data aggregation is done. This information is then transferred to a higher-tiered server which can be termed as a hierarchical processing model. The aggregation process also uses weights to decide the number of times it should draw data from various clients depending on their performance or credibility to handle the non-IID data common with IoT in real-world environments. In order to enhance the privacy of the health data, the framework can also use the method of differential privacy and secure aggregation. Finally, the performance of the proposed model based on accuracy, F1-score and communication efficiency metrics is being tested for worst case scenario, data heterogeneity and adversarial conditions for proving the robustness of the model. Overall, this also provides a way to have a real-time and intelligent ECG analysis that is privacy-preserving and scalable across various connected health devices. It includes low latency, flexibility, and convergence of the models with high security. The attention mechanism is also very useful in improving both the model training and also the federated communication. In sum, this study resolves general concerns that occur based on the integration of edge intelligence and medical data security. In summary, the practicality of the ECG Arrhythmia dataset is verified using this experimental set-up, to facilitate real-world applicability of the same. Fig. 1 demonstrates the working principle of attention federated learning for preserving data using hierarchical IoT device.

A. Data Collection

The St. Petersburg INCART 12-lead Arrhythmia Database is utilized for this study, which is obtained from the open-source Kaggle [25]. The information from this dataset contains various attributes. The dataset consists 75 annotated 30-minute ECG recordings collected from 32 subjects and 257 Hz samples from 12 leads. All the data are grouped as per the AAMI EC57 standard, and divided into five main classes: Normal, Ventricular, Supraventricular, Fusion, and Unknown. This

dataset is pre-processed using normalization, wavelet denoising, and segmentation.

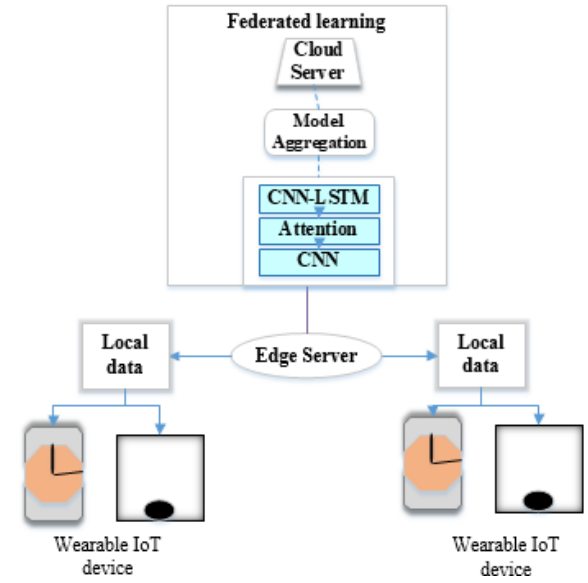

Fig. 1. Workflow of the attention federated learning edge AI using hierarchical IoT device.

B. Data Preprocessing

1) Signal denoising. ECG signals obtained from IoT devices are mostly contaminated by movement, electrical interferences, or shifting baseline noise. It helps to eliminate high-frequency components while storing the spatial extraction data for CNN model. Denoising is beneficial in enhancing signal quality to ensure an accurate classification. Eq. (1) represents the signal denoising formula:

$$\hat{x}\,(t) = x(t) - n(t) \tag{1}$$

In Eq. (1), $\hat{x}\,(t)$ denoised the signal, $x(t)$ notice a noisy signal, and $n(t)$ calculate the noise in the data. For this study, the wavelet denoising method is used for effective denoising. Denoising the wavelets breaks ECG signal into many frequencies sub-bands using the Discrete Wavelet Transform, enabling the removal of high-frequency noise without loss of some clinically relevant characteristics. It applies what is known as 'thresholding' on the wavelet coefficients and then reconstruct the signal with only large coefficients. And with this method, it is easy to separate the ECG signal from all the interfering noise and not distort the morphology of the waveform (top). The formula for wavelet denoising is given in Eq. (2). It is the formula of waveform:

$$\widehat{\omega} = sign(\omega)\,.\max(|\omega| - \lambda, 0 \tag{2}$$

where, ω is the original wavelet coefficient, $\widehat{\omega}$ is the denoised coefficient, and λ is a threshold value.

2) Normalization. Due to multiple sensors, the quality of IoT devices is not in a stable range. To overcome the values to the normal state, a z-score standardization data normalization technique is utilized. This technique is used to transfer data from the CNN layer to the LSTM layer in various time series. In federated learning (FL), normalization takes place to control the privacy and avoid central aggregation of data. It handles various features from different devices. The data focuses on standard deviation 1 and mean 0. The formula for the normalization technique given in Eq. (3). The equation is called normalization formula.

$$z = \frac{x - \mu}{\sigma} \tag{3}$$

where, σ is the standard deviation and μ is the mean.

C. Hybrid 1DCNN-LSTM Model in Healthcare Monitoring

The Long Short-Term Memory (LSTM) integrates with CNN to control the sequential and spatial data. In this study, IoT devices like wearable ECG patches or smartwatches collect motion signals, heart rate, or ECG data. CNN is utilized to extract spatial patterns from the device, and the output of the CNN is transferred to the LSTM model to analyze the sequential data. This technique enhances the identification of stress levels or arrhythmias.

The input layers receive ECG data in a time-series form. The wearable IoT device stores the raw signals and pre-processes the data in its device. The CNN model is used as a feature extractor that collects local patterns. The Convolutional layer extracts the spatial features using the following formula as in Eq. (4):

$$f_i = \sigma(\sum_{j=0}^{k-1} w_j\ .x_{i+j} + b) \tag{4}$$

where, K is the kernal, and σ is the ReLu. The extracted waveforms and increases in heart rate. In federated learning, the trained parameters are sent to the central server. It prevents users from transmitting data and preserves privacy in the edge server. The MaxPooling is used to lower the extracted ECG features. It selects the maximum value from the low-signal device and enhances the model performance by measuring the IoT device sensors in a limited waveform. The pooling is done by the following Eq. (5):

$$y_i = x_i + x_{i+1} +, \ldots\ldots, x_{i+p-1} \tag{5}$$

where, x_i is the input signal, y_i is the pooled output and p for the pooling size. It only allows the related patterns to attention layer and the LSTM layer for analyzing temporal patterns. The architecture of the hybrid 1DCNN-LSTM is given in Fig. 2.

1) LSTM layers. In this study, an attention mechanism only focuses on related time series or important features as input to improve the LSTM performance. It gives higher weights as input for prediction in the edge server. Arrhythmias are identified by a pattern of time, where LSTM defines the temporal information through its gate memory structure. There are three main categories in LSTM. They are: forget gate, input gate, and output gate. Here the cell gate is the memory unit.

a) Forget gate: The forget gate makes the decision which is to delete previous state. It deletes the irrelevant information from the data and only focuses on healthcare patterns. The formula for the forget gate is in Eq. (6):

$$f_t = \sigma\big(W_f\,.\,[h_{t-1}, x_t] + b_f\big) \tag{6}$$

where, σ is Sigmoid activation function, W_f represent the weight matrix for the forget gate, $h_{t-1}$is the before hidden state, x_t is current, and b_f refers to Bias term.

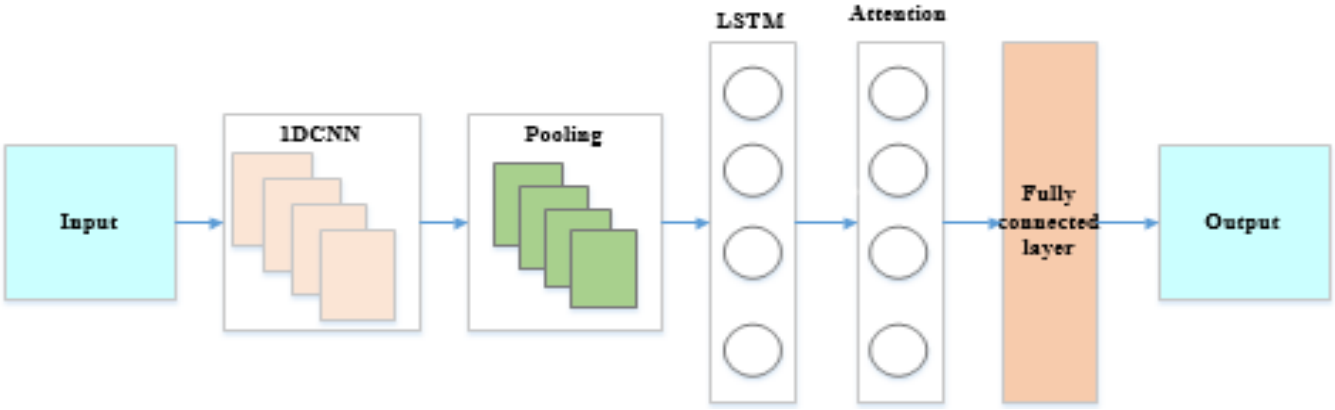

Fig. 2. Architecture representation of hybrid 1DCNN-LSTM.

b) Input gate: The input gate allows new information to be added to the cell gate. Thus, it can add relevant ECG information. The formula is in Eq. (7) and Eq. (8):

$$i_t = \sigma(W_i . [h_{t-1}, x_t] + b_i) \quad (7)$$

$$\widetilde{C_t} = \tanh(W_c . [h_{t-1}, x_t] + b_c) \quad (8)$$

where, W_i, W_c are the weight matrices for the input gate and candidate cell state and b_i, b_c are the bias terms.

c) Cell gate: This state updates the information by combining the current and previous states of cardiac rhythms. Eq. (9) represent the cell gate formula:

$$C_t = f_t * C_{t-1} + i_t * \widetilde{C_t} \quad (9)$$

d) Output gate: The output gate maintains the information from the cell state to the next layers and confirms that the data are related. Eq. (10) represents the output gate formula:

$$h_t = o_t * \tanh(C_t) \quad (10)$$

In Eq. (10), $*$ refers to element-wise multiplication and h_t shows the background changes of the ECG in a time series. The input of the fully connected layer is obtained from the LSTM layers output.

The attention mechanism enhances the understandability and model accuracy by selecting high-weight parameters in wearable IoT devices. The data stored in various edge servers is updated and sent to the global server, like a cloud service. The model is updated from information stored in multiple devices.

$$Attention(Q, K, V) = softmax(\frac{Qk^T}{\sqrt{d_k}}) \quad (11)$$

In Eq. (11), Q, K, and V are the inputs from CNN. The sequence is passed to the LSTM model. When the global model analyzes the updated information, it sends it back to the wearable IoT devices. The process is done recursively until the wearable device gets the updated information.

D. Federated Learning

In this study, federated learning is adopted to create an intelligent ECG arrangement classification model in distributed IoT environments that ensure the privacy of the recorded data from wearable sensors and mobile health monitoring devices. Instead of uploading actual ECG signal to the central server, this network just sends encrypted model update (EMU) to the aggregator. These updates are then merged in hierarchical FL approach to create a global model which gets advantages of all combined devices and thereby patient's data is secure and safe.

A three-tier hierarchical FL architecture is developed for real-time ECG arrhythmia classification based on wearable IoT devices. The structure takes into account of i) personal and wearable devices such as smartwatches or ECG patches, ii) fog nodes such as home or hospital gateways, and iii) a central server for the final fusion of models. Every client applies the locally trained model of CNN-LSTM with the help of the attention mechanism. Original ECG data remain stored locally, thus data privacy is well protected.

1) Attention-based federated averaging. In this study, which focuses on real-time ECG arrhythmia detection from wearable IoT devices, federated learning means that patients' sensitive health information does not leave the device. Instead, the devices engage in a distributed learning process by establishing and developing an individual deep learning model based on the ECG signals. The local training involves the use of these devices and their ability to send their own model updates to the fog server or even a central aggregator that helps in creating a more general model, Hazra et al. (2023). This setup also solves some of the basic issues with IoT healthcare systems, such as heterogeneity of collected data, communication overhead, security and privacy concerns.

FedAvg is also one of the most basic algorithms of FL. In particular, it updates aggregated parameters derived from multiple clients (for example, devices) to derive the new global model. In other situations, it goes through several rounds of communication till an efficient model is arrived at. FedAvg is

one of the well-established approaches to federated learning used to solve machine learning problems in decentralized edge devices without compromising the participants' privacy and with low communication complexity. As discussed earlier, FedAvg is used to update the average of the models iteratively of all the participating clients, for example, wearables. This method enables the server to update its global model easily without the need to download from the clients training data. The FedAvg algorithm starts with the central server having a global model that is to be sent out to all the selected edge devices. Each client then performs offloading to its own data and updates the model on that data for several epochs of gradient descent. The training formula for client-side server is given Eq. (12).

$$\omega_i^{(t+1)} = \omega_i^t - \eta \nabla \mathrm{Li}(\omega_i^t) \quad (12)$$

Here η indicate learning rate, Li is the loss function, and i represent the local data of the client. When the training is finished, it sends the updated information to the server. The weighted average is aggregated by the server. It assigns the weight of the client data based on the trustworthiness, relevance or performance. The formula for the aggregation is given in Eq. (13).

$$\omega^{(t+i)} = \sum_{i=1}^{k} \frac{n_i}{n} . \omega_i^{(t+1)} \quad (13)$$

where, $n_i$is the number of sample clients, n = $\sum n_i$ is the total number of all clients. It is a large dataset with global clients which leads the main process in federated learning. It processes efficiently and secures the data in many wearable IoT devices.

While at the aggregation stage, instead of averaging all updates as it is done in FedAvg, the system calculates the attention score for each client. These scores represent variants related to the only client factors like local model accuracy, loss decrease, the quality of data, or their compliance with previous changes. The scores are generally scaled using the method often known as softmax to force the values of the scores, having value closer to 1 and total up to 1, hence depicting the relative degree of significance or importance. Each client's model update in the communication is next weighted by the attention score given to it; thus, specific high performing or trustworthy clients essentially influence the new received global model. Last of all, it adds the weight into the global model updates collectively and then sends it back to the clients for the next round of training. This process repeats itself providing the opportunity to update this global model with respect to the most informative and reliable contributions, which makes the proposed method highly efficient for non-independent and identically distributed, heterogeneous, and privacy-preserving contexts such as real-time ECG classification in healthcare IoT applications.

Algorithm1. Federated Learning with 1DCNN-LSTM for Preserving Data in Hierarchical IoT Device

Input: Collects raw ECG data from wearable IoT device

Output: Predicted ECG class and trained global model

```
Global_Model ← initialize_CNN_LSTM_model()
Fog_Nodes ← [Fog_Node_1, Fog_Node_2, ..., Fog_Node_M]
Clients ← {Fog_Node_1: [Client_1, Client_2, ...], ..., Fog_Node_M: [Client_N, ...]}
Rounds ← Total_Training_Rounds
for round in range(1, Rounds+1):
  for fog_node in Fog_Nodes:
    Local_Updates ← []
    for client in Clients[fog_node]:
      ECG_Data ← client.collect_ECG_data()
      if ECG_Data is not None:
        ECG_Denoised ← denoise_signal(ECG_Data)
        ECG_Normalized ← normalize_signal(ECG_Denoised)
        ECG_Segmented ← segment_signal(ECG_Normalized)
        Local_Model ← clone(Global_Model)
        Local_Model.train(ECG_Segmented)
        accuracy ← Local_Model.evaluate_accuracy()
        loss ← Local_Model.evaluate_loss()
        if accuracy ≥ Threshold_Accuracy and loss ≤ Threshold_Loss:
          attention_weight ← compute_attention_weight(accuracy, loss)
        else:
          attention_weight ← assign_low_weight()
        Local_Updates.append((Local_Model.parameters, attention_weight))
      else:
        log("Client has no ECG data.")
    if Local_Updates is not empty:
      Fog_Model ← aggregate_models_with_attention(Local_Updates)
      send_to_central_server(fog_node, Fog_Model)
    else:
      log("No updates from clients in this fog node.")
  All_Fog_Models ← collect_models_from_fog_nodes()
  if All_Fog_Models is not empty:
    Global_Model ← aggregate_models_with_attention(All_Fog_Models)
  else:
    log("No models received from fog nodes.")
  for fog_node in Fog_Nodes:
    broadcast_model_to_clients(Global_Model, fog_node)
Trained_Global_Model ← Global_Model
ECG_Classification_Results ← Trained_Global_Model.predict(new_ECG_data)
```

Algorithm 1 shows the federated learning with 1DCNN-LSTM for preserving data in hierarchical IoT device. Fig. 3 shows the workflow of the attention mechanism in a federated learning IoT device. The proposed flowchart allows for implementing an attention into the structure of the hierarchical federated learning process to improve model aggregation. Every edge device runs CNN-LSTM on segmented ECG signals, and assesses the outcome in terms of local validation indices. These scores are utilized in the fog-level and global aggregation to perform an update of the model. This mechanism replaces uniform averaging (FedAvg), which enhances adaptability,

performance-based averaging, and results in better non-IID and noisy ECG settings. The novelty is based on incorporating the layer-level attention for the feature relevancy with the network-level attention for the aggregation reliability which in turn assures better convergence, interpretability as well as the privacy-specific performance across the heterogeneous edge nodes.

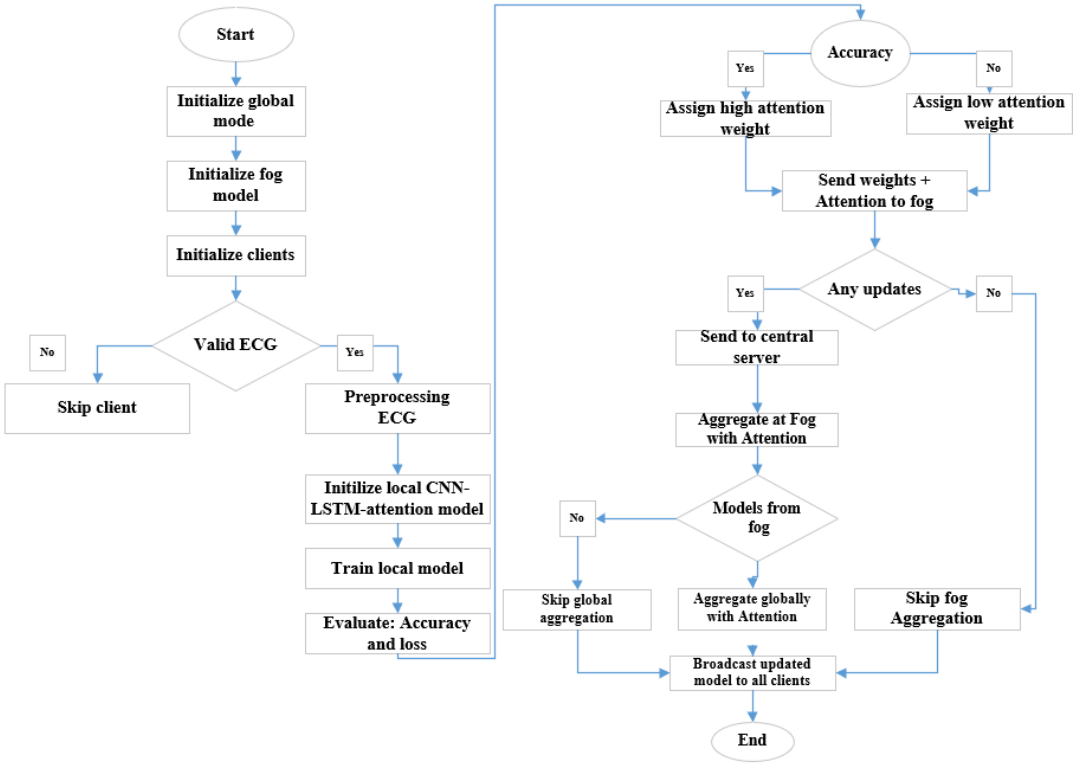

Fig. 3. Flowchart representation of Attention-driven Federated Learning.

V. Result and Discussion

This section provides the assessment and discussion of the approach formulated in the context of ECG arrhythmia classification based on the approach to the attention-driven hierarchical federated learning. Relative measures of performance are typically used for evaluating the model, including correct rate, precision, recall, F1 measure, and ROC-AUC. The redundancy reduction pattern of the local and global models is also established while comparing both models to prove that a hierarchical aggregation enhances the model. The next step aims to compare the convergence speed and the communication efficiency of the proposed model to the standard FedAvg. Till now, confusion matrices and attention heatmaps are helpful in the interpretation of the model's diagnostic emphasis. Furthermore, compare the model with centralized and other models with no attention mechanism included. Finally, an ablation study shows the significance of each part invented in the present work is implemented in the window platform in python CPU/ memory 8GB.

Attention-Driven Hierarchical Federated Learning (HFL) for edge AI privacy in heterogeneous IoT networks is faced with multiple challenges. They include system and data heterogeneity, as devices vary in computing capability and produce non-IID data, causing model convergence to be slow or unstable. The attention mechanism, as it enhances learning focus, adds computational burden—challenging resource-limited devices. Overhead of communication, device dropouts, and privacy-performance trade-offs add to the complexity of training. These problems may lead to unfair models, inefficiency, and decreased scalability. Adaptive and lightweight attention architectures, resilient client selection, model compression, safe aggregation, and privacy-preserving techniques such as differential privacy can be employed to enable consistent and efficient learning in various IoT settings.

A. Model Performance

Model performance in this study concerns the capability of the proposed CNN-LSTM-attention to the accurate and efficient classification of ECG signals into different types of arrhythmia. The 1D-CNN extracts spatial features from the segmented ECG data as the next step, followed by an attention layer which aims at identifying clinically relevant regions from the obtained features. The LSTM in turn, feeds these features in order to capture temporal dependencies and rhythm variations over time. The last layer is composed of output nodes that determine, where the sequence belongs to the arrhythmia classes specified below. There are relevant indicators such as accuracy, zero one loss, precision, recall, F1-score, and control under the receiver operator (AUC) curves, which describe the model's performance in practical large-scale distributed healthcare systems. The figure shows the local model and global model performance. The accuracy of global model is higher than the local in all evaluation metrics.

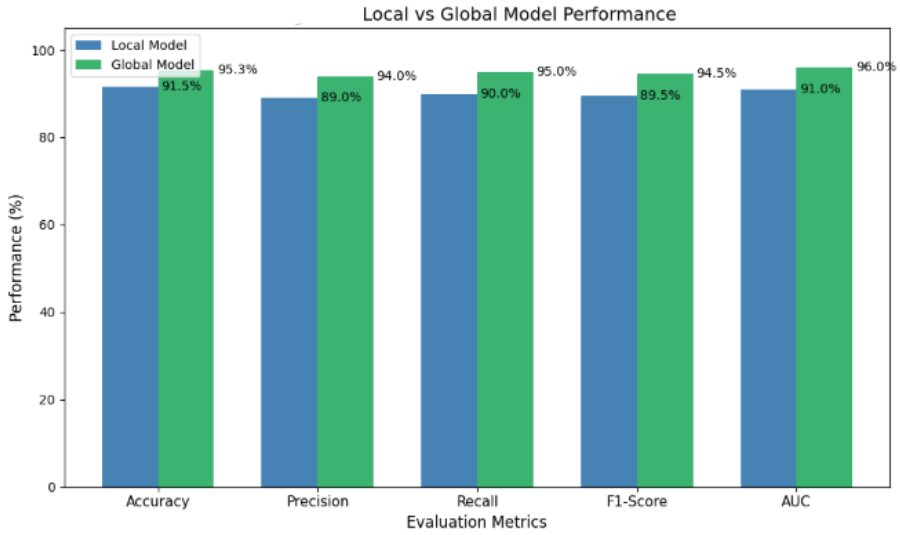

Fig. 4. The performance of local and global models.

Table I shows a summary of the proposed federated learning framework results and Fig. 4 represents the performance metrics of a local and global model. By evaluating the results, it concludes that the global model is better than the local models in terms of all the evaluation measures, namely accuracy, precision, recall, F1-score, and AUC. Even more important, the global model obtained 95.3% of accuracy and 94.5% of F1-score, while the local model obtained just 91.5% for accuracy and 89.5% for F1-score. The results presented in this study shows that not only privacy but also it achieves better predictive performance when the proposed federated framework is used in real-time ECG arrhythmia classification.

B. Multiclass Performance

In this study, the classification of each ECG heartbeat, a total of five classes are recognized: [N], [S], [V], [F], and [Q]. These classes are obtained from multiple annotated beat types of the ECG, thereby transforming them into unified diagnostic classes in order to get more clinical relevance. This classification framework helps in generalizing the outcome of the deep learning model across a variety of arrhythmic as well as, non-arrhythmic ECG signals. The last layer of the CNN-LSTM model proposes the probability distribution of the output layer, and the class with the highest probability is considered to be the prediction result. The multiclass labels are shown in Table II.

C. Confusion Matrix

Fig. 5 represent the predicted values for the multiclass label in healthcare monitoring. In this study, a confusion matrix is computed to predict the performance of the CNN-LSTM-Attention model to classify the ECG beats into two groups of rhythms, namely Supraventricular and Ventricular, and three groups of rhythms, namely Normal, Supraventricular, and Ventricular. The diagonal elements inside the matrix are a correct predicted value of the instances, the row as the actual class and the column as the predicted class. It provides results in terms of the performance for each class and the classification errors that occurred during diagnosis, which will give an insight into the reliability of the diagnostics as well as the areas that require improvement in the model. In Fig. 5, the Confusion Matrix for ECG Arrhythmia Classification of actual class and predicted class, the normal rhythm is 2.00 and the ventricular is 1.00.

D. Multiclass ROC-AUC Curve

Fig. 6 represents the ECG classification in ROC-AUC using multiclass labels. The ROC-AUC graph for the model represents the evaluation of the implementation of the model to distinguish between each class of ECG arrhythmia using a one class to the rest strategy. The graphs are plotted for every class, the curve pointing True Positive Rate, which points out the sensitivity of the screen. The value of AUC near to 1 shows better discriminative ability of the classifier for that class. The accuracy is the highest when the classes have line graphs near the top-left intersection.

TABLE I. PERFORMANCE OF FEDERATED LEARNING EVALUATION

Metric	Local Model (%)	Global Model (%)
Accuracy	91.5	95.3
Precision	89.0	94.0
Recall	90.0	95.0
F1-Score	89.5	94.5
ROC-AUC	91.0	96.0

TABLE II. MULTICLASS CLASSIFICATION PERFORMANCE

Class Label	Class Code
Normal	N
Supraventricular	S
Ventricular	V
Fusion	F
Unknown	Q

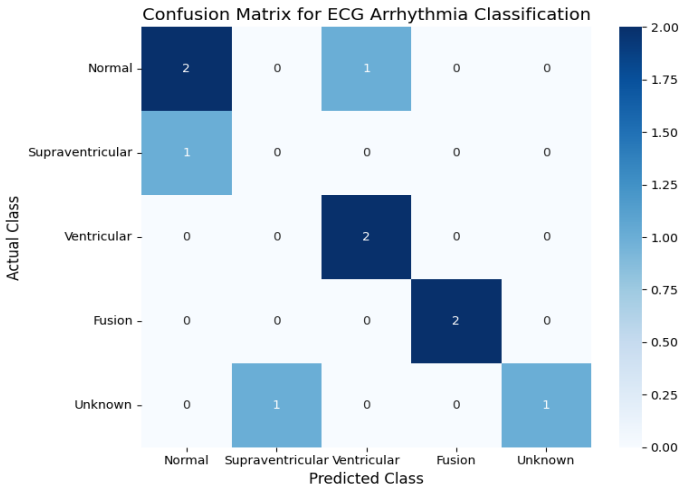

Fig. 5. Confusion matrix for predicted label and true label.

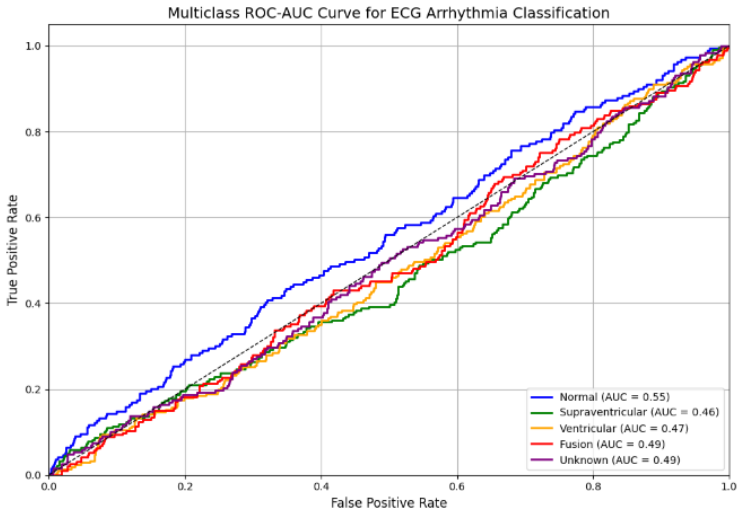

Fig. 6. Multiclass ROC-AUC curve for ECG arrhythmia classification.

E. Comparison of Existing Models

Comparing the models from CNN to the CNN-LSTM-Attention model, the results have been improved in the model's classification capability. In other words, the CNN module successfully captures spatial details of an image, whereas the LSTM encodes temporal aspects of the sequence. Thus, integration of these architectures in the CNN LSTM model helps to integrate spatial and temporal information and increases the accuracy of estimations. The CNN-LSTM-Attention model trained with attention pays more attention to such crucial segments as a result of the proposed architecture; hence, it posted the best performance among the recognized models. These developments support the effectiveness of using attention mechanisms as well as the incorporation of hybrid architectures for reliable and accurate ECG arrhythmia classification. Table III shows the comparison of existing models and proposed model. httpp://www.kaggle.com/code/ahmedashrafhelmi/ecg-classification-rnn-gru-lstm/input. Fig. 7 shows the performance comparison of existing models.

TABLE III. COMPARISON OF THE PROPOSED MODEL AND EXISTING MODEL

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
CNN	97.44	97.56	97.44	97.50
LSTM	97.11	97.40	97.11	97.25
CNN-LSTM	98.22	98.26	98.23	98.24
CNN-LSTM-Attention	98.91	98.98	99.03	99.00
Existing model	97.87	94.20	89.76	87.97

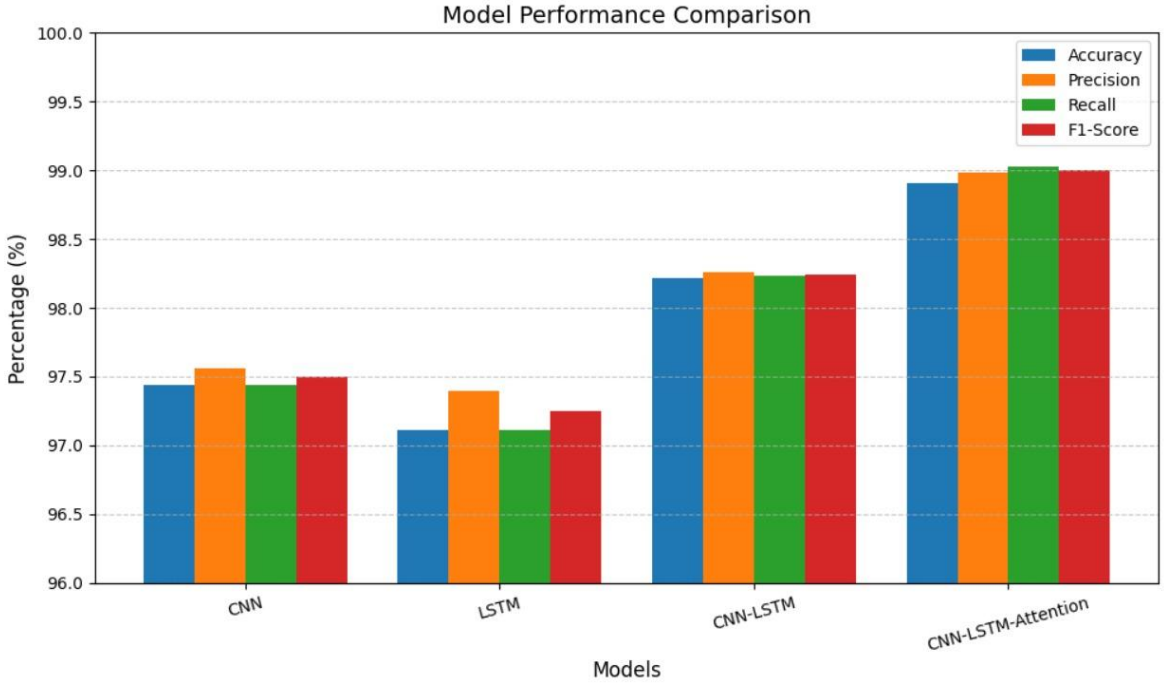

Fig. 7. Performance comparison of existing models.

F. Ablation Study

Fig. 8 shows the overall performance of a model which affects the study. This ablation study shows that each of the components in the proposed architecture is useful in a way and indispensable. The removal of the attention layer resulted in a decrease of model accuracy, reinforcing its function of directing attention to necessary ECG features. Additionally, excluding the LSTM degraded temporal learning, the effect of removing the fog layer on convergence efficiency is not conducive. The last model gathered the highest accuracy, which confirms that the use of CNN with LSTM, attention mechanism, and the hierarchical structure of the model produces the highest results for the classification of arrhythmia.

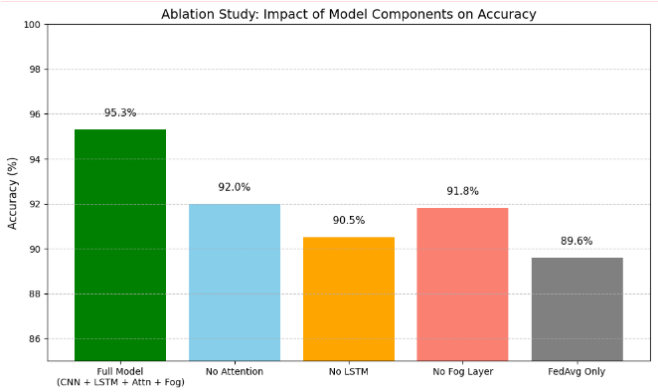

Fig. 8. Performance of each model.

G. Discussion

The proposed attention-driven hierarchical federated learning technique is a great improvement in the classification of ECG arrhythmia as it employs CNN, LSTM, and attention mechanisms. This integration helps the model to detect all spatial and temporal complex determinations of ECG signals to recognize the challenging three types of arrhythmic patterns. The attention mechanism enhances this procedure by focusing on certain parts of the ECG at sharp and boosting the interpretability and classification. This is because federated learning can allow training to be conducted on the device, where data is collected, without sharing that data with other parties to the federation, thus ensuring patient information is protected. The model further validates the practicality of the same for real time monitoring in low-resource settings. There are directions for future work: to achieve better efficiency for the model in various end devices, to consider the adaptive FL approaches to address the data heterogeneity, and to use the Explainable AI in enhancing the clinician's trust in the AI system. Also, applying the described study framework to promptly detect and monitor arrhythmias in wearable devices can greatly improve the prevention of heart conditions.

Although the evidence in favor of the aforementioned attention-driven hierarchical federated learning approach is encouraging, some limitations need to be noted. One of the main issues is the scarcity of computational and energy resources that are often present on most IoT edge devices, which can cause difficulties in processing sophisticated deep learning models such as CNN-LSTM with attention mechanisms. Also, heterogeneous and non-IID data at different clients may cause model convergence problems and less generalization. Communication overhead is another major challenge since federated learning needs to exchange model updates frequently, leading to potential latency and bandwidth usage, particularly in big networks. Additionally, while federated learning maintains privacy with data staying local, it remains susceptible to attacks like gradient leakage and model inversion. Lastly, while attention mechanisms enhance interpretability to some degree, it is still possible for the model to be viewed as a black box,

potentially limiting trust and adoption in sensitive areas such as healthcare.

In order to overcome these challenges, future research should aim to optimize models for deployment on the edge using methods such as model compression, pruning, and quantization. Mitigating data heterogeneity can include applying personalized or cluster-based federated learning techniques that suit better mixed data distributions. Efficiency in communication can be enhanced by means of model update compression, asynchronous training, or selective update methods. Further, augmenting privacy will involve combining sophisticated security mechanisms like differential privacy and secure multi-party computation. Lastly, to enhance transparency and trustworthiness, explainable AI techniques must be incorporated into the structure so that attention-based decision-making can be clearly visualized and end-users such as clinicians or IoT operators can be confident.

VI. Conclusion and Future Work

In this study, a proposed and investigated attention-driven hierarchical federated learning environment for learning ECG arrhythmia classification using the CNN-LSTM-Attention model. The model can incorporate spatial as well as temporal information of the ECG signal to make an accurate classification across various types of arrhythmias. Through integrating with the attention mechanisms, it enables the model to pay attention to some areas of the ECG, which improves its interpretability and accuracy. Federated learning caters to data privacy to prevent raw data from being fed to other clients, instead letting clients train their devices.

The plans for further development include addressing the problems associated with model implementation in constrained devices, further studying of methods of federated learning for handling with the data heterogeneity, as well as integration of the explainability tools to improve the doctors' trust in a model. It is also possible to extend the definition of the framework to utilize wearable technology for the real-time detection of pre-arrhythmias and thus promote the enhancement of preventive cardiologic services. This would require constant testing and validation on other datasets to be able to determine the general applicability of the proposed approach.

References

[1] O. Çalışkan, N. P. Temizel, M. Akay, and B. Mashhoodi, "Typological diversity and morphological continuity in the modern residential fabric: The case of Ankara, Turkey," Habitat Int., vol. 142, p. 102950, 2023.

[2] L. Tinella et al., "Fostering an age-friendly sustainable transport system: A psychological perspective," Sustainability, vol. 15, no. 18, p. 13972, 2023.

[3] R. Verma, "Smart city healthcare cyber physical system: characteristics, technologies and challenges," Wirel. Pers. Commun., vol. 122, no. 2, pp. 1413–1433, 2022.

[4] K. P. Reddy et al., "Association between delayed/forgone medical care and resource utilization among women with breast cancer in the United States," Ann. Surg. Oncol., vol. 32, no. 4, pp. 2534–2544, 2025.

[5] A. Ali, Y. Zhu, and M. Zakarya, "A data aggregation based approach to exploit dynamic spatio-temporal correlations for citywide crowd flows prediction in fog computing," Multimed. Tools Appl., vol. 80, no. 20, pp. 31401–31433, 2021.

[6] A. Singh, S. C. Satapathy, A. Roy, and A. Gutub, "Ai-based mobile edge computing for iot: Applications, challenges, and future scope," Arab. J. Sci. Eng., vol. 47, no. 8, pp. 9801–9831, 2022.

[7] T. Zhang, L. Gao, C. He, M. Zhang, B. Krishnamachari, and A. S. Avestimehr, "Federated learning for the internet of things: Applications, challenges, and opportunities," IEEE Internet Things Mag., vol. 5, no. 1, pp. 24–29, 2022.

[8] J. Zhang et al., "Adaptive federated learning on non-iid data with resource constraint," IEEE Trans. Comput., vol. 71, no. 7, pp. 1655–1667, 2021.

[9] P. Biedermann et al., "Standardizing registry data to the OMOP Common Data Model: experience from three pulmonary hypertension databases," BMC Med. Res. Methodol., vol. 21, pp. 1–16, 2021.

[10] V.-D. Nguyen, S. K. Sharma, T. X. Vu, S. Chatzinotas, and B. Ottersten, "Efficient federated learning algorithm for resource allocation in wireless IoT networks," IEEE Internet Things J., vol. 8, no. 5, pp. 3394–3409, 2020.

[11] H. Touvron et al., "Augmenting convolutional networks with attention-based aggregation," ArXiv Prepr. ArXiv211213692, 2021.

[12] K. Ashok and S. Gopikrishnan, "Statistical analysis of remote health monitoring based IoT security models & deployments from a pragmatic perspective," IEEE Access, vol. 11, pp. 2621–2651, 2023.

[13] K. Ashok and S. Gopikrishnan, "Statistical analysis of remote health monitoring based IoT security models & deployments from a pragmatic perspective," IEEE Access, vol. 11, pp. 2621–2651, 2023.

[14] J. Davis et al., "Methodology and evaluation in sports analytics: challenges, approaches, and lessons learned," Mach. Learn., vol. 113, no. 9, pp. 6977–7010, 2024.

[15] A. Raza, K. P. Tran, L. Koehl, and S. Li, "Designing ECG monitoring healthcare system with federated transfer learning and explainable AI," Knowl.-Based Syst., vol. 236, p. 107763, 2022.

[16] Y. Wang et al., "A novel deep multifeature extraction framework based on attention mechanism using wearable sensor data for human activity recognition," IEEE Sens. J., vol. 23, no. 7, pp. 7188–7198, 2023.

[17] N. A. Chandramouli et al., "Enhanced human activity recognition in medical emergencies using a hybrid deep CNN and bi-directional LSTM model with wearable sensors," Sci. Rep., vol. 14, no. 1, p. 30979, 2024.

[18] J. Zhang et al., "Adaptive federated learning on non-iid data with resource constraint," IEEE Trans. Comput., vol. 71, no. 7, pp. 1655–1667, 2021.

[19] M. Akter, S. Ansary, M. A.-M. Khan, and D. Kim, "Human activity recognition using attention-mechanism-based deep learning feature combination," Sensors, vol. 23, no. 12, p. 5715, 2023.

[20] I. Dirgová Luptáková, M. Kubovčík, and J. Pospíchal, "Wearable sensor-based human activity recognition with transformer model," Sensors, vol. 22, no. 5, p. 1911, 2022.

[21] C. Han, T. Yang, X. Sun, and Z. Cui, "Secure Hierarchical Federated Learning for Large-Scale AI Models: Poisoning Attack Defense and Privacy Preservation in AIoT," Electronics, vol. 14, no. 8, p. 1611, 2025.

[22] M. A. Al-Qaness, A. Dahou, M. Abd Elaziz, and A. Helmi, "Multi-ResAtt: Multilevel residual network with attention for human activity recognition using wearable sensors," IEEE Trans. Ind. Inform., vol. 19, no. 1, pp. 144–152, 2022.

[23] T. Qi, F. Wu, C. Wu, Y. Huang, and X. Xie, "Differentially private knowledge transfer for federated learning. preprint," Rev. Httpsdoi Org1021203rs, vol. 3, 2022.

[24] D. Kumar et al., "Cardiac diagnostic feature and demographic identification (CDF-DI): an IoT enabled healthcare framework using machine learning," Sensors, vol. 21, no. 19, p. 6584, 2021.

[25] S. Sakib, "ECG Arrhythmia Classification Dataset." Accessed: Apr. 15, 2025. [Online]. Available: https://www.kaggle.com/datasets/sadmansakib7/ecg-arrhythmia-classification-dataset

[26] A. Hazra, P. Rana, M. Adhikari, and T. Amgoth, "Fog computing for next-generation internet of things: fundamental, state-of-the-art and research challenges," Comput. Sci. Rev., vol. 48, p. 100549, 2023.