Computer Methods and Programs in Biomedicine 242 (2023) 107821

Contents lists available at ScienceDirect

Computer Methods and Programs in Biomedicine

journal homepage: www.elsevier.com/locate/cmpb

# Weighted aggregation through probability based ranking: An optimized federated learning architecture to classify respiratory diseases

Abdullah Abdul Sattar Shaikh [a], M.S. Bhargavi [a], Pavan Kumar C [b,*]

[a] Department of Computer Science and Engineering, Bangalore Institute of Technology, Bangalore, 560004, Karnataka, India
[b] Department of Computer Science and Engineering, Indian Institute of Information Technology Dharwad, Dharwad, 580009, Karnataka, India

ARTICLE INFO

*MSC:*
68T05
62P10

*Keywords:*
Respiratory disease classification
Federated learning
GANs

ABSTRACT

*Background and Objective*
Respiratory Diseases are one of the leading chronic illnesses in the world according to the reports by World Health Organization. Diagnosing these respiratory diseases is done through auscultation where a medical professional listens to sounds of air in the lungs for anomalies through a stethoscope. This method necessitates extensive experience and can also be misinterpreted by the medical professional. To address this issue, we introduce an AI-based solution that listens to the lung sounds and classifies the respiratory disease detected. Since the research work deals with medical data that is tightly under wraps due to privacy concerns in the medical field, we introduce a Deep learning solution to classify the diseases and a custom Federated learning (FL) approach to further improve the accuracy of the deep learning model and simultaneously maintain data privacy. Federated Learning architecture maintains data privacy and facilitates a distributed learning system for medical infrastructures.
*Methods*
The approach utilizes Generative Adversarial Networks (GAN) based Federated learning approach to ensure data privacy. Generative Adversarial Networks generate new data by synthesizing new lung sounds. This new synthesized data is then converted to spectrograms and trained on a neural network to classify four lung diseases, Heart Attack and Normal breathing patterns. Furthermore, to address performance loss during FL, we also propose a new "Weighted Aggregation through Probability-based Ranking (FedWAPR)" algorithm for optimizing the FL aggregation process. The FedWAPR aggregation takes inspiration from exponential distribution function and ranks better performing clients according to it.
*Results and Conclusion*
A test accuracy of about 92% was achieved by the trained model while classifying various respiratory diseases and heart failure. Additionally, we developed a novel FedWAPR approach that significantly outperformed the FedAVG approach for the FL aggregate function. A patient can be checked for respiratory diseases using this improved learning approach without the need for extensive sensitive data recording or for making sure the data sample obtained is secure. In a decentralized training runtime, the trained model successfully classifies various respiratory diseases and heart failure using lung sounds with a test accuracy on par with a centralized model.

## 1. Introduction

Respiratory Diseases are the leading chronic illness in the world. The World Health Organization (WHO) has reported that about 417,918 people have died globally per annum due to respiratory diseases [1,2]. Furthermore, 24.8 million various medical disabilities were also traced back to asthma in 2016. In 2017, WHO declared chronic respiratory diseases accounted for more than 10% of the globally caused diseases, second to cardiovascular diseases [3].

Diagnosing a Respiratory disease is done through "auscultation" which facilitates listening to internal sounds of air moving inside and outside the lungs using a stethoscope [4]. A patient is diagnosed on the basis of atypical breathing sounds from the lungs by a medical practitioner. Some common abnormal/atypical lung sounds that are

* Corresponding author.
*E-mail addresses:* abdullahshaikh136@gmail.com (A. Abdul Sattar Shaikh), ms.bhargavi@gmail.com (M.S. Bhargavi), pavan@iiitdwd.ac.in (P. Kumar C).

https://doi.org/10.1016/j.cmpb.2023.107821
Received 8 December 2022; Received in revised form 29 August 2023; Accepted 15 September 2023
Available online 21 September 2023
0169-2607/

commonly identified are wheezes, pleural rubs, stridors, and coarse or fine crackles. Each of these lung sounds has distinct characteristics, for example, wheezes and stridors are continuous high-pitched sounds occurring at a frequency range between 400 and 500 Hz, respectively. Lower-pitched wheeze sounds are also known as rhonchi. High-pitched wheeze sounds may be present as a result of inflamed or narrowed bronchial tubules and are thus an indication of asthma or chronic obstructive pulmonary disease [5]. Stridors usually occur due to tracheal or laryngeal edema [6]. On the other hand, crackles are discontinuous high-pitched (fine) or low-pitched (coarse) waves associated with pneumonia, bronchitis, or heart failure conditions [7]. Similarly, pleural rubs are low-pitched rhythmic sounds associated with inflamed lung lining due to pleural effusion.

Since every respiratory disease has specific characteristics corresponding to these atypical lung sounds, a patient can be easily and safely diagnosed by only using a stethoscope. However, although it is regularly used in healthcare, standard pulmonary auscultation has some notable limitations. Firstly, acquiring good auscultatory skills requires extensive training and expertise. Additionally, efficient detection of atypical sounds is sensitive to the level of experience and auditory acuity of the healthcare professional. Even if auscultation is performed by an expert practitioner, abnormal patterns are sometimes overlooked or misinterpreted during the examination [8]. Thus, variability in observations and interpretations may limit the diagnostic effectiveness of such an approach. These challenges have given rise to the significance of computer-aided auscultation systems that can perform automated identification of atypical lung sounds and Respiratory Diseases.

Many researchers have provided AI-based solutions to this problem which are explored in the related work section. The adopted traditional machine learning and deep learning approaches in these works where sensitive medical data is aggregated for training and evaluation can cause a huge privacy issue regarding the personal information of the medical patient. Moreover, the data provided by medical facilities are tightly regulated, and gaining access to them is difficult. Federated Learning (FL) overcomes this issue by training a Machine Learning (ML) model in a decentralized manner [9,10]. In federated learning, instead of the data, the model weights are shared and aggregated while training. This lack of movement of data ensures that the privacy of sensitive medical data is maintained. From our non-exhaustive search, we observed a lack of research in Federated Learning towards Lung sound classification and its optimization since the dataset consisted of imbalance which would adversely affect the performance of the FL models.

In this research work, we address this issue by procedurally generating our own Lung sound spectrograms [11] by employing Generative Adversarial Networks (GAN) [12]. On the surface, GANs consist of two Neural Networks, a generator which, as its name suggests creates a sample dataset from noise, and a “discriminator” determines whether the sample data created by the generator is real or synthetic. These two Neural networks face off with each other in constant back and forth where they try to fool each other and evolve. Eventually, this results in the generator creating data samples from random noise which are almost similar to the real sample although not an exact replica of the dataset. Once the data created by the generator are of adequate quality, we can create Lung Sounds spectrograms of people who “don’t exist”. This avoids the involvement of real-world data during training by generating a large amount of quality data through few real samples thus maintaining the privacy of medical records.

After this generation of data, we subject it to rigorous experimentation using FL architecture. In our testing, we observed that the traditional FedAVG approach to FL had a significant drop in accuracy [13]. This was due to the loss of information while aggregating the weights during training. To rectify and optimize this FL aggregating system, we introduce Weighted Aggregation through Probability-based Ranking (FedWAPR) algorithm. The FedWAPR method gives more priority to the clients having higher accuracy and allots weights to the clients on the basis of probability distribution functions like exponential distribution [14] and Log Cauchy Distribution [15]. Consequently, this approach also increases the converging speed of the FL training process, which in FL training is a major bottleneck.

This custom aggregation method was able to achieve accuracies on par with traditional centralized supervised learning systems and in fewer rounds of aggregation as compared to FedAVG aggregation. The model also optimized the computational and time demand on the system and stood on equal metrics on FL-based training that had more epochs and heavier weights.

## 2. Related works and contributions

Numerous efforts have been made by researchers to explore AI - driven approaches for the classification of lung sounds. Rigorous ML-based testing is done on the JUT datasets created by Fraiwan et al. [16]. Their work mainly concentrates on ensemble-based classifiers like Decision Trees and Linear Discriminants that are compared with traditional ML approaches like Support Vector Machine (SVM) and $k$-Nearest Neighbors ($k$NNs). This work resulted in the conclusion that boosted decision tree models exhibit the best performance with the highest sensitivity at 91.5%.

The work by Khan et al. [17] proposes the implementation of Mel-frequency cepstral coefficient (MFCC) and gamma-tone cepstral coefficient (GFCC) pre-processing techniques before subjecting it to a bagged tree classifier. This work also implemented the model on a Raspberry Pi and subjected it to a 5-fold cross-validation.

Another recent work by Garcia et al. [18] employed variational autoencoders to balance the imbalance classes in the dataset. They employed Mel Spectrogram [11] and CNN’s on cropped and fixed-length audio data acquired from the ICBHI dataset [19].

A Gated Recurrent Unit (GRU) implementation was also implemented by Basu and Rana [20] where the authors subjected the ICBHI dataset to an optimized Convolutional Neural network (CNN) after pre-processing it for MFCC-based features. This work also mentions the requirement of further testing of the model using GANs.

Another Spectrogram-based feature extraction method was carried out by Tariq, Sha and Lee [21] on the ICBHI dataset. The work mainly emphasizes the rigorous study and effects of various normalization techniques on the dataset before subjecting it to training on a custom CNN-based Lung Disease Classification system.

Related work on FL by Dou et al. [22] on classifying COVID-19 which is also a lung-based disease using CT scans has demonstrated that FL tends to better generalize a model and provide better results. In their study, they claimed the requirement for better quality and quantity of data plays a crucial role in the performance of the final model that is created.

Another shortcoming of FL is the skewed distribution of data and not independent and identically distributed (non-IID) among the clients. This issue was studied by Feki et al. [23] while working on the COVID 19 classifier system and concluded that this heterogeneity of data does not affect the model’s performance.

Another novel approach by Tong et al. [24] demonstrated the modeling of small pressure curves which simulated normal and wheezing breathing patterns in the lungs. The authors presented a feature-band attention module that provided promising results for a sound-based lung disease classifier.

Cetinkaya, Akin and Sagiroglu [25] suggested weight pruning and quantization that reduce the communication cost by 10 times. This FL-based CT image classifier classifies different Respiratory diseases like COVID-19, viral or bacterial pneumonia, and lung opacity.

One of the latest studies by Fraiwan et al. [26] used a deep learning model based on CNNs and Bidirectional Long Short-Term Memory (BiLSTM) for classifying lung sounds. With a Cohen’s kappa value of 98.26%, this model had an overall average accuracy of 99.62% when classifying lung sounds into different respiratory conditions.

Smartphone-based breathing sounds are suggested as a promising signal for COVID-19 instances in yet another unique application study by Alkhodari and Khandoker [27]. It also suggests using deep learning as a pre-screening technique for such cases before performing the industry-recognized RT-PCR tests. This model achieved a total accuracy of 94.58% in distinguishing between COVID-19 and healthy patients which demonstrates the approach's potential. This work prepares the road for the adoption of deep learning in COVID-19 diagnostics by outlining it as a quick, time-saving, and cost-free method that complies with social distance regulations during pandemics like COVID-19 but does not violate them.

In their study, Saldanha et al. [28] examined how different classifiers performed when the imbalanced ICBHI dataset was supplemented with artificial audio segments produced by Variational Auto-Encoders(VAE). Utilizing metrics like Fréchet Audio Distance, Cross-Correlation, and Mel Cepstral Distortion, the similarity between the features of the generated audio segments and those of the original were determined. The findings demonstrated that VAE and other unconditional generative models outperformed Conditional VAE in terms of improving the classifiers' performance metrics but when trained on an unbalanced training set, classifiers entirely misclassified some disorders.

Tasar, Yaman and Tuncer [29] developed a high-accuracy sound categorization model based on a nonlinear histogram-based generator. Piccolo pattern, statistical moments, Tunable Q-factor Wavelet Transform (TQWT), Iterative Neighborhood Component Analysis (INCA), and traditional classifiers are used in tandem to achieve this goal. TQWT is used to generate levels in this approach.

Haider and Behera [30] used a computerized Lung Sound (LS) - based approach for classifying asthma and COPD cases. They classified healthy, COPD, and asthma cases utilizing Wavelet entropy and wavelet packet energy properties of LS, several classifiers including a support vector machine, decision tree, k-nearest neighbor, and discriminant analysis (DA). Using the decision tree classifier, the suggested technique has a high classification accuracy of 99.3%.

Kochetov and Filchenkov [31] have explored Lung sound prediction using GANs over the ICBHI dataset. Jayalakshmy and Sudha [32] proposed a method based on Conditional GANs to achieve a balanced dataset for deep learning models. Malygina, Ericheva and Drokin [33] showed that data augmentation with GANs improved the accuracy of pneumonia binary classification tasks, even when the generative network was trained on the same dataset.

Hsu et al. [34] used pre-trained ResNet models as backbone architectures for the classification of adventitious lung sounds and respiratory diseases. They employed techniques such as fine-tuning, co-tuning, stochastic normalization, and data augmentation to address class imbalance and variations in recording device properties. These proposed systems outperformed state-of-the-art lung sound classification systems for both the ICBHI and their multi-channel lung sound dataset.

### 2.1. Our contribution

This work investigates GANs and Federated Learning based architecture for lung sound classification called "auscultation". The aim is to prioritize data privacy which is important in the medical field where the data is tightly regulated and security is highly valued. A new "Fed-WAPR" approach to the Federated Learning Aggregate function has been proposed which has successfully outperformed the FedAVG Aggregation function. We have implemented GAN-generated data samples to further ensure data privacy and security while training. The trained model successfully classifies different respiratory diseases and Heart Failure through lung sounds with a test accuracy of approximately 92% in a decentralized training runtime assuring data privacy.

## 3. System core and design

This section deals with the basic introduction and groundwork of the proposed system. The major components in this system can be broken down into two parts. The first is the procedural generation of data using GANs and the second is the training of this data on our modified FL based architecture.

### 3.1. Generative adversarial networks

Generative Adversarial Networks were first suggested in the work by Gooffellow et al. [12] and consist of two competing neural networks called the Discriminator and the Generator. The generator takes in a random image tensor $z$ and gives an output of the same dimensions as our training dataset. This generated image is then compared with the real image $x$ dataset by the discriminator which tries to classify whether the input image is synthetic or real.

Going into detail, the Generator captures the distribution of data and is trained in such a manner that it tries to maximize the probability of the Discriminator in making a mistake. Whereas the Discriminator is based on the model that estimates the probability that the received input is either from the training data or from the Generator.

This procedure is a constant back and forth of minimax where the Discriminator is trying to minimize its reward $V(D,G)$ and the Generator is trying to minimize the Discriminator's reward or in other words, maximize its loss. It can be mathematically described by the formula as provided in the following equation (1):

$$\min_G \max_D V(D,G) = \mathbb{E}_{x\sim p_{data}(x)}[\log D(x)] + \mathbb{E}_{z\sim p_z(z)}[\log(1 - D(G(z)))] \tag{1}$$

Here, $G$ and $D$ denote the Generator and Discriminator respectively, $p_{data}(x)$ denotes the distribution of real data whereas $P(z)$ denotes the distribution of generated data by the generator. $D(x)$ and $G(z)$ denote the Discriminator network and Generator network respectively.

Once the training procedure is over, the discriminator is turned to idle and the generator weights are frozen and saved. The generator then can be loaded and freely generates numerous realistic images. This is pictorially represented in Fig. 1 denoted as the "generator block".

### 3.2. Federated learning

Federated Learning is a decentralized machine learning paradigm that emphasizes sharing weights instead of data among the different entities involved. Federated learning consists of two major entities namely a global server and clients which can be edge-based devices like smartphones and wearables. The first step in training a Federated model over the clients by downloading the model configuration from the global server. The clients then train the model locally using data present on the device itself. This local training of the model and data isolation in the edge-based devices ensures data privacy. Once the training procedure is completed, the model weights are shared with the global server which aggregates these weights. FedAVG [13] Aggregation at the global model is performed by first equally scaling the weights obtained from the clients by a factor of the reciprocal number of the total number of $n$ clients, i.e., if the total number of clients is 5, the scaling factor would be $\frac{1}{5}$ and if clients are 10, then the scaling factor would be $\frac{1}{10}$. Once all the weights are scaled, they are summated into one global model. This is expressed in Algorithm 1 where $w$ are the weights, $t$ denotes the communication round, $n$ is the client and $K$ is the total number of clients. These global weights are then again downloaded by the clients and the process is repeated until a good accuracy is gained. Consequently, this step also completes one "communication round" in the federated learning system.

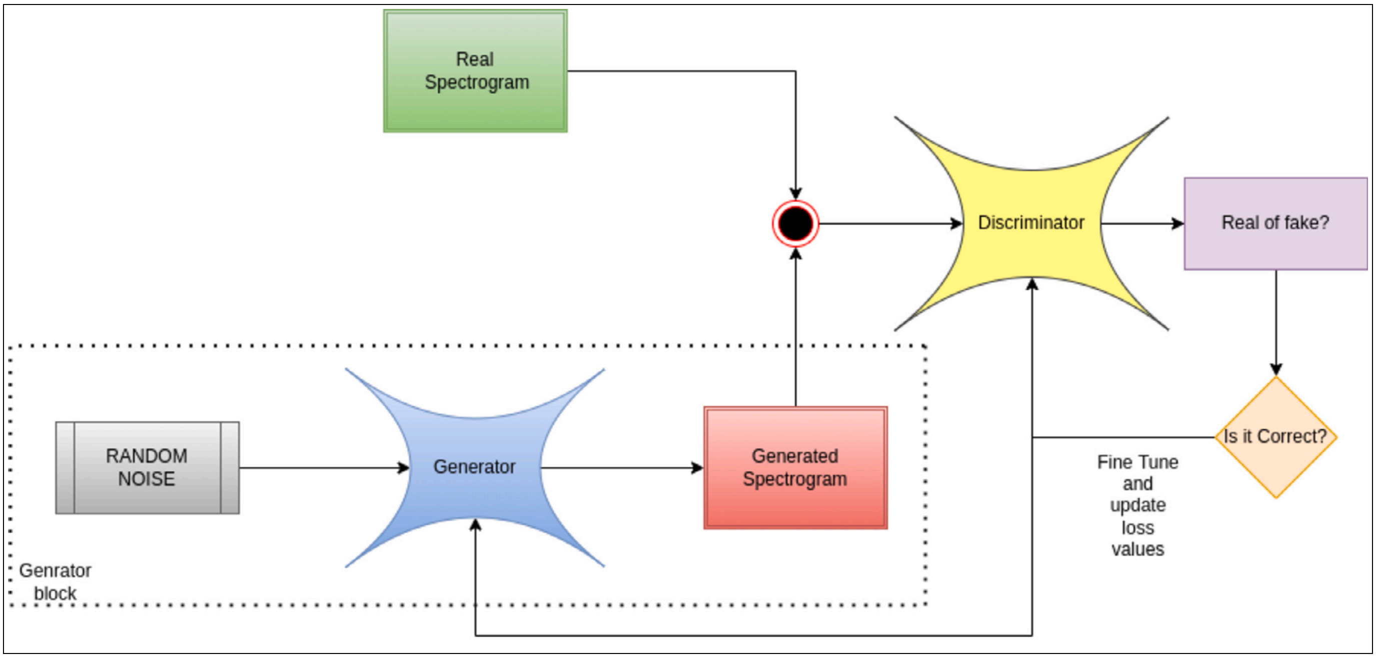


**Fig. 1.** Generative Adversarial Network (GAN) Architecture.

**Algorithm 1** FedAVG.

**Require:** The $K$ clients are indexed by $k$
**Require:** $E$ is the number of local epochs and
**Ensure:** $\eta$ is the learning rate
Server Executes:
Initialize weights $w_0$
**for** `t=1,2,3,...` **do**
  $S_t \leftarrow$ random set of $n$ clients
  **for** `client` $k \in S_t$ `in parallel` **do**
    $w^k_{t+1} \leftarrow ClientUpdate(k, w_t)$
  **end for**
**end for**
$ClientUpdate(k, w)$:
**for** `local epoch 1 to E` **do**
  $w \leftarrow w - loss(w; x, y)$
**end for**
Return $w$ to server

### 3.3. FedWAPR

The FedWAPR Aggregate function works on the basis of ranking the clients based on their accuracies and giving a weighted priority to the client having the highest performance. The algorithm works in the order below

- Once all the clients have completed training on their respective runtimes, they are stacked on the global server in the order of their accuracies, with the client having the highest accuracy ranked at 1.
- A weighted scaling factor is calculated by considering probability distribution functions (PDF) like Log Cauchy [15] and exponential distribution [14] functions as given in Eqns. (2) and (3) respectively.

$$f(x;\lambda) = \begin{cases} \lambda e^{-\lambda x}, & x \geq 0 \\ 0, & x < 0 \end{cases} \tag{2}$$

$$f(x;\mu,\sigma) = \frac{1}{x\pi\sigma\left[1 + (\frac{\ln x - \mu}{\sigma})^2\right]}, x > 0 \tag{3}$$

- In Eqns. (2) and (3), $x$ is the rank of the weight that is selected, $\mu$ and $\sigma$ denote the mean and standard deviation and $\lambda$ is a constant.

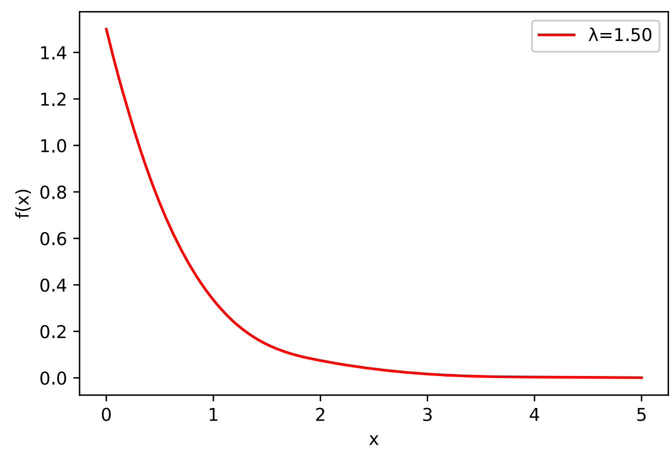


**Fig. 2.** Exponential Probability Density Function.

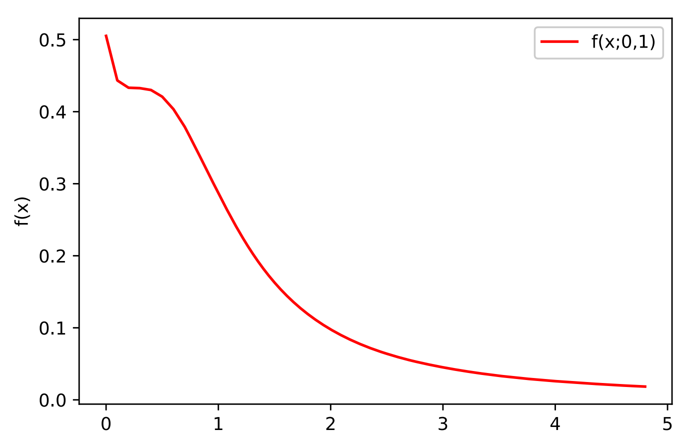


**Fig. 3.** Log Cauchy Probability Density Function.

Exponential Probability Density Function and Log Cauchy Probability Density Function are plotted in Figs. 2 and 3 respectively.

- The reason for selecting these PDFs is that they give a higher weightage to elements closer to zero. So going along with this idea, ranking and scaling the client weights with respect to their accuracies can give a higher priority to the client having higher performance and in turn, increase the performance of the global model while still considering the lower performing clients for better generalization.
- Once the suitable PDF is selected, the $f(x)$ value is calculated for all clients with respect to their ranks. They are then scaled between an interval of [0,1] using the Eqn. (4) which gives us a custom-weighted scaling factor for every weight according to its rank.

$$s_r = \frac{f(r)_r}{\sum_i^k f(i)} \tag{4}$$

where $f$ denotes the PDF selected, $s$ denotes scaling factor, $r$ denotes the rank of the client and $k$ denotes the total number of clients.

- Finally the models are aggregated as given in Eqn. (5) and the global weights are obtained where $w$ denotes weights and $t$ denotes communication round.

$$w_{t+1} = \sum_r^k s_r w_r \tag{5}$$

**Algorithm 2** FedWAPR.

**Require:** The $K$ clients are indexed by $k$
**Require:** $E$ is the number of local epochs
**Require:** $t$ is the selected probability density function
**Ensure:** $\eta$ is the learning rate
Server Executes:
Initialize weights $w_0$ , a pipeline $P$ and weighted scaling factor $A$
**for** `t=1,2,3,...` **do**
  $S_t \leftarrow$ random set of $n$ clients
  **for** `client` $k \in S_t$ `in parallel` **do**
    $w^k_{t+1}, accuracy^k \leftarrow ClientUpdate(k, w_t)$
    Add $w^k_{t+1}, accuracy^k$ to $P$
    Sort $P$ in ascending order of $accuracy^k$
  **end for**
**end for**
**for** `r = 1 to K` **do**
  append $\frac{f(r)}{\sum_{r=0}^K f(i)}$ to $A$
  $w \leftarrow \sum_{r=0}^K A_r w_r$
**end for**
$ClientUpdate(k, w)$:
**for** `local epoch 1 to E` **do**
  $w \leftarrow w - loss(w; x, y)$
**end for**
Return w to server

## 4. Experimental analysis and training

Hardware and Software: The model was trained on Intel core i5, 16 GB RAM using a GTX 1060-Max-Q Graphics card. The dataset was stored on an SSD of size 1TB. The whole training procedure was carried out on a Jupyter notebook with Tensorflow for the creation of the GAN and FL architecture.

### 4.1. Dataset description

To have exhaustive coverage of all lung sound datasets available digitally, we considered the dataset provided by Fraiwan et al. [16]. This dataset was ideal as it combined the publicly available ICBHI dataset [19] and a dataset acquired as part of an ongoing project at King Abdullah University Hospital, Jordan University of Science and Technology (JUT), Irbid, Jordan [16]. The final dataset has a breakdown as given in Table 1.

**Table 1**
Dataset Overview.

| Class | (JUT) dataset samples | ICBHI dataset samples | Total samples |
|---|---|---|---|
| Normal | 105 | 135 | 240 |
| Asthma | 94 | 4 | 98 |
| Heart Failure | 56 | - | 56 |
| Pneumonia | 17 | 148 | 165 |
| Bronchitis (BRON) | 7 | 116 | 123 |
| chronic obstructive pulmonary disease (COPD) | 29 | 773 | 802 |

The ICBHI dataset consists of 920 annotated audio samples from 126 subjects who are diagnosed with the diseases as shown in Table 1. The audios were recorded on stethoscopes like AKGC417L, Meditron, Litt3200, and LittC2SE. The recordings were sampled at 4000 Hz to 44100 Hz which lasted between 10 to 90 seconds. The ICBHI dataset is primarily used for comparison and evaluation in our system whereas the JUT dataset is used as test data in the procedural generation of data using GANS given the fact that this dataset has availability of the Heart Failure Class.

The JUT dataset is recorded in clinical trials of 16 patients diagnosed with Idiopathic Pulmonary Fibrosis (IPF). The lung sounds in this dataset were sampled at 16 kHz within 5 to 30-second intervals to accommodate the full respiratory cycles of the patients. The sensor signals are additionally filtered with a Bessel High-pass filter with a cut-off frequency of 80 Hz and a slope of 24 dB/oct.

We also employ the MNIST dataset [35] to further test the performance of our custom FedWAPR aggregated model. The MNIST (Mixed National Institute of Standards and Technology database), is a large database of handwritten digits collected by the National Institute of Standards and Technology, including 60,000 training images and 10,000 test images. Using the MNIST dataset in the experimental analysis offers numerous advantages within the context of evaluating the custom FedWAPR architecture. The MNIST dataset is a widely recognized benchmark dataset that facilitates rigorous comparison, demonstrating the architecture's adaptability beyond its original training data. By introducing variations in data distribution, MNIST tests the model's robustness and generalizability. Through assessments with different numbers of clients, the dataset evaluates the architecture's scalability in diverse scenarios. Experimentation with hyperparameters using MNIST sheds light on its sensitivity and informs effective parameter tuning. In summary, MNIST's multifaceted utility underpins comprehensive evaluations, highlighting the architecture's versatility, resilience, scalability, and optimization potential.

### 4.2. Pre-processing

By referencing previous works [17,18,21], in order to obtain maximum feature extraction and easier GAN-based generation from the audio data, we preprocess the audio files into Mel spectrograms [11]. A spectrogram can be defined as a visual representation of the various signal strengths in an audio recording over different ranges of frequencies. Fig. 4 demonstrates the waveform of a dataset sample of an asthma patient's breathing which was then converted into a spectrogram using the Python library Matplotlib. The spectrogram was generated using a non-equispaced fast Fourier transform (NFFT) of 2048 and with an 'inferno' themed colormap.

The resulting spectrogram is represented in Fig. 5(a). Similarly, other classes of spectrograms are also represented in Figs. 5(b) and 5(c).

Since a sparsity of image data was observed in the JUT dataset (especially in class Heart Failure), we procedurally generate spectrograms using GANs from already existing data samples as reference. The GAN architecture created, had a generator network architecture as shown in Fig. 6 and a discriminator architecture in Fig. 7.

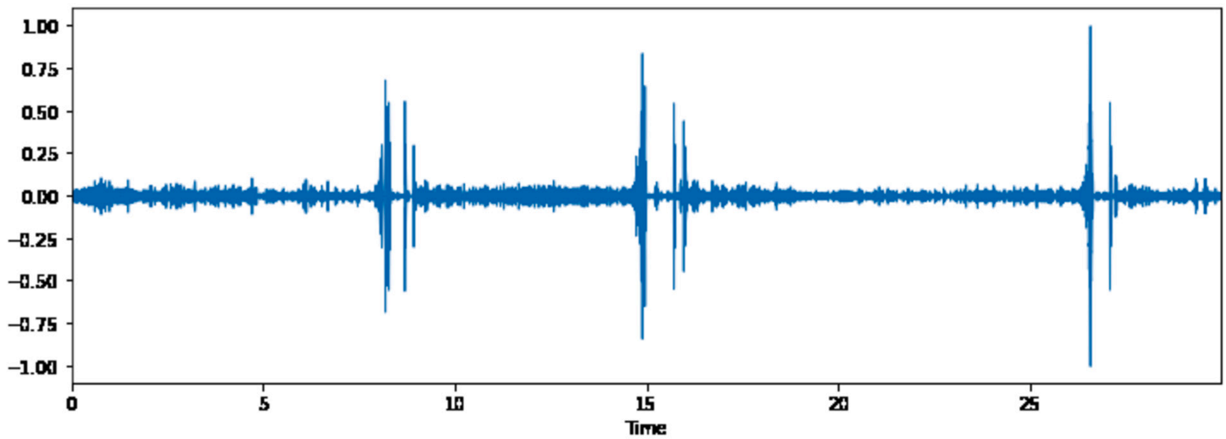


**Fig. 4.** Asthma Audio Waveform.

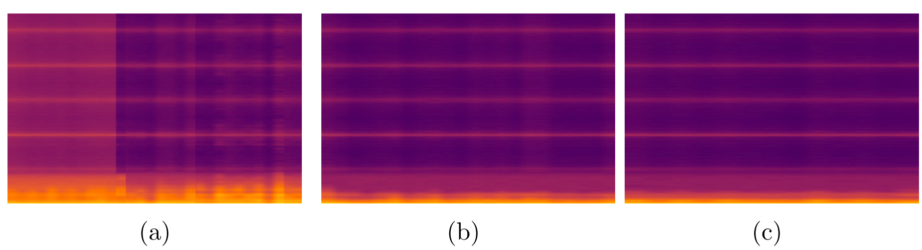


**Fig. 5.** (a) Asthma (b) COPD (c) Normal.

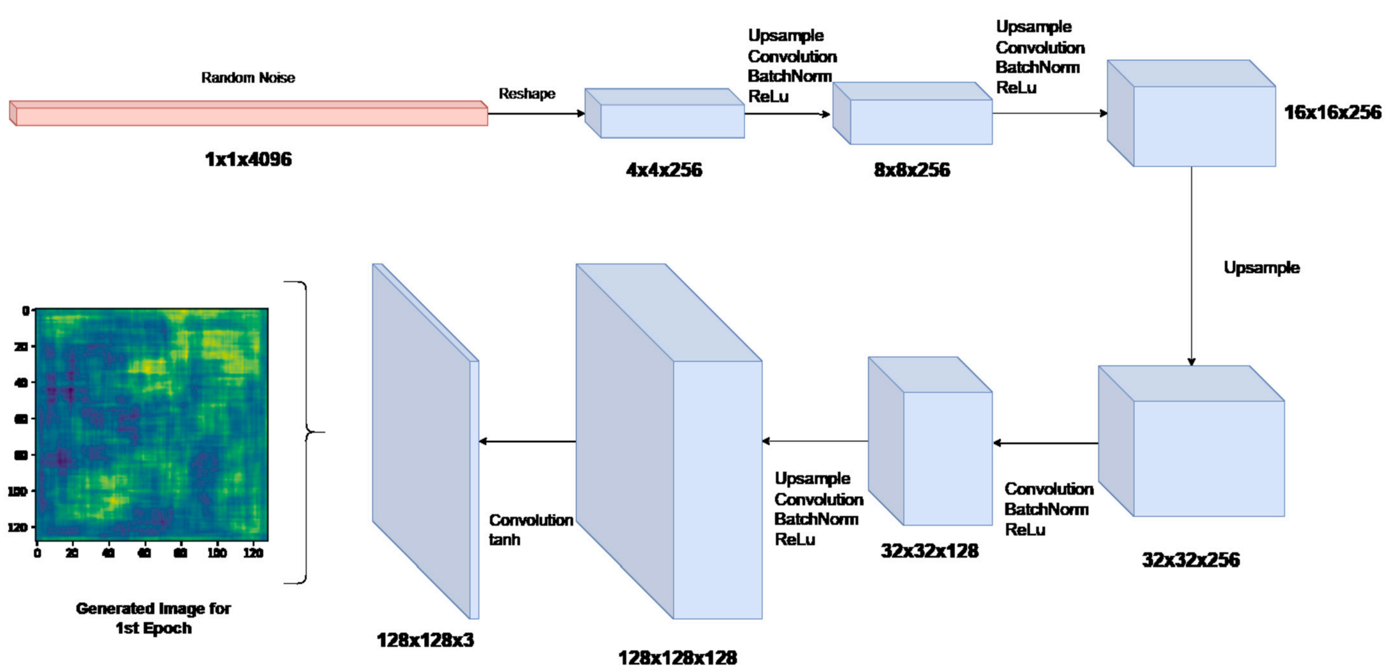


**Fig. 6.** Generator Neural Network Architecture.

The generator takes an input shape of (128,128) three-channel random noise image and was trained to output spectrograms of the same dimensions. Whereas the Discriminator was a binary classifier that took the generated image and classified the image as real or synthetic. Both the architectures utilized the Adam optimizer for its loss. Additionally, this loss function was custom-created as given in Eqn. (1).

The GAN network was tested at various epoch levels. This was carried out to maximize the quality of generated synthetic data to be as similar as possible to the real data. To quantify the value of this newly generated data, we trained a neural network model on the generated data and tested it against the real dataset. This was to verify whether the generated data samples retained key features from their real counterpart and whether a neural network model was able to classify this dataset accurately. Table 2 quantifies at which epoch level the generated images gained adequate quality to be reliable.

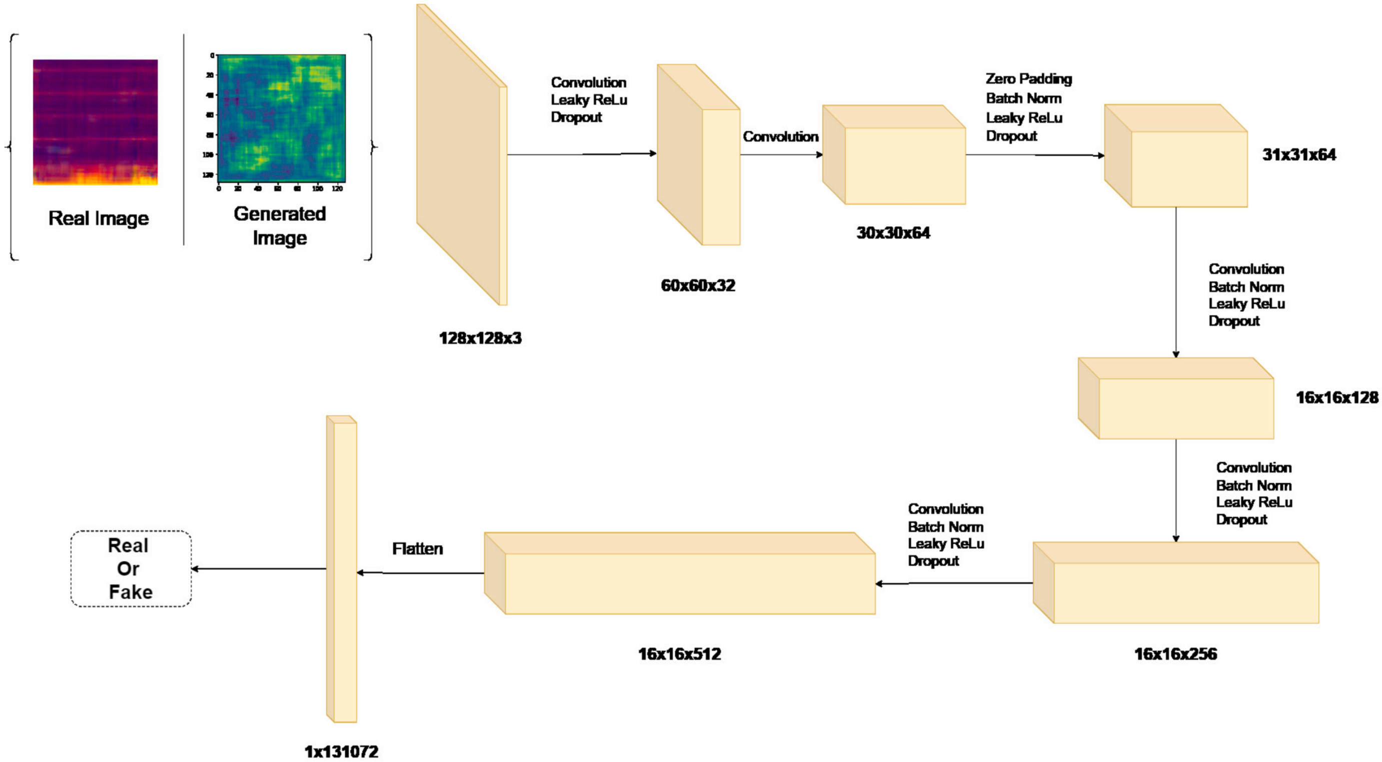


**Fig. 7.** Discriminator Neural Network Architecture.

**Table 2**
GAN model performance with respect to epochs.

| GAN model epochs | Accuracy achieved on original data |
|---|---|
| 500 | 54.3% |
| 1000 | 89.97% |
| 2000 | 92.24% |

From Table 2 we observed an order of diminishing return using the existing dataset. Thus we stop at the 2000th epoch and freeze the Generator weights to create our dataset. Finally, 500 images were generated on this configuration for every data class in the JUT dataset for training on the FL architecture.

### 4.3. Federated learning

The FL architecture was rigorously trained and tested on different architectures for maximum comparative analysis. Each client was subjected to a custom neural network trained for 1 epoch under the "Adam" optimizer. In our testing, the above configuration and the model architecture in Fig. 8 provided the best performance on all configurations of the FL architecture.

Once the training was complete, the convergence of the model accuracy per communication round is denoted in Fig. 9. Here, the generated dataset was first subjected to a traditional centralized model training which resulted in an accuracy of 92.24% at the 394th epoch. After this, the model was subjected to FL training utilizing the FedAVG averaging algorithm (denoted in blue thick line) which resulted in a drop in accuracy to 77%. Another implementation of the FedAVG averaging FL method was to increase the epoch numbers to 15 (orange dotted line) which was able to achieve performance on par with the centralized model. However, this approach of training for 15 epochs exponentially increased the training time of the model to almost 23 hours.

The FedWAPR algorithm was then implemented through the Log Cauchy and exponential PDF functions for 1 epoch. In our testing, it was observed that keeping the $\lambda$ value as 1.5 in the exponential PDF, $\mu$ as 0, and $\sigma$ as 1 in the Log Cauchy function provided the best-weighted rank distribution. Another modification was scaling the ranks of the PDF function by $\frac{1}{8}$ (brown dotted line) and $\frac{1}{16}$ (purple line) (i.e. f(2) becomes f(2/8) in the PDF function according to Eqn. (2)) providing the fastest convergence using FedWAP. Certain exception was noticed for the exponential PDF aggregation where the $\frac{1}{8}$ (red dotted line) scaling performed better than the $\frac{1}{16}$ (green line) scaling. Besides this, a notable improvement observed was in the performance of the Log Cauchy averaging at $\frac{1}{16}$ scaling which was able to perform similarly to the 15 epoch FedAVG model with a training time required for 1 epoch, i.e., approximately 16 minutes.

The key performance analysis of the configuration can be observed in Table 3.

The model was also trained using the MNIST dataset to test the viability of the proposed custom FedWAPR FL architecture on a different dataset and having different numbers of clients. The model was trained until 15 clients as increasing this number gave diminishing results. Once the training was completed, a trend of scaling the PDF ranks by 1/16 provided the best results with an increase in performance for models having a lower number of clients as shown in Figs. 10 and 11.

From Fig. 10 we can also notice that the Log Cauchy distribution also performed the best out of all the configurations given the fact that this model provided a better generalization of models having lower performance. This was possible due to the presence of the slight plateau observed between 0 and 1 as seen in Fig. 3 this prioritized the weights of low/mid performing models better than the exponential PDF model.

The performance breakdown is given in Table 4. It can be observed in Table 4 and Fig. 11, the Log Cauchy PDF has the best performance in terms of either or both max accuracy and converging speed (94% accuracy) compared to the other configurations.

### 4.4. Observations

Some general observations and trends that were noticed are:

- The involvement of GANs in training the model provided an extra layer of data security as we are training with data of people

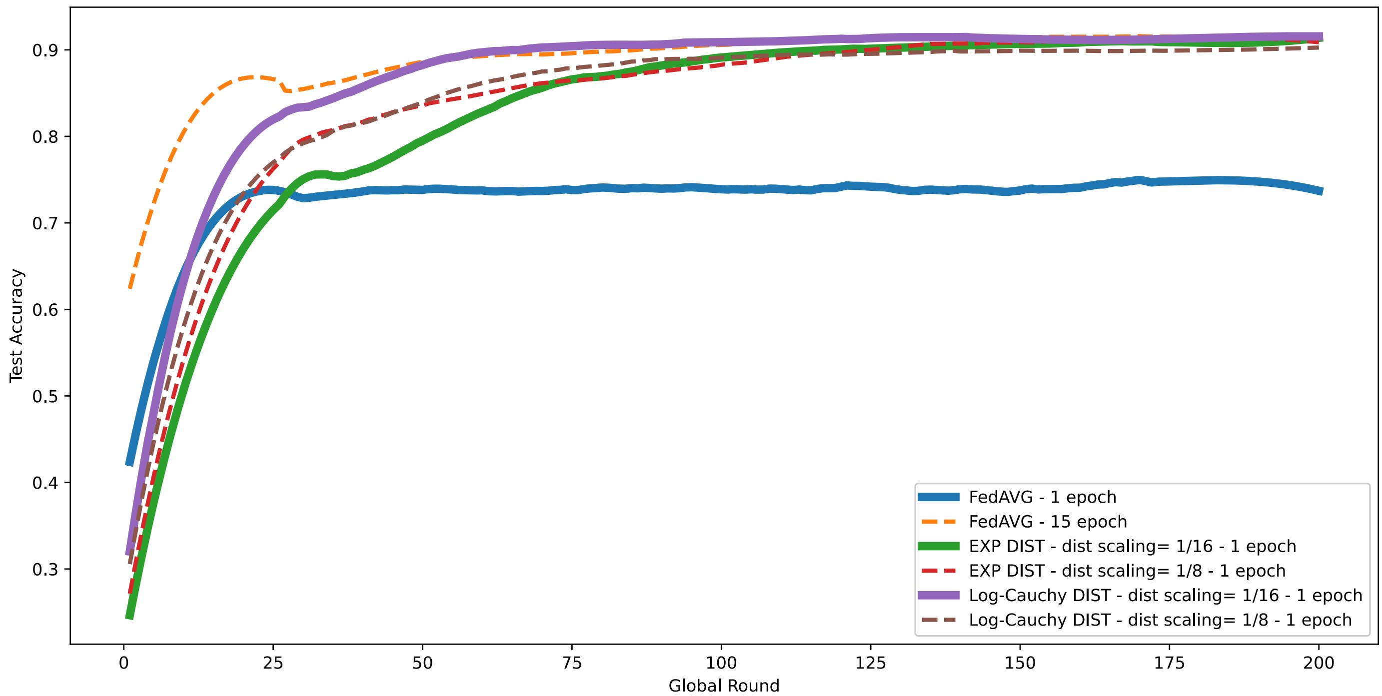


**Fig. 8.** Training trend of best performing models.

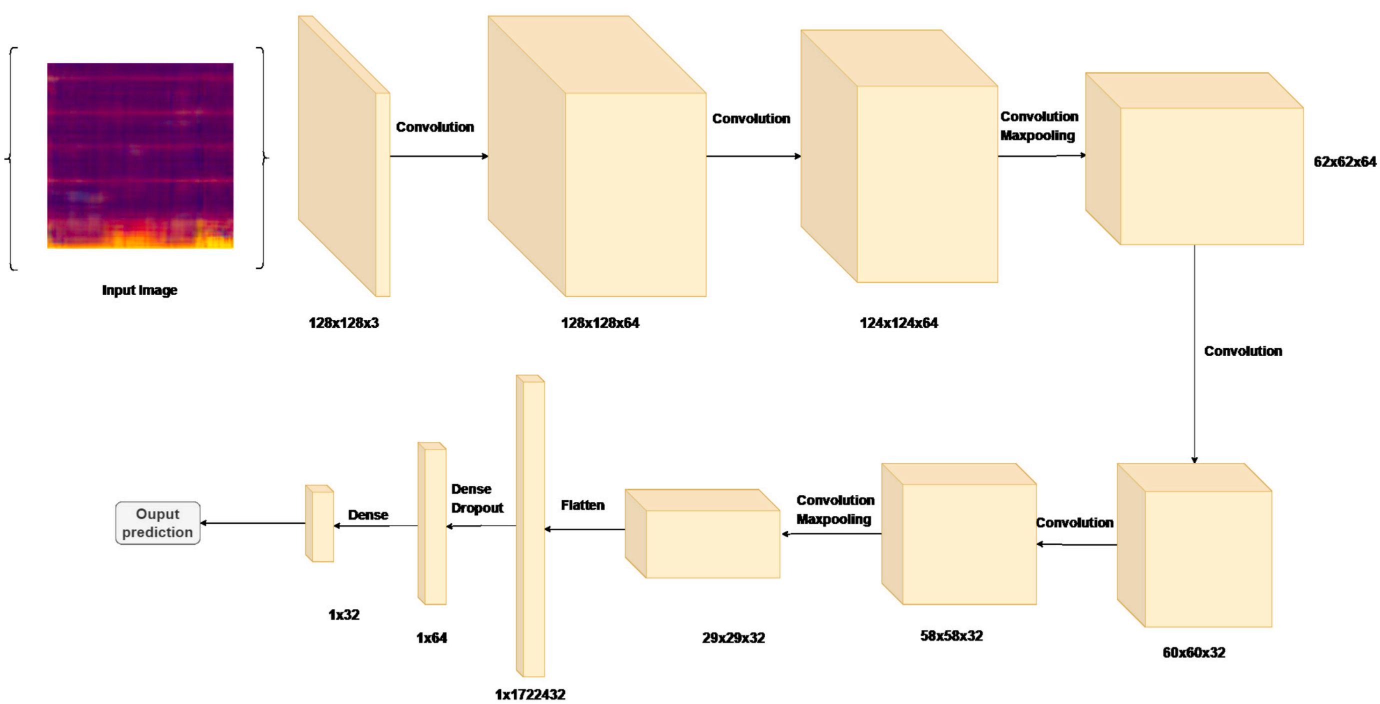


**Fig. 9.** Best performing FL Neural Network Architecture.

**Table 3**
Configured Model Performance.

| Model configuration | Max accuracy achieved | Comm round to reach 85% |
|---|---|---|
| Centralized model | 92.24%(394 epochs) | - |
| FedAVG - 1 epoch | 77.39% | - |
| FedAVG - 15 epoch | 91.78% | 31 |
| Exponential PDF scaled by 1/8 | 92.1% | 54 |
| Exponential PDF scaled by 1/16 | 91.78% | 60 |
| Log Cauchy PDF scaled by 1/8 | 90.6% | 47 |
| **Log Cauchy PDF scaled by 1/16** | **92.1%** | **38** |

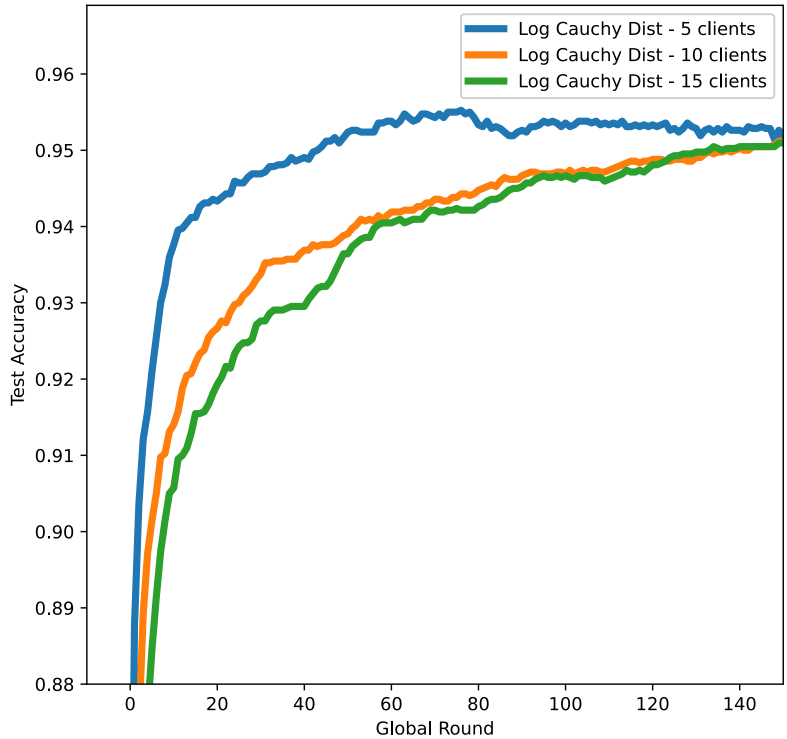


**Fig. 10.** Training trends of best performing models w.r.t. clients on MNIST.

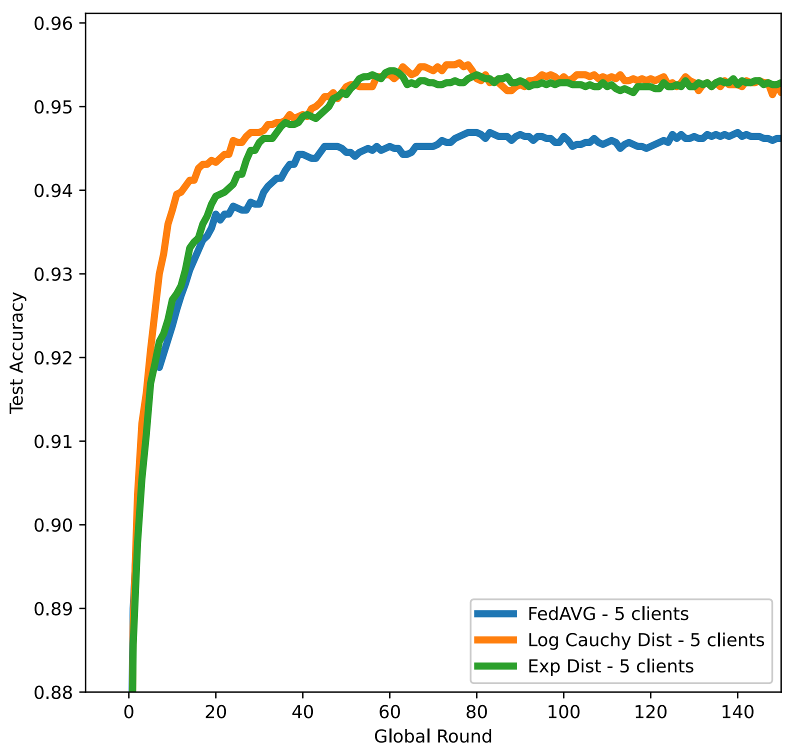


**Fig. 11.** Training trends of best performing models on MNIST.

who "don't exist", thus maintaining the privacy of real users in the process.

- The GAN-generated images enabled a deep learning approach in the JUT dataset due to its lack of quantity of data.
- The Log Cauchy PDF model performed the best to converge faster and reach the highest accuracy at 92.1% when its ranks are scaled to 1/16 as compared to FedAVG which failed to achieve an accuracy above 90%.
- Additionally, the Log Cauchy PDF model took just ≈ 16 minutes to complete 200 global rounds compared to the FedAVG model trained for 15 epochs per client which took ≈ 23 hours to do the same.

**Table 4**
FL model performance with varying clients on MNIST dataset.

| Model configuration | Max accuracy achieved | Comm rounds to reach 94% acc |
|---|---|---|
| FedAVG - 5 client | 94.7% | 32 |
| FedAVG - 10 client | 95.07% | 52 |
| FedAVG - 15 client | 95.1% | 57 |
| Exponential PDF- 5 client | 95.42% | 46 |
| Exponential PDF- 10 client | 94.67% | 78 |
| Exponential PDF- 15 client | 95.11% | 63 |
| **Log Cauchy PDF- 5 client** | **95.52%** | **13** |
| **Log Cauchy PDF- 10 client** | **95.28%** | 52 |
| **Log Cauchy PDF- 15 client** | **95.2%** | 57 |

- This configuration also performed on par with the centralized model performance which had a max accuracy of 92.24% attained at its 394th epoch.
- In the MNIST dataset, the Log Cauchy PDF method achieved a stable accuracy in 19 communication rounds before the FedAVG configuration. This is an increase in performance by almost 60%.
- The exponential PDF method observed increased performance at higher client numbers and when scaled by $\frac{1}{8}$ configurations, although still outweighed by the Log Cauchy PDF by performance.

## 5. Conclusion and future works

This paper addresses the work on a GAN+FL based architecture for lung sound classification to prioritize data privacy which is important in the medical field where the data is tightly regulated and its security is highly valued. Additionally, we have also proposed a novel FedWAPR approach to the FL aggregate function which has successfully and drastically outperformed the FedAVG method. The trained model successfully classifies different respiratory diseases and Heart Failure through lung sounds with a test accuracy of approximately 92% in a decentralized training runtime. Through this optimized federated learning approach, a patient can be examined for respiratory diseases without requiring extensive sensitive data recording and ensuring the obtained data sample is secured. In the future, we aim to test the FedWAPR aggregate function on other FL-based issues like heterogeneous datasets and also subject it to more forms of quantities and qualities of datasets. The algorithm also needs to be tested with the variety of PDF functions available like beta, gamma, and Burr PDF to test which PDF function provides the best performance. Further experimentation on this algorithm's computation requirement and performance needs to be compared with other aggregation methods like FedMA [36], FedBE [37], and FedProx [38].

## Statements of ethical approval

Not applicable

## Declaration of competing interest

The authors of the manuscript entitled "Weighted Aggregation through Probability based Ranking: An Optimized Federated Learning Architecture to Classify Respiratory Diseases" hereby declare no conflict of interest.

## Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

## References

[1] C. Leroyer, L. Perfetti, C. Trudeau, J. L'ARCHEVÊQUE, M. Chan-Yeung, J.-L. Malo, Comparison of serial monitoring of peak expiratory flow and fev1 in the diagnosis of occupational asthma, Am. J. Respir. Crit. Care Med. 158 (3) (1998) 827–832.
[2] W.M. van Aalderen, Childhood asthma: diagnosis and treatment, Scientifica (2012) 2012.
[3] C.D. Mathers, D. Loncar, Projections of global mortality and burden of disease from 2002 to 2030, PLoS Med. 3 (11) (2006) e442.
[4] A. Bohadana, G. Izbicki, S.S. Kraman, Fundamentals of lung auscultation, N. Engl. J. Med. 370 (8) (2014) 744–751.
[5] R.X.A. Pramono, S.A. Imtiaz, E. Rodriguez-Villegas, Evaluation of features for classification of wheezes and normal respiratory sounds, PLoS ONE 14 (3) (2019) e0213659.
[6] H. Pasterkamp, S.S. Kraman, G.R. Wodicka, Respiratory sounds: advances beyond the stethoscope, Am. J. Respir. Crit. Care Med. 156 (3) (1997) 974–987.
[7] S. Reichert, R. Gass, C. Brandt, E. Andrès, Analysis of respiratory sounds: state of the art, Clin. Med. Circ. Respirat. Pulm. Med. 2 (2008) CCRPM–S530.
[8] L. Shi, K. Du, C. Zhang, H. Ma, W. Yan, Lung sound recognition algorithm based on vggish-bigru, IEEE Access 7 (2019) 139438–139449.
[9] B. McMahan, E. Moore, D. Ramage, S. Hampson, B.A. y Arcas, Communication-efficient learning of deep networks from decentralized data, in: Artificial Intelligence and Statistics, PMLR, 2017, pp. 1273–1282.
[10] J. Konečnỳ, H.B. McMahan, D. Ramage, P. Richtárik, Federated optimization: distributed machine learning for on-device intelligence, arXiv preprint arXiv:1610.02527, 2016.
[11] S.S. Stevens, J. Volkmann, E.B. Newman, A scale for the measurement of the psychological magnitude pitch, J. Acoust. Soc. Am. 8 (3) (1937) 185–190.
[12] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, Y. Bengio, Generative adversarial nets, Adv. Neural Inf. Process. Syst. 27 (2014).
[13] X. Li, K. Huang, W. Yang, S. Wang, Z. Zhang, On the convergence of fedavg on non-IID data, in: 8th International Conference on Learning Representations, ICLR 2020, Addis Ababa, Ethiopia, April 26-30, 2020, 2020, pp. 1–26. OpenReview.net, https://openreview.net/forum?id=HJxNAnVtDS.
[14] M. Lugo, The expectation of the maximum of exponentials, Stat 134 (2011).
[15] D.J. Olive, Applied robust statistics, Preprint M-02-006 2008.
[16] L. Fraiwan, O. Hassanin, M. Fraiwan, B. Khassawneh, A.M. Ibnian, M. Alkhodari, Automatic identification of respiratory diseases from stethoscopic lung sound signals using ensemble classifiers, Biocybern. Biomed. Eng. 41 (1) (2021) 1–14.
[17] M.U. Khan, A. Mobeen, S. Samer, A. Samer, Embedded system design for real-time detection of asthmatic diseases using lung sounds in cepstral domain, in: 6th International Electrical Engineering Conference (IEEC 2021), April, 2021 at NEDUET, 2021, pp. 1–5.
[18] M.T. García-Ordás, J.A. Benítez-Andrades, I. García-Rodríguez, C. Benavides, H. Alaiz-Moretón, Detecting respiratory pathologies using convolutional neural networks and variational autoencoders for unbalancing data, Sensors 20 (4) (2020) 1214.
[19] B. Rocha, D. Filos, L. Mendes, I. Vogiatzis, E. Perantoni, E. Kaimakamis, P. Natsiavas, A. Oliveira, C. Jácome, A. Marques, et al., A respiratory sound database for the development of automated classification, in: International Conference on Biomedical and Health Informatics, Springer, 2017, pp. 33–37.
[20] V. Basu, S. Rana, Respiratory diseases recognition through respiratory sound with the help of deep neural network, in: 2020 4th International Conference on Computational Intelligence and Networks (CINE), IEEE, 2020, pp. 1–6.
[21] Z. Tariq, S.K. Shah, Y. Lee, Lung disease classification using deep convolutional neural network, in: 2019 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), IEEE, 2019, pp. 732–735.
[22] Q. Dou, T.Y. So, M. Jiang, Q. Liu, V. Vardhanabhuti, G. Kaissis, Z. Li, W. Si, H.H. Lee, K. Yu, et al., Federated deep learning for detecting Covid-19 lung abnormalities in ct: a privacy-preserving multinational validation study, npj Digit. Med. 4 (1) (2021) 1–11.
[23] I. Feki, S. Ammar, Y. Kessentini, K. Muhammad, Federated learning for Covid-19 screening from chest x-ray images, Appl. Soft Comput. 106 (2021) 107330.
[24] F. Tong, L. Liu, X. Xie, Q. Hong, L. Li, Respiratory sound classification: from fluid-solid coupling analysis to feature-band attention, IEEE Access 10 (2022) 22018–22031.
[25] A.E. Cetinkaya, M. Akin, S. Sagiroglu, A communication efficient federated learning approach to multi chest diseases classification, in: 2021 6th International Conference on Computer Science and Engineering (UBMK), IEEE, 2021, pp. 429–434.

[26] M. Fraiwan, L. Fraiwan, M. Alkhodari, O. Hassanin, Recognition of pulmonary diseases from lung sounds using convolutional neural networks and long short-term memory, J. Ambient Intell. Humaniz. Comput. 13 (10) (2022) 4759–4771.

[27] M. Alkhodari, A.H. Khandoker, Detection of Covid-19 in smartphone-based breathing recordings: a pre-screening deep learning tool, PLoS ONE 17 (1) (2022) e0262448.

[28] J. Saldanha, S. Chakraborty, S. Patil, K. Kotecha, S. Kumar, A. Nayyar, Data augmentation using variationalautoencoders for improvement of respiratory disease classification, PLoS ONE 17 (8) (2022) e0266467.

[29] B. Tasar, O. Yaman, T. Tuncer, Accurate respiratory sound classification model based on piccolo pattern, Appl. Acoust. 188 (2022) 108589.

[30] N.S. Haider, A. Behera, Computerized lung sound based classification of asthma and chronic obstructive pulmonary disease (copd), Biocybern. Biomed. Eng. 42 (1) (2022) 42–59.

[31] K. Kochetov, A. Filchenkov, Generative adversarial networks for respiratory sound augmentation, in: Proceedings of the 2020 1st International Conference on Control, Robotics and Intelligent System, 2020, pp. 106–111.

[32] S. Jayalakshmy, G.F. Sudha, Conditional gan based augmentation for predictive modeling of respiratory signals, Comput. Biol. Med. 138 (2021) 104930.

[33] T. Malygina, E. Ericheva, I. Drokin, Gans' n lungs: improving pneumonia prediction, arXiv preprint, arXiv:1908.00433, 2019.

[34] F.-S. Hsu, S.-R. Huang, C.-W. Huang, C.-J. Huang, Y.-R. Cheng, C.-C. Chen, J. Hsiao, C.-W. Chen, L.-C. Chen, Y.-C. Lai, et al., Benchmarking of eight recurrent neural network variants for breath phase and adventitious sound detection on a self-developed open-access lung sound database—hf_lung_v1, PLoS ONE 16 (7) (2021) e0254134.

[35] L. Deng, The mnist database of handwritten digit images for machine learning research [best of the web], IEEE Signal Process. Mag. 29 (6) (2012) 141–142.

[36] H. Wang, M. Yurochkin, Y. Sun, D. Papailiopoulos, Y. Khazaeni, Federated learning with matched averaging, arXiv preprint arXiv:2002.06440, 2020.

[37] H.-Y. Chen, W.-L. Chao Fedbe, Making Bayesian model ensemble applicable to federated learning, arXiv preprint arXiv:2009.01974, 2020.

[38] T. Li, A.K. Sahu, M. Zaheer, M. Sanjabi, A. Talwalkar, V. Smith, Federated optimization in heterogeneous networks, Proc. Mach. Learn. Syst. 2 (2020) 429–450.