IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 6 October 2024, accepted 17 October 2024, date of publication 23 October 2024, date of current version 11 November 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3485100

RESEARCH ARTICLE

Federated Learning With Dataset Splitting and Weighted Mean Using Particle Swarm Optimization

MOHIT AGARWAL[1], **(Member, IEEE), GARIMA JAISWAL**[1], **(Member, IEEE), ROHIT KUMAR KALIYAR**[1], **(Senior Member, IEEE), AKANSHA SINGH**[1], **KRISHNA KANT SINGH**[2], **S. S. ASKAR**[3], **AND MOHAMED ABOUHAWWASH**[4]

[1]School of CSET, Bennett University, Greater Noida 201310, India
[2]Delhi Technical Campus, Greater Noida 201306, India
[3]Department of Statistics and Operations Research, College of Science, King Saud University, Riyadh 11451, Saudi Arabia
[4]Department of Mathematics, Faculty of Science, Mansoura University, Mansoura 35516, Egypt

Corresponding author: Akansha Singh (akansha1.singh@bennett.edu.in)

This work was supported by King Saud University, Riyadh, Saudi Arabia, through the Researchers Supporting Project, under Grant RSP2025R167.

ABSTRACT Federated learning uses the concept of decentralized training of *n* number of local clients for a small number of epochs say 2-5, and then averaging the learned weights of all local clients, and evaluating on test dataset with the average weights loaded to a global model. The train dataset is split into *n* clusters and each cluster acts as a distributed data for each local model. Each round of weight averaging and then uploading the average weights on each local client for further training is called communication round and it was observed that similar accuracy can be obtained with a lesser amount of training time. In this paper, instead of averaging the weights, a weighted mean concept was developed where the PSO vector helps to find the weight values for the best accuracy of a global model. It was found that PSO can help in two ways by bettering the accuracy and also reducing the training time. The proposed approach can enhance the performance of pre-trained models like AlexNet, VGG16, InceptionV3, and ResNet50 on CIFAR-10 and CIFAR-100 datasets. The maximum increase was found with VGG16 of around 26.01% for CIFAR-10 and 26.84% for CIFAR-100. Similarly, on the Tomato dataset, AlexNet accuracy can be increased by 28.56%. Multi-modal model accuracy on the fake news dataset was also enhanced by 8.21%.

INDEX TERMS Federated learning, particle swarm optimization, optimization, model performance, multimodal.

I. INTRODUCTION

This paper explains the usage of Federated learning for deep learning models showing its advantage in saving training time and getting better performance of models.

A. FEDERATED LEARNING

Machine learning has developed into an amazing tool in many fields, including marketing, finance, healthcare, and education, in the contemporary environment. It stands out due to its support with practical decision-making and unique features for problem-solving. Despite this, there are significant problems with machine learning that prevent its use in practical situations. Unbalanced data and the lack of datasets are two main obstacles to ML algorithms. For instance, in the medical field, there can be a data shortage in the case of rare diseases, which makes it difficult to design precise models. Accounting models may be skewed or lacking if there is a lack of data for particular market situations. Model choice and hyperparameter adjustment are additional restrictions for machine learning algorithms [1]. Depending on the data and issue description, different models

The associate editor coordinating the review of this manuscript and approving it for publication was Mehedi Masud.

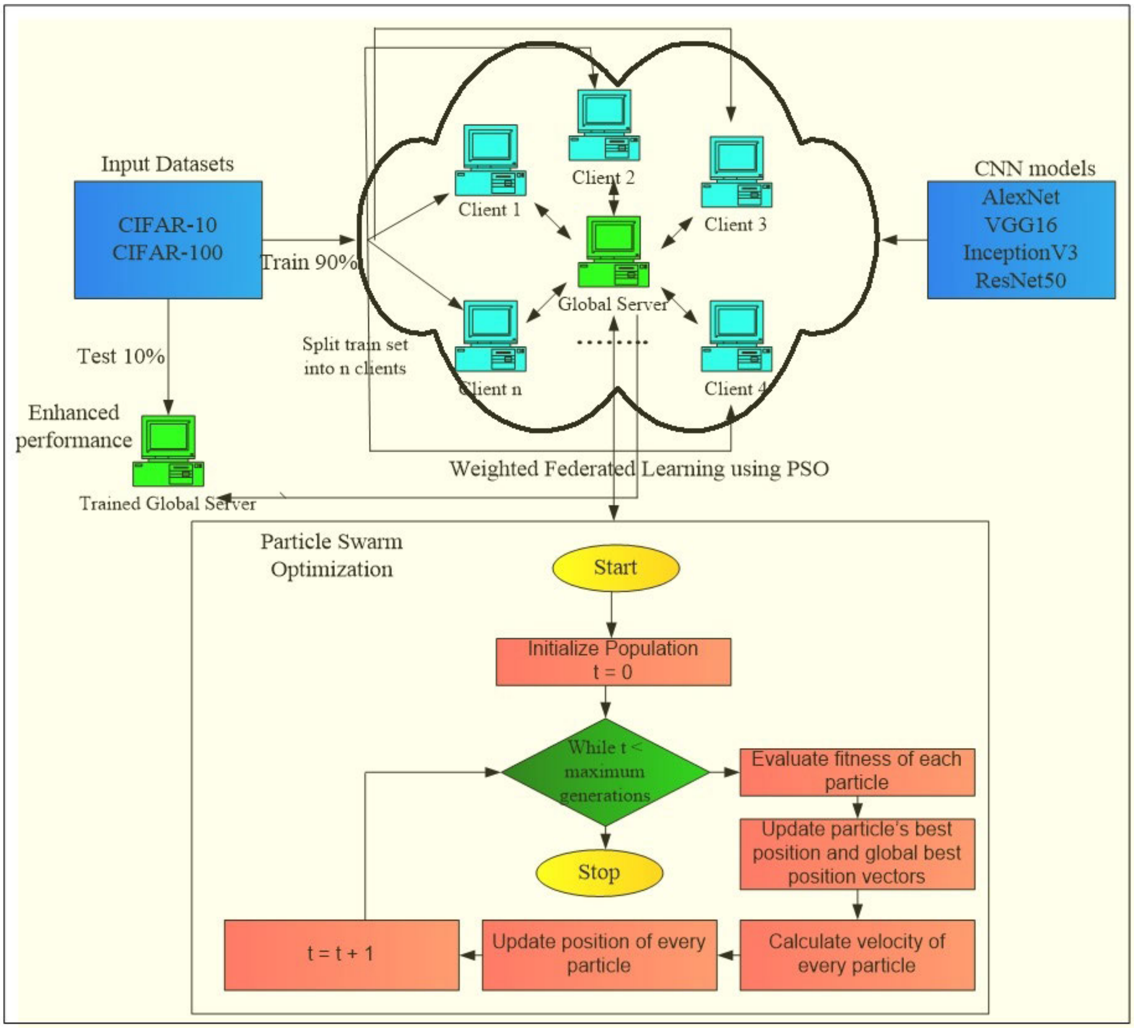


FIGURE 1. Graphical abstract.

may produce different solutions. To achieve the best results, it is essential to choose the right model and adjust the settings. Common restrictions include overfitting and underfitting, which occur when models perform poorly on both training and unobserved data or well on training data but poorly on unobserved data.

A need is also present to handle security and privacy concerns, especially when working with sensitive or private data. By allowing different parties to work together on the training of a common model without disclosing their raw data, federated learning provides a solution to these problems [2]. Privacy is maintained and the chance of data breaches is lower because the raw data is still on the edge devices. Moreover, federated learning makes decentralized computing possible since it uses edge devices to train models rather than a central processing unit [3]. Additionally, because it can be used to train models using data from a large number of edge devices, federated learning is now very scalable.

Federated learning has several advantages that make it a viable answer to machine learning problems. Keeping the raw data on edge devices and only distributing the modified model weights, first protects privacy [4]. Second, it permits decentralized computing, which increases its scalability and effectiveness. Thirdly, it encourages cooperation between various parties, enabling the exchange of knowledge and skills. Fourthly, it allows for real-time model updates, making it appropriate for uses where the data changes quickly.

Several sectors can use federated learning in a variety of ways. Federated learning can be used in the healthcare industry to train models using patient data from many hospitals while protecting patient privacy [1]. It can be used in finance to train models on data from many banks while safeguarding the privacy of the banks and their customers. It can be used in transportation to train models on information from various vehicles, enabling more precise and effective navigation. It can be used in sensors as part of the Internet of Things (IoT) to train models using the data the sensors collect, enabling more intelligent and individualized services [5]. In all, machine learning is a widely used approach used in multiple domains. Privacy protection, distributed computing, scalability, and cooperation are some of the issues that impose restrictions on its usage in real-world scenarios. Federated learning provides potential solutions to these problems thus encouraging secure, reliable, and effective mechanisms for data protection and privacy [4].

B. FEDERATED LEARNING INNOVATIONS

To simulate Federated learning datasets of CIFAR-10, CIFAR-100 [6], tomato [7], and multimodal Fake News [8] were utilized. The dataset was initially randomly divided into ten clusters randomly at run time and each cluster of the dataset was passed as training data for ten local clients. Each local client was trained for 5-10 epochs and their weights average was loaded to a global model which was tested on a test set that was separated from the original dataset before splitting into ten clusters. The last step of weight averaging and testing was called as communicating round and federated learning was executed for a hundred communicating rounds.

The process of dataset division into 10 clusters and training 10 clients was repeated with different numbers of clients such as 2, 5, 7, 10, etc. Particle Swarm Optimization (PSO) [9] was introduced in 1995 to do a weighted averaging of local model weights before loading in the global model. The choice of PSO was made as this is one of the earliest nature-inspired algorithms which has proven its effectiveness over several years. We could not use methods like Genetic Algorithm as it uses binary gene values and in this research continuous decimal values were needed as different weights in range 0-1 were needed for doing weighted averaging of local model weights. Some other heuristics algorithms like Gravitational Search Algorithm [10] were also not considered as later algorithms have taken inspiration from existing original meta-heuristics methods with a slight deviation from them. This process helped to better the accuracy than simple averaging of local model weights. It was found that optimum learning was obtained with Federated Learning and PSO with 10 clients. This could achieve maximum accuracy in 20 communication rounds of nearly 92.06% using the CIFAR-10 dataset and VGG16 pre-trained model. The attainment of good results over CIFAR-10 and CIFAR-100 datasets using pre-trained models helped to diversify the experiments with more datasets and more models. The graphical view of the proposed methodology is shown in Figure 1.

II. BACKGROUND LITERATURE

In this section a brief survey of existing research works on Federated Learning (FL) is discussed. It discusses how research has grown from plain averaging of clients' weights to personalized FL where heterogeneity of data is taken into consideration. It also discusses briefly how PSO has been tried by researchers in FL problems.

McMahan et al. [11] have proposed Federated Learning and tested with five different models including Multilayer Perceptron, Convolution Neural Networks and LSTM networks on four different datasets and found that their proposed model FedAvg could enhance performance and reduce communication rounds cost hundred times as compared to synchronized stochastic gradient descent. In FedAvg proposed model trained clients separately and in each communication round averaged the weights at central server and resent to clients for further training on local data. This helped in data privacy also.

Dinh et al. [12] have proposed a personalized Federated Learning pFedMe which uses Moreau envelopes to optimize personalized individual clients on local data. It samples a subset of clients to receive model weights and uses an additional parameter β to update the weights depending on the previous iteration's global model weights and current average of clients' weights. Authors show that pFedMe achieves quadratic speedup for strongly convex problems.

Ma et al. [13] have proposed a Federated Learning model Layer-wised Personalized Federated learning (pFedLA) which helps in the aggregation of weights layer wise depending on the similarity in clients' datasets. Authors demonstrate on EMNIST, FashionMNIST, CIFAR-10 and CIFAR-100 datasets better performance and reduced communication costs.

Yang et al. [14] have proposed Federated Learning for client-specific Prompt Generation (pFedPG) in Vision Transformers (ViT) based models. Prompts are task specific parameters and are prepended to input tokens of pre-trained ViT. In the case of heterogeneous data prompts learned from separate clients cannot be simply averaged hence a global level prompt generation is also used. This method is effective in reducing training and communication rounds cost.

Zhao et al. [15] have proposed a PSO-based method to maximize the number of clients under the constraint of both latency and bandwidth. The clients should complete the local computation and model upload inside a defined latency. To solve this optimization problem of maximizing clients in limited latency authors have shown PSO can help and give good results on MNIST and Fashion-MNIST datasets.

Torra et al. [16] have discussed a comparison between PSO and FL as optimization problems. The authors discuss that both PSO and FL have different agents that work towards a common optimization goal. In PSO agents share their position without any privacy and authors have proposed a privacy-aware swarm optimization (PAASO) in the Federated Learning context. Results show that privacy does not affect the results but convergence becomes slower.

A comparison of related research works in Federated Learning with their strong points and research gaps is presented in Table 1.

Tian et al. [17] have proposed an improvement of Particle Swarm Optimization using sigmoid-based acceleration coefficients. Authors have used chaotic re-initialization and

TABLE 1. A comparison of previously published literature mentioning briefly their strong points and research gaps.

Author	Strong Points	Research Gaps
McMahan et al. [11]	Authors introduced the parallel decentralized training of clients and termed the concept of Federated Learning.	Authors have used simple averaging of weights at the global model which gives equal weightage to different clients.
T et al. [12]	Authors have subsampled the clients for averaging weights and also considered previous iteration global model weights and obtained quadratic speedup.	Authors have considered a subset of clients in each iteration which may lose some clients' learning.
Ma et al. [13]	Authors have given weightage to each client in the range (0-1) at each hidden layer of the model for averaging. Authors have used similarity between clients to decide the weightage at each layer.	Model accuracy can be compromised if some layers take data from a small percentage of clients but it greatly reduces communication overhead.
Yang et al. [14]	Specific parameters called prompts are prepended at client layers which helps in model personalization in heterogeneous clients.	The number of prompts (K) is determined by experimentally verifying that a very small or large number of prompts can degrade performance and was thus set to 10 for optimal performance.
Zhao et al. [15]	Authors used only those clients that can complete their learning and model upload within defined network bandwidth and latency. PSO helped authors to maximize these clients.	PSO was used for selecting appropriate clients, however it does not play a role in weights averaging.
Torra et al. [16]	Authors have compared PSO and Federated Learning and proposed a privacy-aware combination of two methods and named it PAASO (privacy aware agent swarm optimization).	The proposed method increases privacy but it slows down convergence.
Current Study	This paper proposes a weighted average of clients' weights at the global model of Federated Learning and uses PSO to decide these weight values for each client's averaging share.	This process can enhance the accuracy of CNN models to the state-of-the-art level as compared to using simple averaging

Gaussian mutation which helps to avoid local optima and their approach helps in global convergence to a global optimum. Tian et al. [18] have also proposed a diversity-guided multilevel learning strategy PSO. It also helps to find the global optimum with the help of attractive and repulsive strategies in which the particle's fitness value is compared with the swarm's average fitness value.

As seen authors have used Federated Learning in a wide range of contexts and even used PSO with FL. However, the approach discussed in this research which is the use of PSO for averaging the local clients' weights is not addressed by the authors yet and thus is the main novelty of this article which emphasizes how it can help in a tremendous increase in model performance.

The main contributions of this research article are as mentioned below:

- This research explores the weighted average of weights in Federated Learning
- Particle Swarm Optimization has helped tremendously increase model performance on various Deep Learning Federated models.
- The proposed approach showed increase in around 26% accuracy on CIFAR-10 and CIFAR-100 datasets using VGG16.
- The proposed approach also showed promise in leaf disease detection and fake news classification models.

As mentioned in bulleted contributions above the main innovation in this research is to solve the problem of weights averaging in Federated Learning by giving different preferences to different clients. The clients' weight values which contribute more positively to the global model are given more weightage as compared to clients that have less positive impact. This may be because different local datasets may have different quality of data. Thus this problem is tackled using nature inspired method of Particle Swarm Optimization. It helps to find the amount of weightage of each client by maximizing the test accuracy as the fitness function. On executing these experiments on several models over several datasets it was found that it had a substantial positive impact on the performance of the final trained model.

The rest of the article is organized in the following sections: section III describes a brief of PSO and its implementation with Federated Learning. Section IV gives details of the results of the proposed approach on several datasets and models. Finally, section V discusses the reason for getting good results with the proposed approach and section VI provides an overview of the paper's findings and discusses its future direction.

III. METHODOLOGY

The following section describes the Federated learning for image classification pre-trained models and how their performance can be boosted using Particle Swarm Optimization:

A. FEDERATED LEARNING

The proposed method implements Federated Learning for the classification of standard datasets such as CIFAR-10, and CIFAR-100 on several pre-trained models namely: AlexNet, VGG16, InceptionV3, and ResNet50. In this process dataset is initially divided into train and test in a 90:10 ratio. In the next step the training set is subdivided into 5, 10, or 15 subsets randomly based on the number of clients that are 5, 10, or 15 or any other value. Each subset of the training dataset is assigned to each client separately and the local client starts training on its assigned subset. It trains for a small number of epochs such as 5 to 10. After this, all clients' weights are taken and averaged, and loaded on a global model. This process is called the communication round. The global model can evaluate the average weights on the test set and pass these average weights to all clients that load these weights and further train on their training subset. this process continues for around 100 communication rounds and final test accuracy is recorded for the global model and its weights can be saved for future classification of similar data.

B. FEDERATED LEARNING USING PARTICLE SWARM OPTIMIZATION

1) PARTICLE SWARM OPTIMIZATION

Particle Swarm Optimization (PSO) is a meta-heuristic method to find the near-optimal solution of a problem based on a fitness function. It is derived from a flock of birds to find the best possible position to get food. In this process, a population of particles position are randomly initialized with real numbers between a lower and upper bound. The position of a particle changes based on its velocity. Velocity is derived from the local best and global best position of the population. After a few iterations the fitness value of particles will saturate and give us the best possible solution for any optimization problem.

This approach helped to find the best position vector of parameters starting with default values. PSO is governed by the given eq. 1 and 2.

$$x^i(t+1) = x^i(t) + v^i(t+1) \tag{1}$$

$$v^i(t+1) = w \times v^i(t) + c_1 \times r_1 \times (p^i_{best} - x^i(t)) + c_2 \times r_2 \times (g_{best} - x^i(t)) \tag{2}$$

Here $x^i(t)$ is the position of the i^{th} particle in t^{th} iteration. $v^i(t)$ is particles velocity, p^i_{best} is best position for i^{th} particle, g_{best} is best swarm position. c_1, c_2 are cognitive and social parameters and taken as 1 and 2 for experimental purposes. r_1, r_2 are random numbers between 0 and 1. w_k is the inertia weight introduced by Shi and Eberhart [19] and chosen as 0.5 for experiments.

2) IMPLEMENTATION OF PSO

To implement PSO for optimizing Federated Learning a vector pool is created of vector size equal to the number of clients. The lower bound is chosen as 0 and the upper bound is taken as 1. The PSO algorithm is then integrated with the Federated Learning weight averaging process and vector elements are used to get the weighted average of all clients' weights which is loaded to a global model and tested on a test set. This process repeats till performance saturates and accuracy does not change by less than 0.0001. The averaging of client weights is done as depicted in equation 3 and also in Figure 4. The hyperparameters for PSO and Federated Learning are shown in Table 2. The figure showing the calculation of the next iteration vector in the PSO algorithm is depicted in Figure 3.

$$W^g = \frac{x^i \times W^i}{sum(x^i)} \tag{3}$$

Here W^g is the weighted mean of all client weights loaded to the global model, x^i is the vector element at position i and W^i is the weights of i^{th} client. The initialization of PSO particle position vectors is done as shown in algorithm 1 and the PSO process of finding the best vector is shown in algorithm 2. The process of integrating PSO with Federated learning weighted averaging of weights is shown in algorithm 3. It is also shown in a flow chart form in Figure 4. The 2^{nd} for loop on line 13 is repeated for every PSO vector in the population pool and these iterations are done for a maximum number of PSO iterations say 100 till we get the best set of global weights. The best PSO vector is chosen based on a fitness function which is to maximize the accuracy of the global model in each communication round as shown in equation 4.

$$Maximize\ X = \frac{\sum_{i=1}^{i=N} TP_i}{\sum_{i=1}^{i=N} TP_i + \sum_{i=1}^{i=N} FP_i} \tag{4}$$

Here N is the number of classes for example 10 for CIFAR-10 and 100 for CIFAR-100. Similarly, i is the class number from 1 to N. TP_i is correctly classified samples of class i and FP_i are the misclassifications of samples in class i.

TABLE 2. Hyperparameters for FLPSO.

Parameter	Value
PSO Parameters	
Number of particles	100
Maximum Generations	20
Lower Bound	0
Upper Bound	1
Termination Criteria	Change in fitness < 0.0001
FL Parameters	
Number of FL clients	10
Communication rounds	100
Local clients epochs in 1 round	5
CNN Loss function	categorical_crossentropy
CNN Optimizer	Stochastic Gradient Descent
Learning Rate	0.001

3) SPLITTING OF DATASET

The dataset is initially split into training and testing in a 90:10 ratio. To split the dataset into n clients a random split is done on training data and in 1^{st} step it is split into $1 : (n-1)$

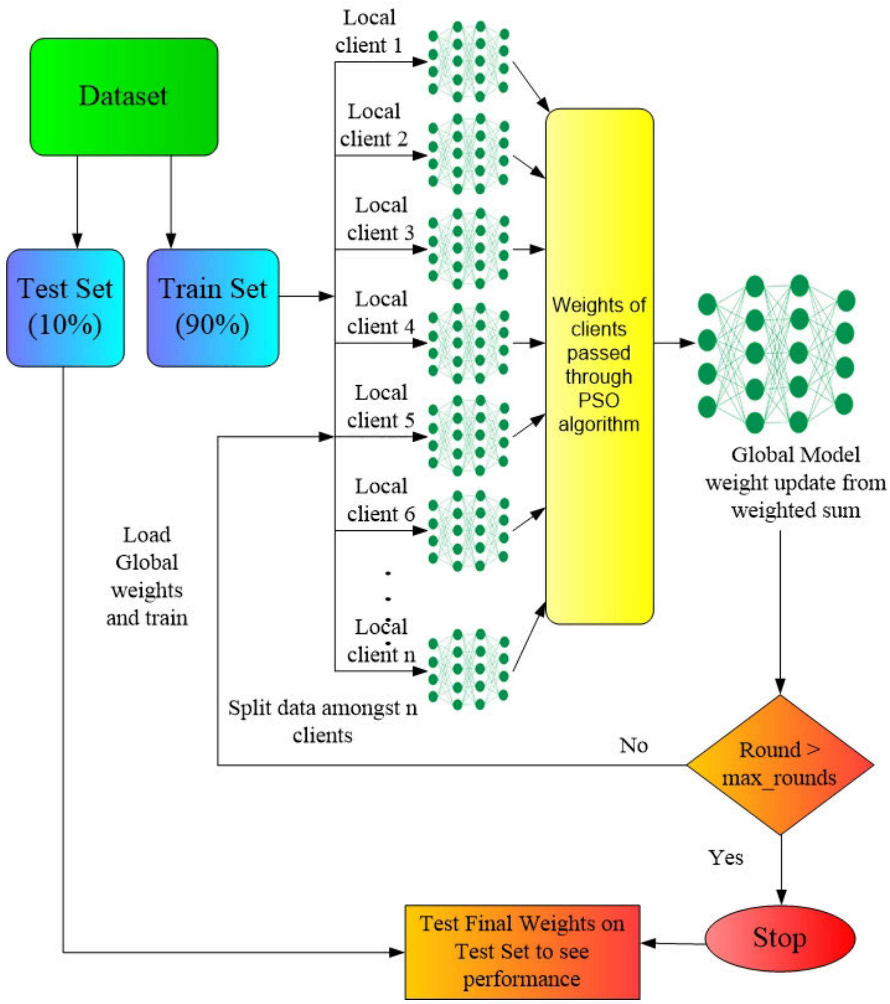


FIGURE 2. Pictorial representation of federated learning with n clients using PSO weighted sum.

x_i	0.23	0.34	0.67	0.12	0.43	0.29	0.22	0.19	0.89	0.87
v_i	0.18	0.11	0.76	0.57	0.24	0.99	0.72	0.76	0.38	0.53
x_best_i	0.19	0.28	0.91	0.38	0.76	0.29	0.38	0.48	0.72	0.62
x_best_g	0.13	0.39	0.81	0.31	0.36	0.67	0.83	0.59	0.39	0.49
c1	1									
c2	2		v_new = 0.5*v_i + c1*r1*(x_best_i - xi) + c2*r2*(x_best_g - xi)							
r1	0.2		x_new = x_i + v_new							
r2	0.6									
v_new	-0.038	0.103	0.596	0.565	0.102	0.951	1.124	0.918	-0.444	-0.241
x_new	0.192	0.443	1.266	0.685	0.532	1.241	1.344	1.108	0.446	0.629
x_new_normalized	0.192	0.443	1	0.685	0.532	1	1	1	0.446	0.629

FIGURE 3. Calculation of next iteration vector in PSO with 10 clients.

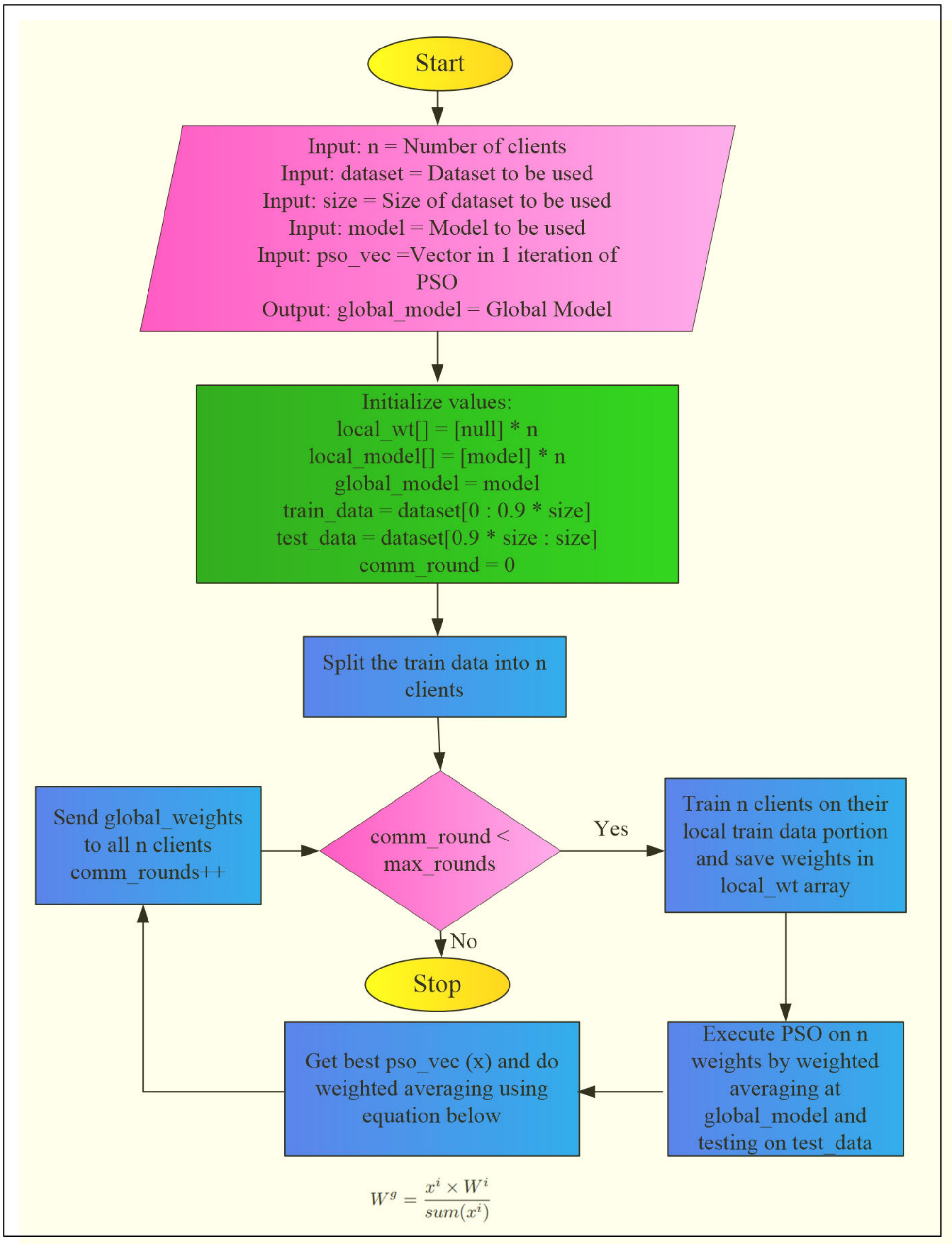


FIGURE 4. Flow chart of the proposed methodology.

ratio. In the next step since we get $(n-1)^{th}$ part as another split which we split it again in $1:(n-2)$ ratio. Similarly, in further steps, we split in ratios $1:(n-3)$, $1:(n-4)$, ..., $1:2$. Like this we get n equal splits of the training dataset to distribute to each n clients.

IV. EXPERIMENTAL RESULTS

All the training experiments using PSO weighted averaging Federated Learning were performed on NVIDIA DGX V100 equipped with 40,600 CUDA cores, 5120 tensor cores, 128 GB of RAM, and an operational speed of 1000 TFLOPS. The testing was done on HP Z440 Workstation with "Intel (R) Xeon (R) CPU ES-1650 v4 @3.6 GHz" processor and 16 GB RAM with x-64 based processor.

The performance of four pre-trained models: AlexNet, VGG16, InceptionV3, and ResNet50 were recorded using the proposed approach using two standard datasets namely: CIFAR-10 and CIFAR-100. The experiments were performed using the train test split of 90:10 and subsequently with 70:30 also to see their comparison. It was found that 90:10 split gave better performance. The reason for better performance with the 90:10 split was that the model was able to generalize more nicely with more training data. In the 70:30 split the training data was less and the testing data was more which caused

Algorithm 1 Generation of Particle Position Vectors

Input: k ▷ Number of clients
Output: A ▷ Matrix of list of vectors

procedure InitParticles(k)
  # Create Initial particle position vectors of PSO
  # Initialize set of initial position vectors population to empty list
  $A \leftarrow []$;
  # Generate say 100 particles each of length k
  int $B[100][k]$
  for $i \leftarrow 0$ *to* 100 **do**
    # Generate random vectors of length k with random values between 0 and 1
    for $j \leftarrow 0$ *to* k **do**
      $B[i][j] = random(0, 1)$
      # random function generates a random positive value between 0 and 1
    $A \leftarrow A \cup B[i]$
  return A

TABLE 3. Comparison statistics of performance on CIFAR-10 dataset.

Model	Plain FL Accuracy (%)	PSO optimized FL Accuracy (%)	
		Train: Test 90:10	Train: Test 70:30
AlexNet	68.31	87.14	83.19
VGG16	66.05	92.06	88.79
InceptionV3	73.45	89.23	86.27
ResNet50	77.94	82.23	79.23

some samples to be wrongly classified for which training could not be sufficiently done in lesser data. These results are discussed in the following subsections:

A. PERFORMANCE ON CIFAR-10

The CIFAR-10 dataset comprises of images from 10 different classes: airplane, automobile, bird, cat, deer, etc. There are 60,000 images of size 32×32. The performance of models on this dataset is described in Table 3. As can be seen from the comparison PSO-based Federated Learning could enhance the performance of the model tremendously. This was caused since the most important weights were selected from clients' learning. The Receiver Operating Characteristic (ROC) Curve provides a good analysis of the model performance, hence a comparison of 4 pre-trained models on the CIFAR-10 dataset using Federated Learning with PSO is shown in Figure 5 using ROC curves.

B. PERFORMANCE ON CIFAR-100

The CIFAR-10 dataset is comprised of images from 100 different classes such as apple, bear, bridge, bed, camel, fox, crab, etc. There are 500 training images in each class of size 32×32 and 100 images in each class in the testing set. The performance of models on this dataset is described in Table 4. As can be seen from the comparison PSO-based Federated Learning could enhance the performance of the model tremendously for the CIFAR-100 dataset also. A comparison of 4 pre-trained models' performance on the CIFAR-100 dataset using Federated Learning with PSO is shown in Figure 6 using ROC curves.

Algorithm 2 Particle Swarm Optimization

Input: c_1 ▷ Social constant
Input: c_2 ▷ Cognitive constant
Input: Max_iter ▷ Maximum Iterations
Input: Num_particles ▷ Number of particles
Output: b^g_{best} ▷ **Best global Particle**

procedure PSO(c_1, c_2, Max_iter, Num_particles)
  # To find the best position vector satisfying fitness criteria
  $x[\,] =$ **INITPARTICLES**()
  # Set Global best to 0
  $f^g_{best} \leftarrow 0$
  # Initialize global best particle to null
  $b^g_{best} \leftarrow []$
  for i *in range*$(0, length(x))$ **do**
    # Set each particle best to 0
    $f^i_{best} \leftarrow 0$
    # Initialize best particle position to null
    $b^i_{best} \leftarrow []$
    # Initialize particle velocity to 0
    $v^i(0) \leftarrow 0$
  for t *in range*$(0, Max\_iter)$ **do**
    for i *in range*$(0, Num\_particles)$ **do**
      # Check fitness of each particle
      if $fitness(x^i(t)) > f^g_{best}$ **then**
        $f^g_{best} \leftarrow fitness(x^i(t))$
        $b^g_{best} \leftarrow x^i(t)$
      if $fitness(x^i(t)) > f^i_{best}$ **then**
        $f^i_{best} \leftarrow fitness(x^i(t))$
        $b^i_{best} \leftarrow x^i(t)$
      # Generate two random numbers between 0 and 1: r_1, r_2
      # Compute the velocity and position vectors of the i^{th} particle for the next iteration of PSO.
      $v^i(t+1) = w \times v^i(t) + c_1 \times r_1 \times (p^i_{best} - x^i(t)) + c_2 \times r_2 \times (g_{best} - x^i(t))$
      $x^i(t+1) = x^i(t) + v^i(t+1)$
  return b^g_{best}

TABLE 4. Comparison statistics of performance on CIFAR-100 dataset.

Model	Plain FL Accuracy (%)	PSO optimized FL Accuracy (%)	
		Train: Test 90:10	Train: Test 70:30
AlexNet	40.28	50.42	47.56
VGG16	43.28	70.12	68.73
InceptionV3	41.39	68.25	66.35
ResNet50	32.93	56.35	53.83

Algorithm 3 Federated Learning Using PSO

Input: n ▷ Number of clients
Input: dataset ▷ Dataset to used
Input: size ▷ Size of dataset to used
Input: model ▷ Model to used
Input: pso_vec ▷ Vector in 1 iteration of PSO
Output: global_model ▷ **Global Model in 1 communication round**

procedure FLPSO
  $data[] = [null] * n$
  $local\_wt[] = [null] * n$
  $local\_model[] = [model] * n$
  $global\_model = model$
  $train\_data = dataset[0 : 0.9 * size]$
  $test\_data = dataset[0.9 * size : size]$
  $size = size * 0.9$
  # Split the dataset into n sets
  for $i \leftarrow 0$ *to* n **do**
    $data[i] = train\_data[i * size/n : (i + 1) * size/n]$

  $global\_wt \leftarrow 0$
  # Train i^{th} model on i^{th} spit of dataset
  for $i \leftarrow 0$ *to* n **do**
    $model[i].train(data[i])$
    $local\_wt[i] \leftarrow get\_weights(model[i])$
  $pso\_vec =$ **PSO**()
  # Multiply i^{th} model weights with i^{th} element of PSO vector
  for $i \leftarrow 0$ *to* n **do**
    $local\_wt[i] \leftarrow local\_wt[i] \times pso\_vec[i]$
    $global\_wt \leftarrow global\_wt + local\_wt[i]$ ▷ Do a weighted average of weights

  $global\_wt \leftarrow global\_wt/sum(pso\_vec)$
  $global\_model.set\_weights(global\_wt)$
  return global_model

The graph of accuracy increase with communication rounds for VGG16 using CIFAR-10 and CIFAR-100 datasets is shown in Figure 7.

C. PERFORMANCE OF AlexNet ON TOMATO DATASET

Since the PSO-based Federated learning showed promising results with CIFAR-10 and CIFAR-100 datasets, an attempt was made to enhance the performance of AlexNet on tomato classes from PlantVillage [7] dataset. The comparison of results with plain Federated Learning and PSO-based Federated Learning with different numbers of clients are provided in Table 5. As clearly found Federated Learning was able to enhance the accuracy of plain AlexNet but PSO-based Federated Learning with ten clients gave the maximum performance known so far on this dataset. The performance deteriorated with decreasing the number of clients and was low with only two clients.

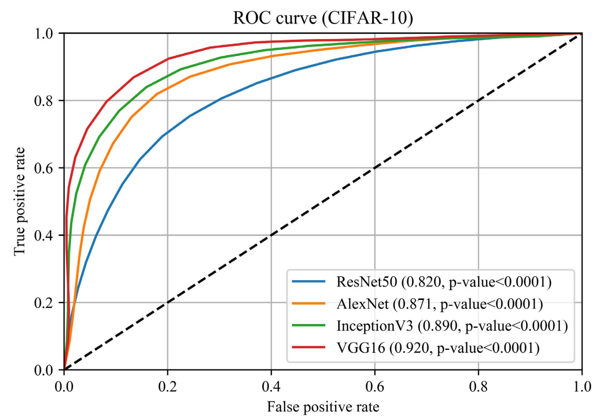


FIGURE 5. ROC curve of 4 pre-trained models using FL and PSO on CIFAR-10 dataset.

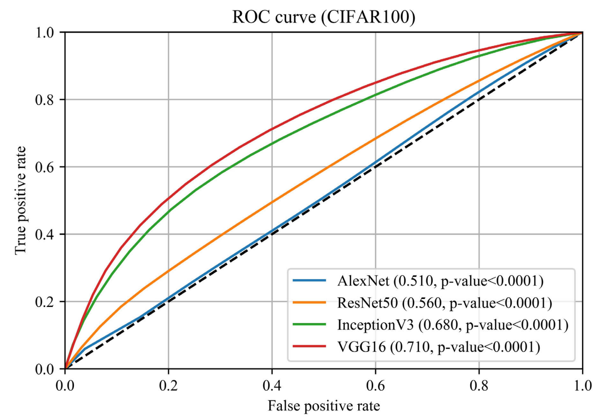


FIGURE 6. ROC curve of 4 pre-trained models using FL and PSO on CIFAR-100 dataset.

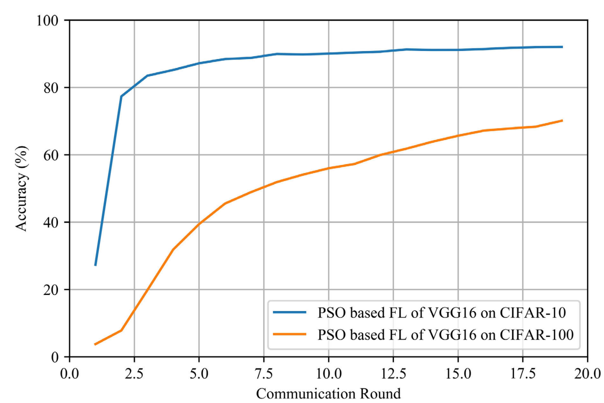


FIGURE 7. Graph showing an increase in test accuracy with FL communication rounds using PSO for VGG16.

The graph of accuracy increase with communication rounds for AlexNet using the Tomato dataset for different numbers of clients is shown in Figure 8. The confusion matrix of final weights obtained from Federated Learning using PSO is shown in Figure 9. ROC curve showing performance of

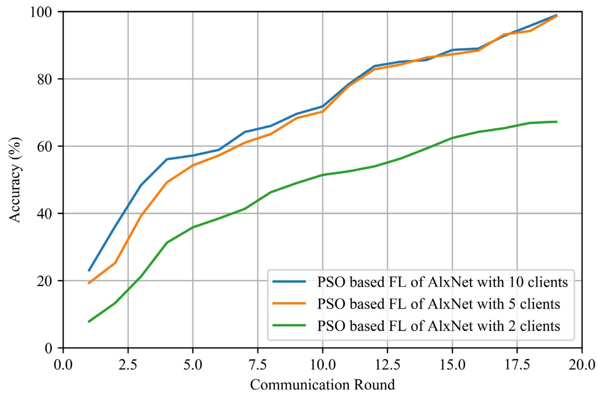


FIGURE 8. Graph showing an increase in test accuracy with FL communication rounds using PSO for AlexNet using tomato dataset.

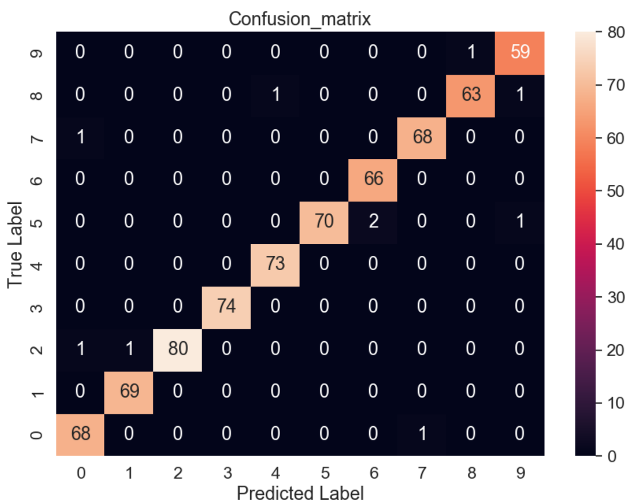


FIGURE 9. Confusion matrix from the final trained global model after federated learning using PSO with ten clients.

TABLE 5. Comparison statistics of AlexNet performance on tomato dataset.

Model	Accuracy (%)	
	Train: Test 90:10	Train: Test 70:30
Plain AlexNet	70.34	67.29
Plain FL	90.9	87.46
Ten clients PSO+FL	98.9	96.72
Five clients PSO+FL	98.6	85.68
Two clients PSO+FL	67.23	65.38

AlexNet on Tomato dataset without Federated Learning and using Federated Learning with PSO on 2, 5, and 10 clients in shown in Figure 10.

The activation images of a sample tomato leaf images for 2, 5, and 10 clients are shown in Figure 11 - 16.

D. PERFORMANCE OF MULTIMODAL CNN MODEL ON FAKE NEWS DATASET

In a further step, the PSO-based Federated Learning was used to enhance the accuracy of the multimodal model with dual input of text and images. The model had a text Embedding layer along with a CNN model with 3 convolutional layers. On applying PSO-based Federated Learning the results were found encouraging with performance jumping from 69.99% accuracy to 78.2%. The experiments when repeated with train test ratio as 70:30 then original and FLPSO accuracy noticed

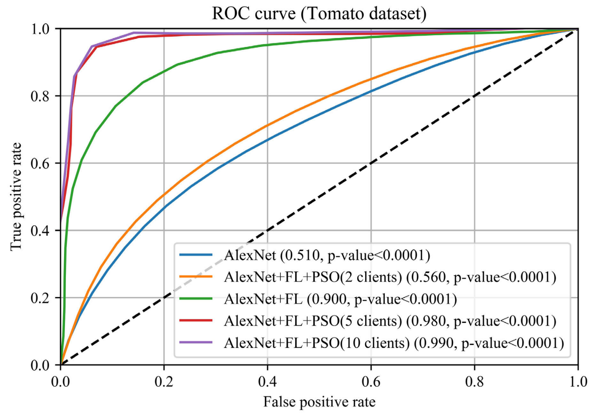


FIGURE 10. ROC curve of AlexNet using FL and PSO on tomato dataset with a different number of clients.

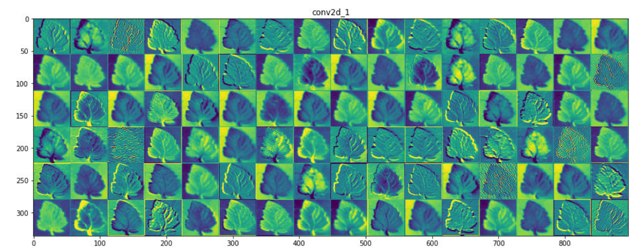


FIGURE 11. Figure showing activation images in 1st convolution layer of PSO-based FL on AlexNet with 10 clients.

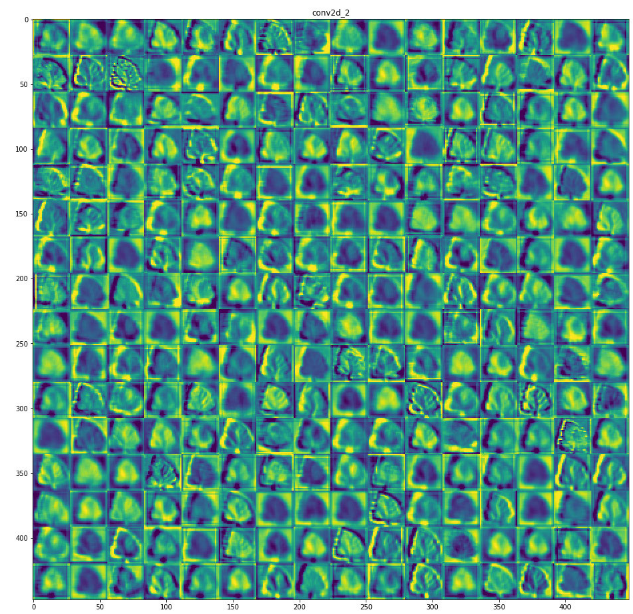


FIGURE 12. Figure showing activation images in 2nd convolution layer of PSO-based FL on AlexNet with 10 clients.

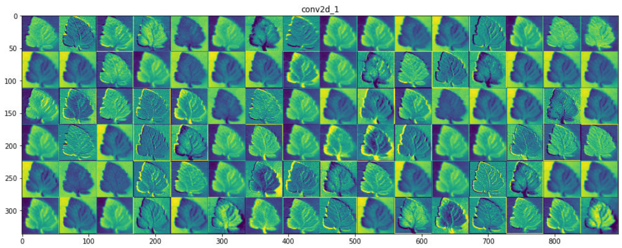


FIGURE 13. Figure showing activation images in 1^{st} convolution layer of PSO-based FL on AlexNet with 5 clients.

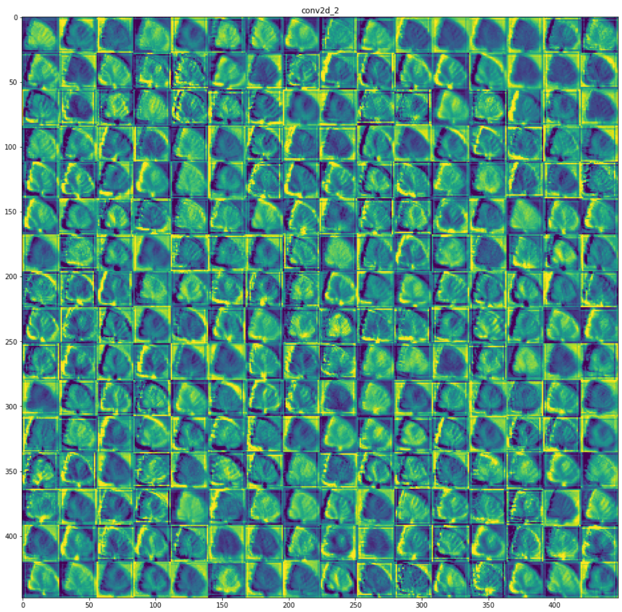


FIGURE 14. Figure showing activation images in 2^{nd} convolution layer of PSO-based FL on AlexNet with 5 clients.

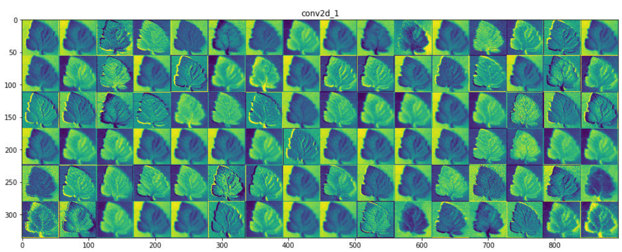


FIGURE 15. Figure showing activation images in 1^{st} convolution layer of PSO-based FL on AlexNet with 2 clients.

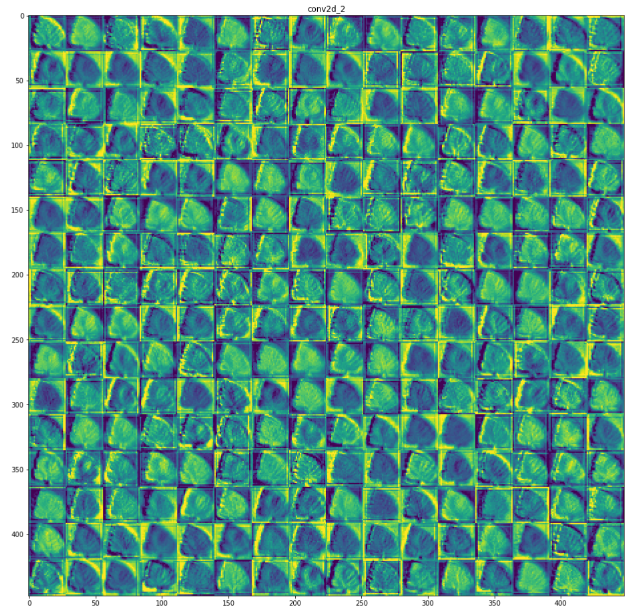


FIGURE 16. Figure showing activation images in 2^{nd} convolution layer of PSO-based FL on AlexNet with 2 clients.

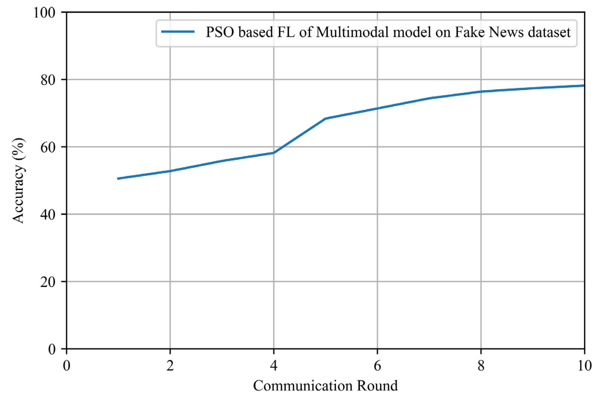


FIGURE 17. Graph showing an increase in test accuracy with FL communication rounds using PSO for Multimodal model on Fake News dataset.

were: 65.83% and 76.39%. Thus it was found with a high degree of confidence that PSO-based Federated Learning was a reliable model to enhance the performance of most of the deep learning models. The graph of accuracy for the proposed approach on the Fake News dataset [8] is shown in Figure 17. The block-level architecture diagram of the multimodal model is also shown in Figure 18.

E. BENCHMARKING WITH EXISTING RESEARCH

This section gives a comparison of existing research work on Federated Learning with the proposed methodology. The comparison can be seen in Table 6. As clearly seen from the benchmarking table the results obtained by the proposed research could surpass almost all the cited research works in the last 2-3 years.

V. DISCUSSION

On analyzing the reason of the better performance of FL using PSO it can be argued that some local models may be contributing in a better way than other local clients and thus simple averaging of weights will not boost the accuracy due to equal contribution from all local models. However, PSO helps in finding which model can enhance the accuracy by a larger degree and thus gives more weightage to those clients and less to the clients that are not aiding in boosting the model performance. It also helps by reducing the communication rounds as in simple averaging the negatively contributing clients will need more time to be trained so that the overall accuracy of the FL model is good.

FIGURE 18. Architecture diagram of multimodal model.

TABLE 6. Benchmarking of proposed research performance with prior research works.

Method	DL model	Dataset	Accuracy (%)
Federated-Communication Efficient, Shah and Lau [20]	VGG16	CIFAR-10	86.16
FedNAG, Yang et al. [21]	VGG11	CIFAR-10	$\sim$ 89.5 (From Graph)
Federated-DOMO, Xu and Huang [22]	VGG16	CIFAR-10	85.54
Federated-AdaCFL, Gong et al. [23]	VGG16	CIFAR-10	86.1
Blockchain-based federated, Kalapaaking et al. [24]	VGG16	CIFAR-10	87.4
FedGB, Yang and Sun [25]	VGG16	CIFAR-10	87.79
This Study	VGG16	CIFAR-10	**92.06**
Federated-DOMO, Xu and Huang [22]	VGG16	CIFAR-100	62.47
FedNAG, Yang et al. [21]	VGG11	CIFAR-100	$\sim$ 66.5 (From Graph)
This Study	VGG16	CIFAR-100	**70.12**

As seen from the performance of four pre-trained models and a multi-modal deep learning model, this approach can increase the performance by a considerable amount.

VI. CONCLUSION

In this research article it was found that if any problem can be formulated in terms of a meta-heuristics-based optimization problem then we will surely benefit from this approach. For example, in this research the problem was noticed that all clients in Federated Learning may not contribute equally to the optimum global weights. Hence it was felt we can do weighted averaging with fitness function being the test accuracy. Since it needed continuous real-values for doing weighted averaging we needed a continuous style nature-inspired algorithm. Since the Genetic Algorithm works on binary values 0-1, hence we chose PSO so that we can optimize in continuous real values also. In future we can try more such real-valued optimization algorithms like Differential Evolution, Whale Optimization, etc.

This research article will bring a new dimension to Federated Learning used in deep learning models. The integration of PSO with FL for the weighted average of local clients's weights can help to enhance the model performance tremendously as evident from four pretrained models on CIFAR-10 and CIFAR-100 datasets. The validation of these results on tomato and multimodel fake news datasets confirmed that it is a sure way of bettering the model performance by a considerable amount. A comparison of VGG16 model accuracy on CIFAR-10 and CIFAR-100 with existing research works using Federated Learning shows that the proposed approach gives a huge promise in the field of Federated Learning. Future research will aim to show proposed approach can also be used on image segmentation models such as FCN, UNet, SegNet, etc.

REFERENCES

[1] N. Rieke, J. Hancox, W. Li, F. Milletarì, H. R. Roth, S. Albarqouni, S. Bakas, M. N. Galtier, B. A. Landman, K. Maier-Hein, S. Ourselin, M. Sheller, R. M. Summers, A. Trask, D. Xu, M. Baust, and M. J. Cardoso, "The future of digital health with federated learning," *npj Digit. Med.*, vol. 3, no. 1, p. 119, Sep. 2020.

[2] L. Li, Y. Fan, M. Tse, and K.-Y. Lin, "A review of applications in federated learning," *Comput. Ind. Eng.*, vol. 149, Nov. 2020, Art. no. 106854.

[3] C. Zhang, Y. Xie, H. Bai, B. Yu, W. Li, and Y. Gao, "A survey on federated learning," *Knowl.-Based Syst.*, vol. 216, Mar. 2021, Art. no. 106775.

[4] Q. Li, B. He, and D. Song, "Model-contrastive federated learning," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 10708–10717.

[5] S. Savazzi, M. Nicoli, and V. Rampa, "Federated learning with cooperating devices: A consensus approach for massive IoT networks," *IEEE Internet Things J.*, vol. 7, no. 5, pp. 4641–4654, May 2020.

[6] A. Krizhevsky, V. Nair, and G. Hinton. (2017). *The CIFAR-10 Dataset (2014)*. [Online]. Available: https://www.cs.toronto.edu/kriz/cifar.html

[7] D. P. Hughes and M. Salathe, "An open access repository of images on plant health to enable the development of mobile disease diagnostics," 2015, *arXiv:1511.08060*.

[8] K. Nakamura, S. Levy, and W. Yang Wang, "R/fakeddit: A new multimodal benchmark dataset for fine-grained fake news detection," 2019, *arXiv:1911.03854*.

[9] R. E. J. Kennedy, "Particle swarm optimization," in *Proc. Int. Conf. Neural Netw.*, Perth, WA, Australia, Nov. 1995, pp. 1942–1948.

[10] K. Jiao, J. Chen, B. Xin, L. Li, Y. Zheng, and Z. Zhao, "Three-dimensional path planning with enhanced gravitational search algorithm for unmanned aerial vehicle," *Robotica*, vol. 42, no. 7, pp. 2453–2487, Jul. 2024.

[11] B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. Y. Arcas, "Communication-efficient learning of deep networks from decentralized data," in *Proc. 20th Int. Conf. Artif. Intell. Statist.*, vol. 54, A. Singh and J. Zhu, Eds., Fort Lauderdale, FL, USA, 2017, pp. 1273–1282.

[12] C. T. Dinh, N. H. Tran, and T. D. Nguyen, "Personalized federated learning with Moreau envelopes," in *Proc. NIPS*, Dec. 2020, pp. 21394–21405.

[13] X. Ma, J. Zhang, S. Guo, and W. Xu, "Layer-wised model aggregation for personalized federated learning," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 10082–10091.

[14] F.-E. Yang, C.-Y. Wang, and Y.-C.-F. Wang, "Efficient model personalization in federated learning via client-specific prompt generation," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2023, pp. 19159–19168.

[15] Z. Zhao, J. Xia, L. Fan, X. Lei, G. K. Karagiannidis, and A. Nallanathan, "System optimization of federated learning networks with a constrained latency," *IEEE Trans. Veh. Technol.*, vol. 71, no. 1, pp. 1095–1100, Jan. 2022.

[16] V. Torra, E. Galván, and G. Navarro-Arribas, "PSO + FL = PAASO: Particle swarm optimization + federated learning = privacy-aware agent swarm optimization," *Int. J. Inf. Secur.*, vol. 21, no. 6, pp. 1349–1359, Dec. 2022.

[17] D. Tian, X. Zhao, and Z. Shi, "Chaotic particle swarm optimization with sigmoid-based acceleration coefficients for numerical function optimization," *Swarm Evol. Comput.*, vol. 51, Dec. 2019, Art. no. 100573.

[18] D. Tian, Q. Xu, X. Yao, G. Zhang, Y. Li, and C. Xu, "Diversity-guided particle swarm optimization with multi-level learning strategy," *Swarm Evol. Comput.*, vol. 86, Apr. 2024, Art. no. 101533.

[19] Y. Shi and R. Eberhart, "A modified particle swarm optimizer," in *Proc. IEEE Int. Conf. Evol. Comput. IEEE World Congr. Comput. Intell.*, May 1998, pp. 69–73.

[20] S. M. Shah and V. K. N. Lau, "Model compression for communication efficient federated learning," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 34, no. 9, pp. 1–15, Dec. 2021.

[21] Z. Yang, M. Chen, K.-K. Wong, H. V. Poor, and S. Cui, "Federated learning for 6G: Applications, challenges, and opportunities," *Engineering*, vol. 8, pp. 33–41, Jan. 2022.

[22] A. Xu and H. Huang, "Coordinating momenta for cross-silo federated learning," in *Proc. AAAI Conf. Artif. Intell.*, vol. 36, 2022, pp. 8735–8743.

[23] B. Gong, T. Xing, Z. Liu, J. Wang, and X. Liu, "Adaptive clustered federated learning for heterogeneous data in edge computing," *Mobile Netw. Appl.*, vol. 27, no. 4, pp. 1520–1530, Aug. 2022.

[24] A. P. Kalapaaking, I. Khalil, M. S. Rahman, M. Atiquzzaman, X. Yi, and M. Almashor, "Blockchain-based federated learning with secure aggregation in trusted execution environment for Internet-of-Things," *IEEE Trans. Ind. Informat.*, vol. 19, no. 2, pp. 1703–1714, Feb. 2023.

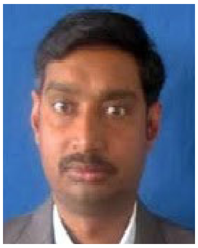
[25] Z. Yang and Q. Sun, "A dynamic global backbone updating for communication-efficient personalised federated learning," *Connection Sci.*, vol. 34, no. 1, pp. 2240–2264, Dec. 2022.

MOHIT AGARWAL (Member, IEEE) received the B.Tech. degree in CSE from IIT Delhi, in 1995, the M.Tech. degree from ABES-EC, Ghaziabad, in 2016, and the Ph.D. degree from Bennett University, Greater Noida, in 2021. He has vast experience in MNCs in India and the USA. He is currently an Associate Professor with Bennett University. He has good programming skills in Python and C#.NET. He has published more than 54 papers in SCIE and Scopus-indexed journals and conferences and has more than 1250 citations with an H-index of 17 and an i10-index of 26. His research interests include machine learning, deep learning, image processing, and model compression.

GARIMA JAISWAL (Member, IEEE) received the Distinguished Ph.D. degree from Indira Gandhi Delhi Technical University for Women, Delhi. With over a decade of experience, she is a recognized expert in the fields of document forgery detection, hyperspectral imaging, and machine learning. Her profound knowledge and practical skills have led her to be a trailblazer in the application of hyperspectral imaging for document analysis and revealing imperceptible alterations. In addition, her proficiency in machine learning empowers her to develop advanced algorithms and models, enhancing the accuracy, and efficiency of models. Her passion for these domains and academic rigor have firmly established her as a renowned authority, contributing significantly to the enhancement of document security and authenticity. With an impressive track record, she has published more than 30 research papers in prestigious SCI, SCIE, and Scopus-indexed journals and conferences.

ROHIT KUMAR KALIYAR (Senior Member, IEEE) is currently an Assistant Professor with the School of Computer Science Engineering and Technology, Bennett University, Greater Noida, India. Formerly, he was a Team Lead-Senior Software Engineer with HCL Technologies, from June 2015 to June 2017. Previously, he was a Research Assistant with Shiv Nadar University, Greater Noida, where he was actively involved in the development of the first successful big data analytics center with IBM and the health systems with Apollo Hospital. His research interests include fake news detection, rumor detection, and the analysis of abusive content and sentiment analysis for fake content. He is currently developing a real-time fake news tracking system using deep learning techniques.

AKANSHA SINGH received the B.Tech., M.Tech., and Ph.D. degrees in computer science, and the Ph.D. degree in image processing and machine learning from IIT Roorkee. Currently, she is an Associate Professor with the School of Computer Science and Engineering, Bennett University, Greater Noida, India. She has also undertaken government-funded projects as the principal investigator. Her research interests include image processing, remote sensing, the IoT, and machine learning. She served as an associate editor and a guest editor for several journals.



KRISHNA KANT SINGH received the B.Tech., M.Tech., M.S., and Ph.D. degrees in image processing and machine learning from IIT Roorkee. He has wide teaching and research experience. He is currently the Director of the Delhi Technical Campus, Greater Noida, India. He has authored more than 140 research articles in Scopus and SCIE-indexed journals of repute. He has also authored 25 technical books. He is serving as a member of the editorial board for *Applied Computing and Geoscience* (Elsevier). He is an Associate Editor of *Journal of Intelligent and Fuzzy Systems* (SCIE Indexed) and IEEE Access (SCIE Indexed) and a Guest Editor of *Open Computer Science* and *Wireless Personal Communications*.

S. S. ASKAR received the B.Sc. degree in mathematics and the M.Sc. degree in applied mathematics from Mansoura University, Egypt, in 1998 and 2004, respectively, and the Ph.D. degree in operation research from Cranfield University, U.K., in 2011. He has been an Associate Professor with Mansoura University, Egypt, since 2016. He joined King Saud University, in 2012, where he has been a Professor with the Department of Statistics and Operation Research. His research interests include game theory and its applications, including mathematical economy, dynamical systems, and network analysis.

MOHAMED ABOUHAWWASH received the B.Sc. and M.Sc. degrees in statistics and computer science from Mansoura University, Mansoura, Egypt, in 2005 and 2011, respectively, and the joint Ph.D. degree in statistics and computer science from Michigan State University, USA, and Mansoura University, Egypt, in 2015. He is currently an Associate Professor with the Department of Mathematics, Faculty of Science, Mansoura University, Egypt. He was a recipient of the Best Master's and Ph.D. Thesis Awards from Mansoura University, in 2012 and 2018, respectively.