IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 10 June 2023, accepted 27 July 2023, date of publication 31 July 2023, date of current version 8 August 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3300221

RESEARCH ARTICLE

Metaheuristics Algorithm-Based Minimization of Communication Costs in Federated Learning

MOHAMED AHMED ELFAKI[1], HAYA MESFER ALSHAHRANI[2], KHALID MAHMOOD[3], RANA ALABDAN[4], MOFADAL ALYMANI[5], (Member, IEEE), HUSSAIN ALSHAHRANI[1], ABDELWAHED MOTWAKEL[6], AND AMANI A. ALNEIL[7]

[1]Department of Computer Science, College of Computing and Information Technology, Shaqra University, Shaqra 11911, Saudi Arabia
[2]Department of Information Systems, College of Computer and Information Sciences, Princess Nourah bint Abdulrahman University, Riyadh 11671, Saudi Arabia
[3]Department of Information Systems, College of Science and Art at Mahayil, King Khalid University, Abha 62529, Saudi Arabia
[4]Department of Information Systems, College of Computer and Information Science, Majmaah University, Al Majma'ah 11952, Saudi Arabia
[5]Department of Computer Engineering, College of Computing and Information Technology, Shaqra University, Shaqra 11911, Saudi Arabia
[6]Department of Information Systems, College of Business Administration in Hawtat bani Tamim, Prince Sattam bin Abdulaziz University, Al-Kharj 16278, Saudi Arabia
[7]Department of Computer and Self Development, Preparatory Year Deanship, Prince Sattam bin Abdulaziz University, Al-Kharj 16278, Saudi Arabia

Corresponding author: Khalid Mahmood (kasgr@kku.edu.sa)

The authors extend their appreciation to the Deanship of Scientific Research at King Khalid University for funding this work through Large Groups Project under grant number (RGP2/112/43). Princess Nourah bint Abdulrahman University Researchers Supporting Project number (PNURSP2023R237), Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia. The Authors would like to thank the Deanship of Scientific Research at Majmaah University for supporting this work under Project No. R-2023-526. This study is supported via funding from Prince Sattam bin Abdulaziz University project number (PSAU/2023/R/1444).

ABSTRACT The Federated learning (FL) technique resolves the issue of training machine learning (ML) techniques on distributed networks, including the huge volume of modern smart devices. FL clients frequently use Wi-Fi and have to interact in unstable network surroundings. However, as the present FL aggregation approaches receive and send a large number of weights, accuracy can be decreased considerably in unstable network surroundings. Therefore, this study presents a Quantum with Metaheuristics Algorithm Based Minimization of Communication Costs in Federated Learning (QMAMCC-FL) technique. The presented QMAMCC-FL technique is designed a federated hybrid convolutional neural network with a gated recurrent unit (HCNN-GRU) model with a quantum Aquila optimization (QAO) algorithm. The QMAMCC-FL technique upgrades the global model via weight collection of the learned model, which is commonly used in FL. The proposed model can be employed to increase the performance of network communication and reduce the size of data transmitted from clients to servers such as smartphones and tablets. The experimental analysis of the QMAMCC-FL approach is tested, and the outcomes show better performance over other existing models.

INDEX TERMS Metaheuristics, federated learning, communication cost, quantum computing, deep learning.

I. INTRODUCTION

In classical Machine Learning (ML) techniques, data were stored and collected by a centralized server or one node and then utilized for testing and training [1], [2]. The accuracy and efficiency of the methods rely upon various aspects, namely the amount of training data, computational power, quality etc. [3]. However, new techniques may need the management and acquisition of a huge volume of data that, in some fields, such as healthcare, may undergo trouble by adopting centralized structures. Centralizing and transmitting information certainly increases several legal, administrative, and ethical problems [4], [5], predominantly relevant to data security and privacy, per the General Data Protection Regulation (GDPR). Federated Learning (FL) has developed a technological solution to overcome these challenges [6], [7]. FL authorizes collaborative learning by not centralizing data. It is a new ML technique intended to solve the problem of data island

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

whereas protecting data security. It includes many nodes or clients, like organizations, and mobile devices, coordinated with more than one centralized server for decentralized ML settings [8]. In the FL technique, all nodes describe their local method related to local information and transfer it to the centralized server. The centralized server aggregates every local model and describes one global method [9]. In FL, knowledge can be interchanged by transferring the methods rather than the data of users in such a way it is not vulnerable to data breaches on the network.

Even though the FL has various advantages, communication efficiency remains the main concern [10], [11]. Various rounds of transmission among the FL server and the participants are needed to attain some target accuracy [12]. This indicates that for all rounds, a vast amount of data, GigaBytes, can spread through the network. The high dimensions of the updates may lead to a training bottleneck and lead extraordinary transmission costs [13], [14]. Currently, authors are making great endeavours to enrich the transmission efficacy of FL. However, they generally suffer from heavy performance sacrifice if the compression ratios are largely demanded [15].

This study presents a Quantum with Metaheuristics Algorithm Based Minimization of Communication Costs in Federated Learning (QMAMCC-FL) technique. The presented QMAMCC-FL technique is designed a federated hybrid convolutional neural network with a gated recurrent unit (HCNN-GRU) model using a quantum Aquila optimization (QAO) algorithm. The QMAMCC-FL technique upgrades the global model via weight collection of the learned model, which is widely utilized in FL. The QMAMCC-FL technique attains increased robustness in unstable networks with enhanced communication efficiency. Moreover, the amount of data exploited during network communication is decreased, and the performance is improved. The experimental analysis of the QMAMCC-FL method is tested under various aspects.

II. RELATED WORKS

Jain and Sharma [16] introduced a technique called Improved Quantum Salp Swarm Algorithm (IQSSA) that enriches the SSA by including the principles of Quantum computing (QC) for increasing convergence speed. Moreover, Quantum-inspired Salp Swarm GWA (QSSGWA) entrenches SSA with GWO to enrich global besteyhw solutions, and quantum operators are utilized for initializing the population. The presented techniques accomplish tasks with budget constraints and user-defined deadlines. Moreover, the penalty costs are implemented and developed in the case of deadline violation. An enhanced SMA having a dynamic quantum rotation gate (DQRG) and opposition-based learning (OBL) was modelled in [17]. To be Specific, for the first time, 2 systems were utilized. The OBL and DQRG are utilized concurrently to enhance the original SMA's sturdiness. The DQRG devises an adaptive parameter control method related to the fitness for obtaining a balance betwixt exploration and exploitation when a comparison is made to original quantum rotation gates. The OBL strategy prevents falling into the local optima and enriches population diversity.

Bhattacharyya et al. [18] presented a quantum metaheuristic technique based on the behavior of sperm whales to optimum thresholding of gray-level imageries. Outcomes were illustrated on 4 test images with 3 threshold levels. Xu et al. [19] devise a lattice-related multi-use secret-sharing method to prevent dispensing novel secret shares to every participant in all rounds of FL whereas attaining post-quantum security. In contrast, this innovative tool permits all participants for updating their secret shares locally whereas preserving the privacy of participants' gradients against quantum assaults. At last, this study implements this novel confidential sharing method for building a lattice-related FL protocol LaF.

In [20], the FL system is presented in the DL of medicinal approaches in IoT-related healthcare systems. Cryptographic primitives, including homomorphic encryption and masks, were implemented to secure local models and avoid adversaries from inferring private healthcare datasets through several assaults like model reconstruction assault or model inversion assault, etc. Kang and Ahn [21] presented a technique to distribute a method with a structure different to the server model, dispensing a process appropriate for customers with various data sizes and training a server model utilizing rebuilt system trained by the user.

Vaiyapuri et al. [22] establish an FL-based IDS utilizing bird swarm algorithm-based FS with classification (FLIDS-BSAFSC) approach in an IoT platform. The presented approach primarily executes a min–max normalized system for pre-processing the IoT data. Eventually, a social group optimizer system with the KELM approach was utilized to identify several types of classes. Abasi et al. [23] examine a federated GWO (FedGWO) technique for reducing data communication. The proposed technique enhances the efficiency under unstable network surroundings with transfer score rules instead of every client model weight.

Połap and Woźniak [24] present a hybridization of this kind of training with metaheuristics. The metaheuristic technique was adjusted for managing the whole method and for analysing the better approaches for minimizing attacks on this kind of collaboration. A novel solution was examined concerning its applications to the problem of image classifiers utilizing typical CNNs, and one of the famous metaheuristic approaches. In [25], a new federated probabilistic predicting approach to solar irradiation was presented dependent upon DL, variational Bayesian inference, and FL. During this method, the trained data can be saved and calculated from local IoT devices, only predicting approaches are shared.

Xie et al. [26] developed an asynchronous measurement-device-independent quantum key distribution technique for surpassing the secret key capacity even without phase tracking and locking. Gu et al. [27] adopted the reference approach for proving the security of effective four-phase measurement-device-independent QKD by the use of laser pulses over significant source imperfections. It is exhibited

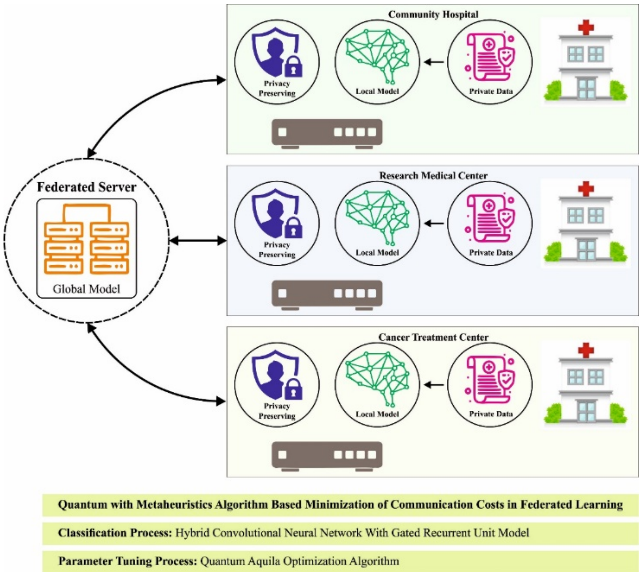

FIGURE 1. The overall procedure of the QMAMCC-FL system.

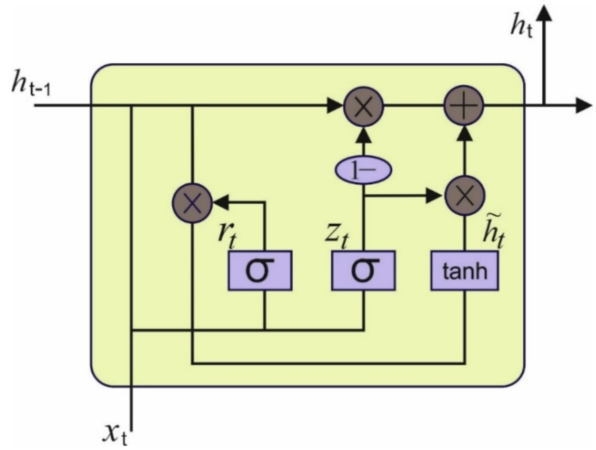

FIGURE 2. The architecture of GRU.

that the practicability of the method can be accomplished by a proof-of-principle experimental implementation with a 20 dB channel loss. The proposed model considerably enhances the secured key rate and the communication distance compared to the existing QKD protocol with imperfect devices.

III. THE PROPOSED MODEL

The study introduced a new QMAMCC-FL method to reduce communication costs in the FL technique. The presented QMAMCC-FL technique designed a federated HCNN-GRU model with the QAO algorithm. The QMAMCC-FL technique upgrades the global model via weight collection of the learned model, which is commonly used in FL. Fig. 1 demonstrates the working procedure of the QMAMCC-FL method.

Communicating the client updates to the peer-to-peer or server is essential in the design of the FL system. There are two transmission frameworks in FL: synchronous and asynchronous transmission. In synchronous transmission, the clients' model updates are aggregated only when they complete their local training and convey their updates to the server. In asynchronous transmission, any client can be aggregated with the global model at t time and doesn't wait for others to complete their training. FedAvg is a well-known optimization algorithm aggregating the model amongst synchronous transmission techniques. In the presented method, every client undergoes a certain amount of training epochs to obtain the local updates on its data using optimizers such as Stochastic Gradient Descent (SGD). Afterwards every round of training, the client transfers the weight and bias to the server, where they are aggregated (using two-stream FL) for making a global model. The core element to reduce the count of transmission rounds between servers and clients is increasing the count of local training epochs at all the clients. The greater the number of local epochs while training, the fewer times weight are transmitted with the server. Firstly, every sequence with N observations is regarded as data at every client, and this sequence of data is given as input to the HCNN-GRU architecture. Fig. 2 showcases the infrastructure of GRU.

The HCNN-GRU architecture transmitted to the client will be trained on the local series and updated with its parameters. These updated parameters from every client will aggregate at the server and continue training for a few rounds as the data is limited. The handling data is the only possible distinction with time-series data compared to classification or regression problems while implemented as a prototype. The individual time series data could not be split between clients; approaches like sliding window were exploited to input the data while training every client sequentially. The HCNN-GRU architecture is compared to the centralized model for confirming the working of time series problems in a federated setting. Moreover, to discover an effective method to communicate between the peer-to-peer or server and the client transmission, a set of active clients was selected for sending the initial model with randomly initialized weight. Every client receives the model, and the training begins at the local client such that the data remain with the client and is not revealed. Afterwards, the initial round of training is finished, and each client model's weight is upgraded. Subsequently, each parameter is shared with the server at the beginning, and every client parameter is stored in the corresponding list on the server. Next, the value from every list was aggregated for updating the global model.

A. TWO-STAGE FL MODEL

Compared to other SGD-based algorithms, the FL model reduces communication rounds and overcomes the problems of Non-IID data distribution. FL approach is longer needed around 10 × transmission rounds while data is distributed in a Non-IID manner [28]. Considering the client in the FL setting was typically in the form of smart IoT devices, mobile phones and wearable devices, the connection between them and the servers is unreliable or frequently slow. Consequently, the transmission plays the bottleneck role in the optimization technique, and cost remains the principal constraint. In the presented method, the single model has been replaced with the two-stream model to be trained on all the clients in the FL setting has been explored. More formally, Θ^G and Θ^L are considered the model parameter, that is, the biases and weights of every layer in the local and global models correspondingly. Assume $X^t = \{x_i^t\}_{i=1}^{n_t}$ and $Y^t = \{y_i^t\}_{i=1}^{n_t}$ represent the training dataset and corresponding label on client t, where n_t represent a number of instances. The global model can get through the server in the first place of the present round and be fixed in subsequent training methods, whereas the local model can be initialized with a parameter of the global model ($\Theta^L \Leftarrow \Theta^G$) and later trained on the local dataset X^t and label Y^t by minimalizing the loss function as

$$L\left(\Theta^L|\Theta^G, X^t, Y^t\right) = L_{cls} + L_{MMD} \quad (1)$$

$$L_{cls} = \frac{1}{n^t}\sum\nolimits_{i=1}^{n^t} J\left(\theta^L\left(x_i^t\right), y_i^t\right) \quad (2)$$

$$L_{MMD} = \lambda MMD^2\left(\theta^G\left(X^t\right), \theta^L\left(X^t\right)\right) \quad (3)$$

where Θ^G and Θ^L represents the model parameters, i.e. bias and weight of every layer in the local and global models. $\theta^G(X^t)$ and $\theta^L(X^t)$ symbolize the outcome of global and local models with respective input X^t. $J(\theta(x), y)$ represents a typical classification loss, for example, the cross-entropy loss function in this experiment. L_{MMD} signifies MMD loss between outputs of global and local methods calculated. This term can be weighted using coefficient λ.

FL model is a cycle procedure of learning local representation, combining knowledge from multiple clients and learning again. In other words, global models contain additional knowledge from different clients, whereas local models learn a better representation of local data. By minimalizing MMD loss terms between the output of global and local processes, we force the local model to learn further knowledge from others along with the data representations on the existing client; thereby, the convergence of the training model has been accelerated, especially minimizing the transmission round.

B. QAO-BASED COMMUNICATION COST MINIMIZATION

The QAO transmits the optimal score (loss or accuracy) to the server using the QAO algorithm for transmitting the trained models. AO is a population-based optimized technique that stimulates Aquila's social activities for catching prey [29]. Like other metaheuristic algorithms, this approach initiates by establishing the primary population X with N number of agents. The following formula was used for implementing the process.

$$X_{ij} = r_1 \times (UB_j - LB_j) + LB_j, i = 1, 2, Nj = 1, 2, \ldots, Dim \quad (4)$$

Now, UB_j and LB_j are the upper and lower boundaries of search space. $r_1 \in [0, 1]$ is the randomly generated number, and Dim indicates the dimension of the agent. In the presented method, the next phase is to do exploration and exploitation till a better solution is attained. The X_b optimum agent and (X) average agent is exploited in the exploration, and the mathematical formula is provided in Eqs. (5) and (6):

$$X_i(t+1) = X_b(t) \times \left(\frac{1-t}{T}\right) + (X_M(t) - X_b(t) * rand), \quad (5)$$

$$X_M(t) = \frac{1}{N}\sum\nolimits_{i=1}^{N} X(t), \forall j = 1, 2, \ldots, Dim \quad (6)$$

The exploration process is controlled based on $(\frac{1-t}{T})$. The maximum amount of generations is indicated as T. The exploration process uses Levy flight (Levy (D) distribution and X_b to upgrade the solution using the following equations:

$$X_i(t+1) = X_b(t) \times Levy(D) + X_R(t) + (y - x) * rand, \quad (7)$$

$$Levy(D) = s \times \frac{u \times \sigma}{|v|^{\frac{1}{\beta}}}, \sigma = \left(\frac{\Gamma(1+\beta) \times sine\left(\frac{\pi\beta}{2}\right)}{\Gamma\left(\frac{1+\beta}{2}\right) \times \beta \times 2^{\left(\frac{\beta-1}{2}\right)}}\right) \quad (8)$$

where $s = 0.01$ and $\beta = 1.5$.u and v demonstrate the random number. X_R is a randomly selected agent. Furthermore, y and x variables are used for stimulating spiral shape:

$$y = r \times cos(\theta), x = r \times sin(\theta) \quad (9)$$

$$r = r_1 + U \times D_1, \theta = -\omega \times D_1 + \theta_1, \theta 1 = \frac{3 \times \pi}{2} \quad (10)$$

where $\omega = 0.005$ and $U = 0.00565$.$r_1 \in [0, 20]$ is a randomly generated integer. based on X_b and X_M, the initial method improves the agent in the exploitation phase:

$$X_i(t+1) = (X_b(t) - X_M(t)) \times \alpha - rnd + (UB \times rnd + LB) \times \delta \quad (11)$$

Now, $UB = (UB - LB)$, α, and δ denote exploitation adjustment parameters. $rnd \in [0, 1]$ is a random number:

$$X_i(t+1) = QP \times X_b(t) - GX - G_2 \times Levy(D) + rnd \times G_1 \quad (12)$$

$$GX = (G_1 \times X(t) \times rnd)$$
$$QP(t) = t^{\frac{2 \times rnd() - 1}{(1-T)^2}} \quad (13)$$

Moreover, G_1 is the motion used to track the best solution by using the subsequent formula:

$$G_1 = 2 \times rnd() - 1, G_2 = 2 \times \left(1 - \frac{t}{T}\right) \quad (14)$$

In Eq. (14), rnd signifies a random value. Furthermore, G_2 shows a variable that minimalized from two to zero:

$$G_2 = 2 \times \left(1 - \frac{t}{T}\right) \quad (15)$$

TABLE 1. Accuracy analysis of the QMAMCC-FL method with other systems under the CIFAR-10 dataset. [31], [32].

No. of Epochs	QMAMCC-FL	PSO-FED	ACO-FED	GWO-FED	CSO-FED	Traditional Federated Learning
0	0	0	0	0	0	0
10	51.25	16.68	12.57	19.55	19.55	25.51
20	65.48	27.97	24.89	39.67	35.16	56.92
30	80.48	36.80	39.26	53.84	50.35	75.20
40	81.51	52.20	54.46	62.06	62.26	77.87
50	80.48	58.98	62.67	66.57	66.78	78.69
60	81.31	59.18	64.31	69.86	72.53	79.10
70	80.07	58.36	63.70	71.30	72.53	79.51
80	81.31	60.00	64.52	71.50	73.96	78.89
90	82.74	59.80	64.52	70.88	75.40	79.30
100	84.59	60.41	64.72	71.50	76.63	78.69

The QAO algorithm is developed based on quantum computing. The small modules of datasets used in quantum computation can be known as a Q-bit or quantum bit. A quantum bit dissimilar to a conventional bit can be in "1 state 0" or superposition among them [30]. To maximize stochastic modules to initialize seed (population), all seeds were described by the one Q- bit termed as Q-seed. The state of Q-bit (Ψ) is described by:

$$\Psi = \bigcup_{j=1}^{n} |\psi_j(t)\} = [\alpha_j \beta_j]^T$$
$$j = 1, 2, \cdots, n \tag{16}$$

In Eq. (16), α and β are arbitrary integers that characterized state possibility. $|\alpha|^2 and |\beta|^2$ signify the possibility that Q-bit $|\psi$ relies upon "0" and "1" states. Then, they satisfy the relation $|\alpha|^2 + |\beta|^2 = 1$. Primarily, the mediocrity of range [0, 1] can be chosen as the first population.

$$\psi_1 = \frac{1}{\sqrt{2}} \left[1 \;\; 1 \right] \tag{17}$$

After initialization, a randomly generated seed can be normalized between zero and one:

$$|\psi\rangle_{[0,1]} = \frac{|\psi\rangle - \min(|\psi\rangle)}{Max(|\psi\rangle) - \min(|\psi\rangle)}$$

where $\min(|\psi)$ and $\mathrm{Max}(|\psi\rangle)$ indicated lower and upper limits.

IV. RESULTS AND DISCUSSION

To examine the performance of the proposed method, a wide range of experiments were conducted to compute the convergence rate and accuracy and experiments in unstable network surroundings. At first, MNIST and CIFAR-10 datasets are used for the accuracy benchmarks and reviewed the cost of data transmission between clients and servers. Then, the accuracy of the proposed method was investigated under different network surroundings.

Algorithm 1 Pseudocode of AO Algorithm

```
Initialize parameter
WHILE (ending condition is not satisfied) do
  Define the value of the fitness function
  X_best(t) = Calulate optimum solution by using fitness value
  for (i= 1, 2 . . . ,N) do
  Upgrade the mean value of the existing solution X_M(t).
  Upgrade the x, y, G_1, G_2, Levy(D), and so on.
    if τ ≤ (2/3)*T then
    if rand ≤ 0.5 then
      Expanded exploration (X_1)
      Upgrade existing solution
      if Fitness (X_1(t+1)) <Fitness(X(t)) then
       X(t) = (X_1(t + 1))
      if Fitness (X_1(t+1)) <Fitness(X_best(t)) then
      X_best(T) = X_1(T+1)
    end if
    end if
  else
    Narrowed exploration (X_2) }
    Upgrade existing solution.
    if Fitness (X_2(t+1)) <Fitness(X(t)) then
        X(T) = (X_2(T + 1))
    if Fitness (X_2(t+1)) <Fitness(X_best(t)) then
      X_best(t) = X_2(t+1)
    end if
  end if
end if
else
if rand ≤ 0.5 then
{Expanded exploitation (X_3) }
Upgrade existing solution
 if Fitness (X_3(t+1)) <Fitness(X(t)) then
          X(t) = (X_3(t + 1))
        if Fitness (X_3(t+1)) <Fitness(X_best(t)) then
          X_best(T) = X_3(T+1)
        end if
      end if
else
Narrowed exploitation (X_4)
Upgrade existing solution
  if Fitness (X_4(t+1)) <Fitness(X(t)) then
      X(T) = (X_4(T+1))
    If fitness (X_4(t+1)) <Fitness(X_besf(t)) then
          X_best(t) = X_4(t+1)
      end if
      end if
      end if
      end if
      end for
  end while
  Return optimum solution (X_best).
```

In this section, a brief set of experiments were undertaken to highlight the betterment of the QMAMCC-FL technique using existing techniques [31], [32]. Table 1 reports the $accu_y$ examination of the QMAMCC-FL technique with other models on the CIFAR-10 dataset. The outcomes identified that the QMAMCC-FL system had displayed maximum

TABLE 2. Accuracy analysis of the QMAMCC-FL method with other systems under the MNIST dataset.

Accuracy (%) MNIST Dataset						
No. of Epochs	QMAMCC-FL	PSO-FED	ACO-FED	GWO-FED	CSO-FED	Traditional Federated Learning
0	0	0	0	0	0	0
10	59.28	22.80	19.70	29.64	25.67	44.58
20	75.77	32.21	31.90	48.00	43.12	68.51
30	81.08	43.32	48.61	59.02	56.71	76.27
40	81.87	56.06	59.92	65.18	66.31	78.47
50	81.77	59.67	63.26	69.26	68.50	79.10
60	83.02	58.40	64.59	70.98	72.99	79.38
70	83.07	58.46	64.08	70.98	73.09	79.69
80	82.24	59.64	64.66	71.06	75.08	78.80
90	84.86	60.53	64.50	71.33	75.95	77.99
100	84.86	60.53	64.50	71.33	75.95	77.99

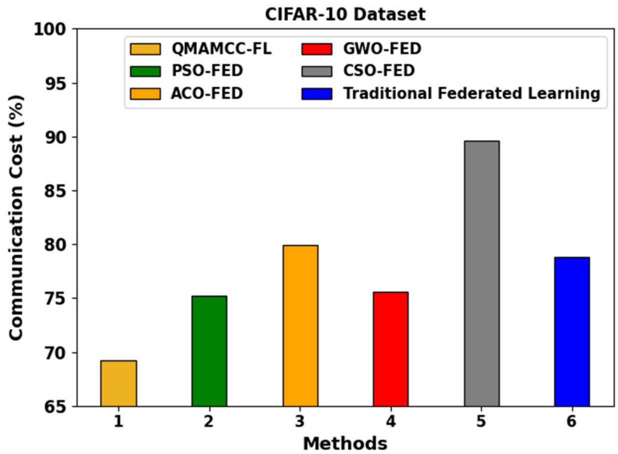

FIGURE 3. Communication cost analysis of QMAMCC-FL method under the CIFAR-10 dataset.

$accu_y$ under all epochs. For example, with 10 epochs, the QMAMCC-FL method has offered an increased $accu_y$ of 51.25% while the PSO-FED, ACO-FED, GWO-FED, CSO-FED, and TFL models have gained reduced $accu_y$ of 16.68%, 12.57%, 19.55%, 19.55%, and 25.51% respectively. Next, with 20 epochs, the QMAMCC-FL method has offered a higher $accu_y$ of 65.48% while the PSO-FED, ACO-FED, GWO-FED, CSO-FED, and TFL approaches have reached decreased $accu_y$ of 27.97%, 24.89%, 39.67%, 35.16%, and 56.92% correspondingly. Then, with 40 epochs, the QMAMCC-FL technique provided a maximum $accu_y$ of 81.51% while the PSO-FED, ACO-FED, GWO-FED, CSO-FED, and TFL models achieved lower $accu_y$ of 52.20%, 54.46%, 62.06%, 62.26%, and 77.87% correspondingly. Meanwhile, with 60 epochs, the QMAMCC-FL approach has offered a superior $accu_y$ of 81.31% while the PSO-FED, ACO-FED, GWO-FED, CSO-FED, and TFL models have gained minimal $accu_y$ of 59.18%, 64.31%, 69.86%, 72.53%, and 79.10% correspondingly. Eventually, with 80 epochs, the QMAMCC-FL technique has offered a maximum $accu_y$ of 81.31% while the PSO-FED, ACO-FED, GWO-FED, CSO-FED, and TFL methods have gained a reduced $accu_y$ of 60%, 64.52%, 71.50%, 73.96%, and 78.89% respectively. Finally, with 100 epochs, the QMAMCC-FL system has offered an enhanced $accu_y$ of 84.59% while the PSO-FED, ACO-FED, GWO-FED, CSO-FED, and TFL models have attained lesser $accu_y$ of 60.41%, 64.72%, 71.50%, 76.63%, and 78.69% correspondingly.

Table 2 reports the $accu_y$ investigation of the QMAMCC-FL method with other approaches on the MNIST dataset. The results identified that the QMAMCC-FL approach had displayed maximal $accu_y$ under all epochs. For the sample, with 10 epochs, the QMAMCC-FL method has offered a higher $accu_y$ of 59.28% while the PSO-FED, ACO-FED, GWO-FED, CSO-FED, and TFL approaches have achieved minimal $accu_y$ of 22.80%, 19.70%, 29.64%, 25.67%, and 44.58% correspondingly. Afterwards, with 20 epochs, the QMAMCC-FL system has offered an increased $accu_y$ of 75.77% while the PSO-FED, ACO-FED, GWO-FED, CSO-FED, and TFL models have gained reduced $accu_y$ of 32.21%, 31.90%, 48%, 43.12%, and 68.51% correspondingly. Afterwards, with 40 epochs, the QMAMCC-FL methodology has obtainable enhanced $accu_y$ of 81.87%. In contrast, the PSO-FED, ACO-FED, GWO-FED, CSO-FED, and TFL algorithms have obtained lesser $accu_y$ of 56.06%, 59.92%, 65.18%, 66.31%, and 78.47% correspondingly. In the meantime, with 60 epochs, the QMAMCC-FL methodology has provided an increased $accu_y$ of 83.02% while the PSO-FED, ACO-FED, GWO-FED, CSO-FED, and TFL methods have reached decreased $accu_y$ of 58.40%, 64.59%, 70.98%, 72.99%, and 79.38% correspondingly. Followed by, with 80 epochs, the QMAMCC-FL technique has obtainable increased $accu_y$ of 82.24% while the PSO-FED, ACO-FED, GWO-FED, CSO-FED, and TFL models have gained reduced $accu_y$ of 59.64%, 64.66%, 71.06%, 75.08%, and 78.80% correspondingly. Lastly, with 100 epochs, the QMAMCC-FL technique has accessibly increased $accu_y$ of 84.86% while the PSO-FED, ACO-FED, GWO-FED, CSO-FED, and TFL models have gained reduced $accu_y$ of 60.53%, 64.50%, 71.33%, 75.95%, and 77.99% correspondingly.

Fig. 3 exhibits a communication cost assessment of the QMAMCC-FL technique with recent models on the CIFAR-10 dataset. The figure indicated that the CSO-FED technique had shown poor performance with a maximum cost of 89.64%. Next, the ACO-FED and TFL models have obtained slightly improvised outcomes with costs of 79.89% and 78.83%, respectively. The PSO-FED and GWO-FED models have reported considerable costs of 75.18% and 78.83%, respectively. However, the QMAMCC-FL technique has shown maximum outcomes with a lower cost of 69.21%.

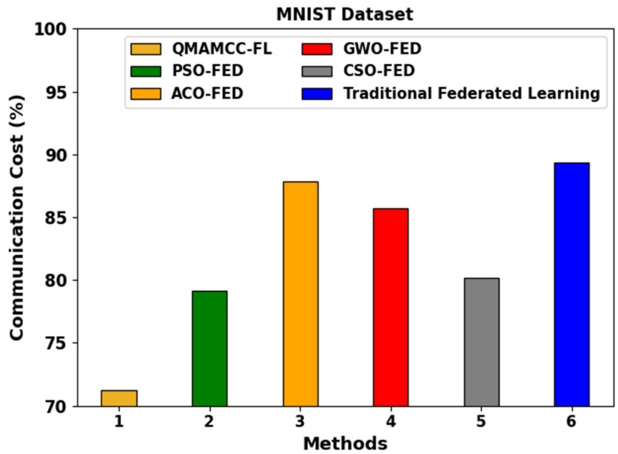

FIGURE 4. Communication cost analysis of QMAMCC-FL approach under MNIST dataset.

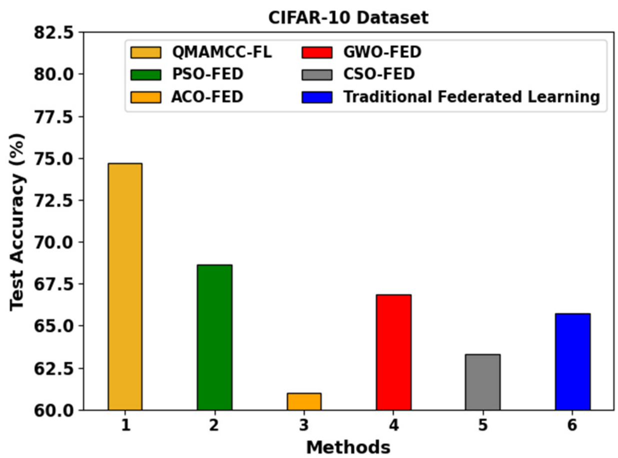

FIGURE 5. Test accuracy analysis of the QMAMCC-FL method under the CIFAR-10 dataset.

Fig. 4 demonstrates a communication cost analysis of the QMAMCC-FL system with recent models on the MNIST dataset. The figure stated that the CSO-FED technique had demonstrated worse performance with a maximal cost of 80.15%. Afterwards, the ACO-FED and TFL models gained somewhat improvised outcomes with costs of 87.82% and 89.34%, correspondingly. Followed by this, the PSO-FED and GWO-FED approaches have managed to report considerable costs of 79.13% and 85.72%, correspondingly. However, the QMAMCC-FL algorithm has demonstrated maximal outcomes with a lower cost of 71.21%.

Fig. 5 displays a test accuracy investigation of the QMAMCC-FL algorithm with recent models on the CIFAR-10 dataset. The figure shows that the ACO-FED system exhibited the least performance with a minimal test accuracy of 61%. Next, the CSO-FED and TFL methods have attained somewhat improvised outcomes with a test accuracy of 63.28% and 65.74%, respectively. Similarly, the PSO-FED and GWO-FED models have managed to report considerable costs of 68.63% and 66.85%, correspondingly. At last, the QMAMCC-FL approach has outperformed higher outcomes with a higher test accuracy of 74.69%.

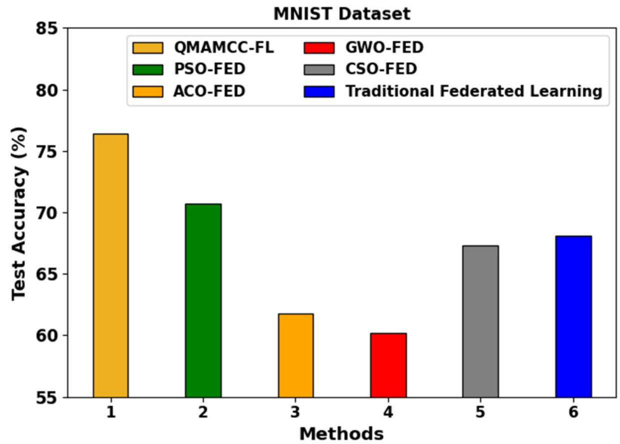

FIGURE 6. Test accuracy analysis of QMAMCC-FL approach under MNIST dataset.

Fig. 6 showcases a test accuracy assessment of the QMAMCC-FL method with recent methods on the MNIST dataset. The figure pointed out that the ACO-FED system has shown poor performance with a minimal test accuracy of 61.78%. Next, the CSO-FED and TFL algorithms have obtained somewhat improvised outcomes with a test accuracy of 67.33% and 68.14%, correspondingly. Following, the PSO-FED and GWO-FED models have managed to report considerable costs of 70.71% and 60.23% correspondingly. Eventually, the QMAMCC-FL technique exhibited maximal outcomes with a superior test accuracy of 76.38%. These results confirmed the enhanced outcomes of the QMAMCC-FL technique.

V. CONCLUSION

This study developed a new QMAMCC-FL system to reduce communication costs in the FL technique. The presented QMAMCC-FL technique designed a federated HCNN-GRU model with the QAO algorithm. The QMAMCC-FL technique upgrades the global model via weight collection of the learned model, which is widely used in FL. The QMAMCC-FL technique attains increased robustness in unstable networks with enhanced communication efficiency. Moreover, the amount of data exploited in the network communication is decreased, and the performance is improved. The experimental analysis of the QMAMCC-FL technique is tested, and the results show better performance over other existing approaches. The proposed method is used to enhance the performance of network communication and decrease the size of data transmitted from clients to the servers such as smartphones and tablets. In the future, hybrid metaheuristic approaches can be devised to improvise the communication cost in the FL further.

ACKNOWLEDGMENT

The authors extend their appreciation to the Deanship of Scientific Research at King Khalid University for funding this work through Large Groups Project under grant number (RGP2/112/43). Princess Nourah bint Abdulrahman University Researchers Supporting Project number (PNURSP2023R237), Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia. The Authors would like to thank the Deanship of Scientific Research at Majmaah University for supporting this work under Project No. R-2023-526. This study is supported via funding from Prince Sattam bin Abdulaziz University project number (PSAU/2023/R/1444).

REFERENCES

[1] C. Wu, F. Wu, L. Lyu, Y. Huang, and X. Xie, "Communication-efficient federated learning via knowledge distillation," *Nature Commun.*, vol. 13, no. 1, pp. 1–8, Apr. 2022.

[2] Y. Li, R.-G. Zhou, R. Xu, J. Luo, W. Hu, and P. Fan, "Implementing graph-theoretic feature selection by quantum approximate optimization algorithm," *IEEE Trans. Neural Netw. Learn. Syst.*, early access, Jul. 21, 2022, doi: 10.1109/TNNLS.2022.3190042.

[3] K. Bonawitz, F. Salehi, J. Konecný, B. McMahan, and M. Gruteser, "Federated learning with autotuned communication-efficient secure aggregation," in *Proc. 53rd Asilomar Conf. Signals, Syst., Comput.*, Nov. 2019, pp. 1222–1226, doi: 10.1109/IEEECONF44664.2019.9049066.

[4] A. R. Javed, M. A. Hassan, F. Shahzad, W. Ahmed, S. Singh, T. Baker, and T. R. Gadekallu, "Integration of blockchain technology and federated learning in vehicular (IoT) networks: A comprehensive survey," *Sensors*, vol. 22, no. 12, p. 4394, Jun. 2022.

[5] M. Chen, Z. Yang, W. Saad, C. Yin, H. V. Poor, and S. Cui, "Performance optimization of federated learning over wireless networks," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2019, pp. 1–6, doi: 10.1109/GLOBECOM38437.2019.9013160.

[6] X. Wu, Y. Zhang, M. Shi, P. Li, R. Li, and N. N. Xiong, "An adaptive federated learning scheme with differential privacy preserving," *Future Gener. Comput. Syst.*, vol. 127, pp. 362–372, Feb. 2022.

[7] A. Alferaidi, K. Yadav, Y. Alharbi, W. Viriyasitavat, S. Kautish, and G. Dhiman, "Federated learning algorithms to optimize the client and cost selections," *Math. Problems Eng.*, vol. 2022, pp. 1–9, Apr. 2022, doi: 10.1155/2022/8514562.

[8] I. Hegedũs, G. Danner, and M. Jelasity, "Decentralized recommendation based on matrix factorization: A comparison of gossip and federated learning," in *Proc. Joint Eur. Conf. Mach. Learn. Knowl. Discovery Databases*, 2019, pp. 317–332, doi: 10.1007/978-3-030-43823-4_27.

[9] A. Hammoud, A. Mourad, H. Otrok, O. A. Wahab, and H. Harmanani, "Cloud federation formation using genetic and evolutionary game theoretical models," *Future Gener. Comput. Syst.*, vol. 104, pp. 92–104, Mar. 2020.

[10] H.-L. Yin, Y. Fu, C.-L. Li, C.-X. Weng, B.-H. Li, J. Gu, Y.-S. Lu, S. Huang, and Z.-B. Chen, "Experimental quantum secure network with digital signatures and encryption," 2021, *arXiv:2107.14089*.

[11] A. Fadaeddini, B. Majidi, and M. Eshghi, "Privacy preserved decentralized deep learning: A blockchain based solution for secure AI-driven enterprise," in *Proc. Int. Congr. High-Perform. Comput. Big Data Anal.*, 2019, pp. 32–40, doi: 10.1007/978-3-030-33495-6_3.

[12] N. H. Tran, W. Bao, A. Zomaya, M. N. H. Nguyen, and C. S. Hong, "Federated learning over wireless networks: Optimization model design and analysis," in *Proc. IEEE Conf. Comput. Commun.*, Apr. 2019, pp. 1387–1395, doi: 10.1109/INFOCOM.2019.8737464.

[13] X. Yao, T. Huang, C. Wu, R.-X. Zhang, and L. Sun, "Federated learning with additional mechanisms on clients to reduce communication costs," 2019, *arXiv:1908.05891*.

[14] M. Chen, Z. Yang, W. Saad, C. Yin, H. V. Poor, and S. Cui, "A joint learning and communications framework for federated learning over wireless networks," *IEEE Trans. Wireless Commun.*, vol. 20, no. 1, pp. 269–283, Jan. 2021.

[15] M. Asad, A. Moustafa, and M. Aslam, "CEEP-FL: A comprehensive approach for communication efficiency and enhanced privacy in federated learning," *Appl. Soft Comput.*, vol. 104, Jun. 2021, Art. no. 107235.

[16] R. Jain and N. Sharma, "A quantum inspired hybrid SSA–GWO algorithm for SLA based task scheduling to improve QoS parameter in cloud computing," *Cluster Comput.*, pp. 1–24, Sep. 2022, doi: 10.1007/S10586-022-03740-X.

[17] Y. Zhang, S. Du, and Q. Zhang, "Improved slime mold algorithm with dynamic quantum rotation gate and opposition-based learning for global optimization and engineering design problems," *Algorithms*, vol. 15, no. 9, p. 317, Sep. 2022.

[18] S. Bhattacharyya, S. Dey, J. Platos, V. Snasel, and T. Dutta, "Multilevel quantum sperm whale metaheuristic for gray-level image thresholding," in *Proc. Int. Conf. Innov. Comput. Commun.*, 2020, pp. 775–788, doi: 10.1007/978-981-15-1286-5_69.

[19] P. Xu, M. Hu, T. Chen, W. Wang, and H. Jin, "LaF: Lattice-based and communication-efficient federated learning," *IEEE Trans. Inf. Forensics Security*, vol. 17, pp. 2483–2496, 2022.

[20] L. Zhang, J. Xu, P. Vijayakumar, P. K. Sharma, and U. Ghosh, "Homomorphic encryption-based privacy-preserving federated learning in IoT-enabled healthcare system," *IEEE Trans. Netw. Sci. Eng.*, early access, Jun. 30, 2022, doi: 10.1109/TNSE.2022.3185327.

[21] D. Kang and C. W. Ahn, "Communication cost reduction with partial structure in federated learning," *Electronics*, vol. 10, no. 17, p. 2081, Aug. 2021.

[22] T. Vaiyapuri, S. Algamdi, R. John, Z. Sbai, M. Al-Helal, A. Alkhayyat, and D. Gupta, "Metaheuristics with federated learning enabled intrusion detection system in Internet of Things environment," *Exp. Syst.*, vol. 40, no. 5, Jun. 2023, Art. no. e13138.

[23] A. K. Abasi, M. Aloqaily, and M. Guizani, "Grey wolf optimizer for reducing communication cost of federated learning," in *Proc. IEEE Global Commun. Conf.*, Dec. 2022, pp. 1049–1154.

[24] D. Połap and M. Woźniak, "Meta-heuristic as manager in federated learning approaches for image processing purposes," *Appl. Soft Comput.*, vol. 113, Dec. 2021, Art. no. 107872.

[25] X. Zhang, F. Fang, and J. Wang, "Probabilistic solar irradiation forecasting based on variational Bayesian inference with secure federated learning," *IEEE Trans. Ind. Informat.*, vol. 17, no. 11, pp. 7849–7859, Nov. 2021.

[26] Y.-M. Xie, Y.-S. Lu, C.-X. Weng, X.-Y. Cao, Z.-Y. Jia, Y. Bao, Y. Wang, Y. Fu, H.-L. Yin, and Z.-B. Chen, "Breaking the rate-loss bound of quantum key distribution with asynchronous two-photon interference," *PRX Quantum*, vol. 3, no. 2, Apr. 2022, Art. no. 020315.

[27] J. Gu, X.-Y. Cao, Y. Fu, Z.-W. He, Z.-J. Yin, H.-L. Yin, and Z.-B. Chen, "Experimental measurement-device-independent type quantum key distribution with flawed and correlated sources," *Sci. Bull.*, vol. 67, no. 21, pp. 2167–2175, Nov. 2022.

[28] X. Yao, C. Huang, and L. Sun, "Two-stream federated learning: Reduce the communication costs," in *Proc. IEEE Vis. Commun. Image Process. (VCIP)*, Dec. 2018, pp. 1–4, doi: 10.1109/VCIP.2018.8698609.

[29] S. Mahajan, L. Abualigah, A. K. Pandit, and M. Altalhi, "Hybrid Aquila optimizer with arithmetic optimization algorithm for global optimization tasks," *Soft Comput.*, vol. 26, no. 10, pp. 4863–4881, May 2022.

[30] N. Razmjooy and M. Ramezani, "An improved quantum evolutionary algorithm based on invasive weed optimization," *Indian J. Sci. Res*, vol. 4, no. 2, pp. 413–422, 2014.

[31] S. R. Tandra, "Personalized and communication cost reduction models in federated learning," Doctoral dissertation, Faculty Graduate Stud. Res., University of Regina, Regina, SK, Canada, 2022.

[32] S. Park, Y. Suh, and J. Lee, "FedPSO: Federated learning using particle swarm optimization to reduce communication costs," *Sensors*, vol. 21, no. 2, p. 600, Jan. 2021.