Computers, Materials & Continua

Tech Science Press

https://doi.org/10.32604/cmc.2026.079577

ARTICLE

Fed-HOER: Federated Hybrid-Optimized Emotion Recognition Framework Using DBO-FLA Metaheuristic Optimization

Mohammed Shukur Alfaras[1,2,*], Oguz Karan[3], Sefer Kurnaz[1] and Ayca Kurnaz Turkben[4]

[1]Department of Electrical and Computer Engineering, Engineering College, Altinbas University, Istanbul, Turkey
[2]Information and Communications, Planning Department, Babil Education Directorate, Ministry of Education, Hillah, Babil, Iraq
[3]Department of Research and Development, Siemens A.S, Istanbul, Turkey
[4]Department of Computer Engineering, Faculty of Engineering and Natural Sciences, Rumeli University, Istanbul, Turkey

*Corresponding Author: Mohammed Shukur Alfaras. Email: 223720622@ogr.altinbas.edu.tr or moalfarras@gmail.com

Received: 23 January 2026; Accepted: 11 May 2026; Published: 15 June 2026

ABSTRACT: Despite deep learning's high precision in emotion identification, centralized training is associated with privacy and scalability concerns. The privacy-preserving federated learning model, Federated Hybrid-Optimized Emotion Recognition (Fed-HOER), introduced in this paper is an auto-tuning hyperparameters optimizer based on a hybrid Dung Beetle Optimizer-Fick's Law Algorithm (DBO-FLA) optimizer. The global and local searches are optimized at two levels, and validation loss is minimized by 22%–24% without sharing raw data. The experiments on Extended Cohn–Kanade (CK+), Japanese Female Facial Expressions (JAFFE), and Karolinska Directed Emotional Faces (KDEF) exhibit a high generalization rate with a mean accuracy of 98.14. The findings demonstrate that Fed-HOER is statistically significantly better than baseline configurations. The results show that the suggested framework offers a favorable trade-off between predictive accuracy and privacy protection, which is why it can be used in the healthcare, educational, and other emotion-related fields.

KEYWORDS: Emotion recognition; affective computing; federated learning; Dung Beetle Optimizer (DBO); Fick's Law Algorithm (FLA); hybrid metaheuristic optimization; privacy preservation; convolutional neural networks (CNN)

1 Introduction

Emotions are highly powerful in human interaction, and they can be communicated by means of text, speech, and facial expressions [1]. Of these, facial expressions give the most abundant social and emotional information. They play a key role in nonverbal communication, which drives the creation of automated facial emotion recognition (FER) systems of posed and spontaneous expressions [2–5]. FER has also gained increasing attention in computer vision, human-robot interaction, social robotics, and other systems that need the correct perception of emotions [6–8].

Automatic emotion recognition systems still face major challenges, despite significant progress. The large intra- and intercultural variability usually causes overfitting and poor generalization [9,10]. Conventional feature-based approaches like Principal Component Analysis (PCA), Local Binary Patterns (LBP) [11], and Histogram of Oriented Gradients (HOG) [12,13] have the disadvantage of often struggling to generalize to unseen conditions [14], whereas deep Convolutional Neural Networks (CNNs), despite their greater power, are susceptible to cross-dataset generalization problems [15,16]. Other challenges are due to an imbalance in classes [17], diversity in demographics [18,19], and the variations in expression intensity [20].

Copyright © 2026 The Authors. Published by Tech Science Press. This work is licensed under a Creative Commons Attribution 4.0 International License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In addition, most of the current research is based on the standard CNNs or simple transfer learning without additional optimization [21], which demonstrates an evident gap in the hybrid optimization strategies and systematic benchmarking of deep models using metaheuristic optimizers to achieve robust FER.

To overcome these constraints, the proposed study has the following contributions:

- Architecture Design: CNN, MobileNet, and Visual Geometry Group (VGG) architectures are implemented, and their parameters are modified to suit emotion classification.
- The Initial Stages of Optimization: The combination of the individual and early-stage hybrid optimizers, Walrus and Snake, significantly enhances the model's accuracy.
- Comparative Analysis: A deeper analysis of optimized and non-optimized models is done to elaborate on how optimization affects the recognition of emotions.
- Advanced Hybrid Optimization: A new hybrid optimizer is suggested, which exhibits more efficiency in the learning process and is more resilient against classification.

In contrast to the previous literature, which uses fixed hyperparameters and a single architecture, this paper uses a systematic framework of combining various deep learning models with different bio-inspired optimizers and their hybrids. Four metaheuristics, namely Walrus, Snake, Dung Beetle, and Fick Law, are tested and two new hybrids, Snake + Dung Beetle Optimizer (DBO) and DBO + Fick's Law Algorithm (FLA), are proposed to improve the global exploration and local refinement, which has not previously been investigated for FER. The robust generalization is achieved through subject-wise cross-validation and the accuracy, macro-F1, and per-class measures are used to compare baseline, single-optimizer, and hybrid methods. It is evaluated on a variety of architectures, such as CNNs, MobileNets, and VGG16, and compares the proposed hybrids with the existing ones, such as Particle Swarm Optimization–Genetic Algorithm (PSO-GA) and Weighted Quantum-behaved Particle Swarm Optimization (WQPSO).

1.1 Motivation and Research Gap

Despite progress in federated learning and emotion recognition, several gaps remain at their intersection:

- Limited privacy-preserving FER: Most FER systems still rely on centralized data aggregation, conflicting with regulations such as General Data Protection Regulation (GDPR) and Health Insurance Portability and Accountability Act (HIPAA).
- Suboptimal hyperparameter optimization: Federated FER models often use manual tuning or grid search, which is inefficient and may miss optimal configurations.
- Restricted dataset diversity: Many studies are evaluated on a single dataset, limiting generalization across demographics and acquisition conditions.
- Single-algorithm metaheuristics: Hybrid metaheuristic optimization is underexplored, missing the complementary strengths of multiple algorithms.
- Incomplete analysis of federated dynamics: Convergence behavior under heterogeneous client data is not yet thoroughly characterized.

1.2 Contributions

This work addresses the above limitations through the following contributions:

(1) Federated FER architecture: We introduce Federated Hybrid-Optimized Emotion Recognition (FedHOER), a privacy-sensitive federated learning architecture of facial emotion recognition, which can train a model jointly with decentralized clients without exchanging raw facial images. The framework is developed to work with heterogeneous and non-IID data.

(2) Two-stage hybrid optimization strategy: In our proposed hybrid optimization strategy based on client-side hyperparameter tuning, DBO is used to perform global search of the search space, and FLA is used to narrow down the candidate solutions by doing local search. It is a coordinated global-local mechanism that is specifically modified to the federated environment with the aim of enhancing convergence stability.
(3) Systematic testing of optimization strategies: We perform a comparative study of the optimization strategies consisting of baseline, single-optimizer, and hybrid-optimizer settings on various FER datasets involving empirically determined insights on the efficacy of hybrid metaheuristic optimization on the condition of federated and heterogeneous data.
(4) Federation training and convergence analysis: We examine the convergence behavior of the suggested framework per communication round and show that the training dynamics are stable and the performance increases are consistent in the sampling-size-weighted FedAvg aggregation scheme.

2 Related Works

Facial emotion recognition (FER) has been studied using traditional techniques, deep learning systems, transfer learning, and metaheuristic-based optimization frameworks. To emphasize the changes in approaches in this field, the literature will be classified into four groups below.

2.1 Traditional Methods

The initial FER was based on handcrafted features and a traditional classifier. An example is Lopes et al. (2017) [21], who used convolutional neural networks with specialized preprocessing to augment expression-specific features, achieving up to 96.76% accuracy on CK+. These methods, as illustrated in Table 1, were cautious in their data preparation, but were disadvantaged by the fact that they used relatively small and homogeneous datasets.

Table 1: Comparison of representative facial emotion recognition (FER) studies with respect to dataset, method, and optimization strategy.

Study	Dataset(s)	Model/Framework	Optimization Method
Lopes et al. [21]	CK+	CNN with preprocessing	Manual tuning
Fei et al. [22]	5 benchmarks	AlexNet + LDA	Manual tuning
Jain et al. [23]	Cartoon dataset (8k images)	ResNet-50, MobileNetV2, InceptionV3, VGG16	Transfer learning
Kim et al. [24]	Driving dataset	CRNN + LFA + PP2 encryption	Hybrid pipeline
Haq et al. [25]	Real-world images	MobileNet-V1	Transfer learning
Martvel et al. [26]	Cat facial videos	Landmark-based deep learning pipeline	Facial landmark detection + temporal modeling
Sathya and Sudha [27]	EMOTIC, FR	CNN + GRU + LSTM + ANFIS	Ensemble/fuzzy optimization

(Continued)

Table 1 (continued)

Study	Dataset(s)	Model/Framework	Optimization Method
Vaijayanthi and Arunnehru [28]	MUG, GEMEP	DNN (microexpressions)	Swarm-based feature selection
Kasar et al. [29]	FER-2013, CK48, Legend	Modified CNN	Manual tuning
Akrout [30]	CK+, FER-2013, JAFFE	Dual Attention Residual U-Net	FEW-FE + CSMO
Mu et al. [31]	RAF, KDEF	ResNet-50 + QIFABC	Quantum Firefly + ABC
Unnisa and Ganesan [32]	Multiple FER datasets	CNN + XGBoost	Hybrid PSO–GA

2.2 Deep Learning-Based Approaches

Deep neural networks have made a great contribution to FER performance. Previous work has been AlexNet with Linear Discriminant Analysis (LDA) to low-resource mental health assessment [22], Mask R-CNN and VGG16 to cartoon FER with 96% accuracy [23], CRNN-based emotion-aware driving systems [24], and a pose-robust MobileNet-V1 with 97.9% accuracy [25]. Recent works have applied deep learning to animal emotion recognition [26] and large-scale CNN-based FER on FER-2013. Robustness is also enhanced by ensemble and optimization techniques, and several CNN-optimized and hybrid networks achieve accuracies of over 99% [27–30].

2.3 Transfer Learning Techniques

Transfer learning facilitates successful FER by adapting ImageNet-trained models, including ResNet-50, MobileNetV2, InceptionV3, and VGG16, and VGG16 tends to perform the best [23]. MobileNet-V1 has been improved by enhancements that have enhanced the results on real-world low-resolution data [25]. Task-specific refinement has also been shown to be advantageous with hybrid feature-selection models with pretrained CNNs recording high multi-pose FER performance, surpassing 98% accuracy on standard datasets [31].

2.4 Metaheuristics in FER

Metaheuristic optimization has been extensively used for FER to select features and tune models, such as PSO-, GA-, and quantum-inspired hybrids [32]. Although there are previous hybrid optimizers, in the federated case, most methods are centralized or have fixed hyperparameters. This paper overcomes these drawbacks by combining federated learning with a hybrid DBO-FLA optimizer that allows the optimization of hyperparameters of various datasets without exchanging sensitive information.

3 Proposed Methodology

The Fed-HOER model suggested allows the use of privacy-preserving and decentralized emotion recognition with high performance and multiple clients. It has three key elements, as illustrated in Fig. 1, namely (1) client-side CNN-based emotion recognition, (2) two-stage hybrid DBO-FLA hyperparameteroptimizer,

and (3) a federated learning architecture based on FedAvg. Every client trains a CNN on its emotion data, which supports the differences in demographics and acquisition conditions, and retains raw data locally to meet privacy needs, including GDPR and HIPAA.

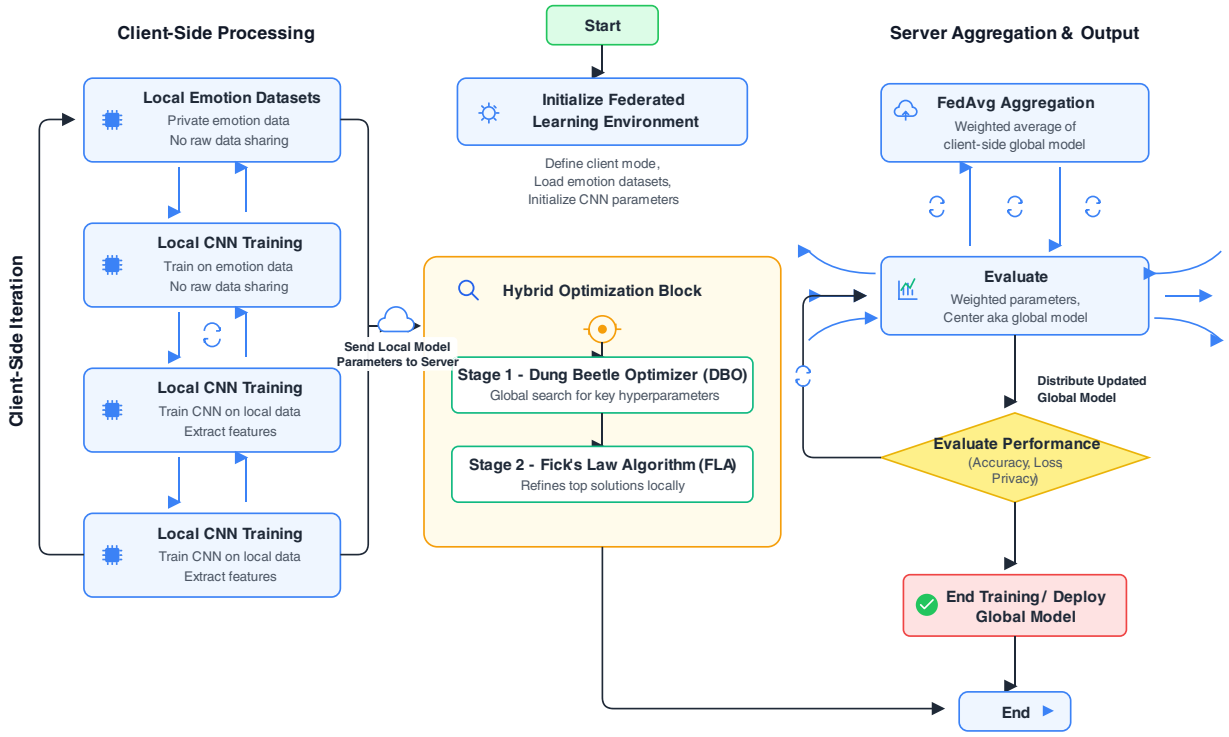

Figure 1: Fed-HOER flowchart: federated CNN training with hybrid DBO–FLA optimization and aggregation workflow.

Hyperparameters (e.g., learning rate, filter size, dropout) are optimized through the hybrid DBO-FLA approach where DBO is used to conduct global exploration, and FLA is used to optimize the best solutions in local settings to enhance convergence and generalization to heterogeneous clients. The model parameters are exchanged with a central server on a periodic basis in which FedAvg averages them with dataset-size-weighted averaging and re-distributes the global model. Fed-HOER achieves good performance of emotion recognition and user privacy by combining local training, hybrid optimization, and federated aggregation.

3.1 Datasets

Fed-HOER is tested on three popular FER benchmarks, namely JAFFE, CK+, and KDEF, which are selected due to their demographic, acquisition condition, and variability diversity as presented in Table 2.

Table 2: Characteristics of the three emotion-recognition datasets used in this study.

Dataset	Total Samples	Number of Classes	Demographics	Format/Notes
JAFFE	213	7	Japanese females (10 subjects)	Grayscale static posed images
CK+	≈593 sequences (peak frames ~327 labelled)	7	Mixed gender	Mostly frontal static or peak frames from sequences
KDEF	4900 images (subset ~490 commonly used)	7	Swedish males and females (35 + 35), multiple angles	Color/multi-angle photos

3.1.1 JAFFE (Japanese Female Facial Expressions)

JAFFE This dataset comprises of 213 gray-scale facial images of 10 Japanese female models displaying six universal emotions and a neutral pose. In the Fed-HOER framework, the JAFFE dataset is "low-diversity, high-precision" challenge. The dataset's demographic homogeneity and limited size make it the main test case for the Fick's Law Algorithm (FLA) part of the hybrid optimizer. This dataset measures the framework's capacity to attain high accuracy and avoid overfitting in a high-precision, low-diversity scenario with limited subjects and facial structural diversity. This guarantees that the DBO-FLA optimization can regularise the model even when it is trained on small-scale and culturally specific datasets.

3.1.2 CK+ (Extended Cohn–Kanade)

CK+ The CK+ dataset contains 593 emotion sequences from 123 participants (mixed gender) ranging in age from 18–50 years. Unlike other static databases, CK+ is inherently dynamic; each sequence represents a temporal sequence from a neutral state to the peak expression. To remain consistent with the static CNN architecture of the Fed-HOER, and retain these high-intensity features, the sequences were transformed into a "temporal-to-static" format. In particular, the last three frames—corresponding to the peak expression—of each sequence were selected. This guarantees that the model can learn from the most expressive facial features, thus solving the problem of high-intensity expression recognition in a controlled pose and lighting environment.

The key distinction is the complexity and diversity of the dataset. JAFFE is a small, static, simple, and culturally homogeneous dataset, suitable for testing accuracy. But CK+ is more diverse, bigger, and contains temporal sequences, so some preprocessing (such as turning temporal data into static images) is required so the Fed-HOER model can work for both simple, small datasets and complex, larger datasets.

3.1.3 KDEF (Karolinska Directed Emotional Faces)

KDEF consists of 4900 color images of 70 actors (35 male, 35 female) portraying seven emotions from multiple viewing angles. This study uses a common subset of 490 frontal and near-frontal images. The multi-view setup and heterogeneous demographics make KDEF a challenging benchmark for generalization.

Within the federated setting, each dataset (JAFFE, CK+, and KDEF) is divided into several federated clients, making each dataset to have more than one client. JAFFE offers 213 images, CK+ some 600 peak frames and KDEF about 490 multi angle images, constituting a heterogeneous benchmark of some

1300 samples in seven categories of emotion. This configuration presents inter-dataset and intra-dataset heterogeneity since the datasets are not acquired under the same conditions, have different demographics, and pose variability, and each client has a different sub-set of local data. JAFFE captures the conditions of control, CK+ captures the demographic heterogeneity and KDEF captures the pose and appearance heterogeneity to allow a complete assessment of Fed-HOER with non-IID and diverse data distributions.

3.2 Image Preprocessing and Experimental Setup

3.2.1 Image Preprocessing Pipeline

To ensure consistent input dimensions and data quality across JAFFE, CK+, and KDEF, a unified preprocessing pipeline was applied before training Fig. 2. Images were converted to grayscale to reduce noise and computation, resized to 224×224, and normalized to the $[0-1]$ range to stabilize training. The grayscale channel was then replicated to form a three-channel pseudo-RGB image compatible with CNNs. The procedure is summarized in Algorithm 1.

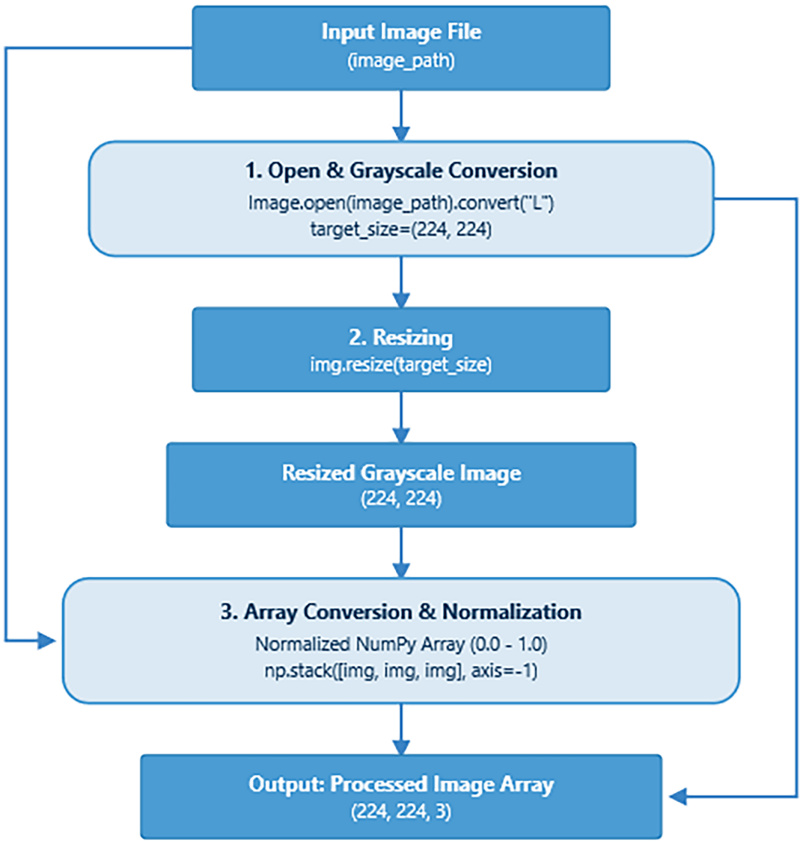

Figure 2: Preprocessing pipeline: grayscale conversion, resizing, normalization, and channel stacking for CNN's input images.

Algorithm 1: Image preprocessing pipeline

Input: Image I, target size (224, 224)
Output: Preprocessed tensor X
1. Load image I
2. Convert to grayscale
3. Resize to 224 × 224
4. Normalize: I ← I/255
5. Replicate channels → 3-channel image
6. Return X

The implementation of grayscale-to-three-channel conversion ensures consistency of input shapes across all datasets and models while maintaining compatibility with deep learning architectures. While this approach minimizes the impact of color variance—particularly for the KDEF dataset—future work will explore the comparative benefits of native RGB processing.

3.2.2 Data Partitioning and Sampling Strategy

To ensure the scientific credibility of the Fed-HOER framework and to eliminate any sampling biases, we adopted a Subject-Independent Cross-Validation approach. This approach, as opposed to random splitting, guarantees that all images of a given subject are included in one and only one of the training, validation, or testing partitions. This approach ensures that facial identity of subjects is not present in both training and testing phases, thus removing the possibility of data leakage and confirming that the model learns common emotional features and not the facial identity of subjects.

An 80/10/10 split was used to divide the global data. 80% was allocated to the three clients' local training. 10% was set aside as a Global Validation Proxy that is used by the server-side DBO-FLA optimizer to evaluate the fitness function (validation loss) and adjust hyperparameters without needing access to the training data. The remaining 10% was reserved as a Hold-out Test Set, which was not used by either the local learners or the metaheuristic optimizer, to report on the final test.

To ensure that no client has a bias, we applied Stratified Random Sampling to allocate data to Client 1, Client 2 and Client 3. This ensures that the three clients have an equal percentage of the seven emotion classes (Angry, Disgust, Fear, Happy, Sad, Surprise, and Neutral), thereby avoiding client-level bias towards a certain emotion class and enabling the global model weights to converge.

3.3 Federated Learning Setup

Every dataset (JAFFE, CK+, and KDEF) is divided into three federated clients in this research which makes a total of nine clients. The partitioning is carried out through the random shuffling of each dataset and the division of the dataset into equilibrium-sized and non-overlapping subsets. It has its own local dataset which is further divided into training, validation, and testing set in a stratified 60:20:20 proportion.

The image level is used to partition the dataset in stratified sampling, and the subject identity is not specifically considered during the split. Consequently, the same subject may be represented in the training, validation, and testing sets, and this brings a risk possibility of subject leakage. However, there are no identical samples that are common in the splits, and all the evaluations are made on unknown images. This limitation is acknowledged, and further efforts will be directed towards subject-independent partitioning schemes to offer a more stringent evaluation of generalization.

This architecture is representative of real-life distributed settings, e.g., hospitals, universities, or research laboratories, where organizations jointly learn a model without exchanging raw data and thus maintain privacy within the federated learning paradigm.

Training in Fed-HOER follows the Federated Averaging (FedAvg) algorithm. As illustrated in Fig. 3, the central server initializes the global model parameters. $w^{(0)}$ and distributes them to all clients. During each communication round r, client i receives the global model $w^{(r)}$. Moreover, performs local training on its private dataset, producing updated local parameters according to the standard local update rule.

$$w_i^{(r+1)} = w^{(r)} - \eta \nabla L_i \left(w^{(r)} \right) \tag{1}$$

where η represents the learning rate, and $\mathcal{L}_i(w^{(r)})$ denotes the local loss function computed on client i's dataset.

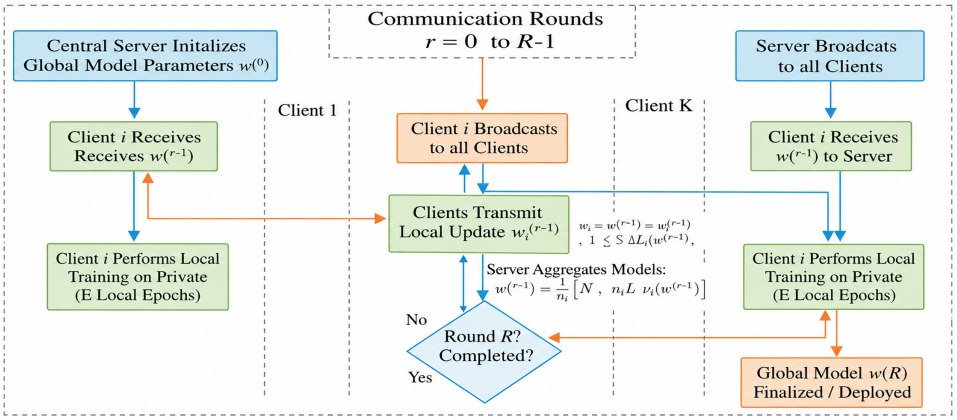

Figure 3: FedAvg workflow: client updates, server aggregation, broadcasts across communication rounds until convergence.

After completing E local epochs of training, each client transmits its locally updated parameters $w_i^{(r+1)}$ to the central server. The server then performs model aggregation using a weighted average, formulated as:

$$w^{(r+1)} = \sum_{i=1}^{K} \frac{n_i}{N} w_i^{(r+1)} \tag{2}$$

where n_i denotes the number of samples at client i, $N = \sum_{i=1}^{K} n_i$ is the total number of samples across all clients, and K is the total number of participating clients. A global model is redistributed to all the clients as $w^{(r+1)}$ and the process is repeated after each R communication round until convergence is reached. The local epochs and communication rounds were empirically set to E = 20 and R = 10, respectively, since preliminary experiments showed that the scheme converged and that the communication overhead was efficient.

The datasets of the current research are heterogeneous with respect to the sample size, distribution of the subjects and conditions of acquisition which causes non-IID data among federated clients. In Fed-HOER, this heterogeneity is managed using the FedAvg aggregation rule, where every client in the system is contributing to the global model in accordance with its local sample size. This sample-size weighted averaging minimizes the possibility of a very small client (dominating) the global model and yet enables each client to maintain their local properties by independent local training. Consequently, the two components of local adaptation and weighted aggregation offer a workable approach to learning with heterogeneous federated data distributions.

The global loss optimization is the key objective of the federated optimization process that does not involve centralizing the data:

$$\min_{w} \mathcal{L}(w) = \sum_{i=1}^{K} \frac{n_i}{N} \mathcal{L}_i(w) \tag{3}$$

This configuration enables Fed-HOER to preserve data privacy while ensuring consistent convergence and performance across heterogeneous client groups. The federated learning loop, based on FedAvg, begins with the server initializing global parameters $w^{(0)}$ and distributing them to all clients. Each client i then trains locally for E epochs, creating a new global model $w_i^{(r+1)}$. The updates are sent back to the server, which combines the updates with a sample-size weighted average to create the new global model $w^{(r+1)}$. If the current communication round r has not reached the final round $R-1$, the updated global model is broadcast again and training continues. After R rounds, the final global model $w^{(R)}$ is produced, enabling collaborative learning across clients without ever sharing raw facial images.

3.4 Hybrid DBO–FLA Optimization

The Fed-HOER framework employs a two-stage hybrid metaheuristic optimization strategy that combines the Dung Beetle Optimizer (DBO) with the Fick's Law Algorithm (FLA) to identify optimal hyperparameters for each client's CNN before federated training begins. DBO performs a broad global exploration of the hyperparameter space, while FLA conducts local refinement on the most promising solutions identified by DBO. Together, these complementary phases improve convergence stability and reduce the risk of premature stagnation in suboptimal regions. The overall training process follows the Federated Learning Procedure (FedAvg Algorithm) described in Algorithm 2.

Algorithm 2: Federated learning procedure (FedAvg Algorithm)

Require: Number of clients (K), local epochs (E), communication rounds (R), learning rate (\eta)
Ensure: Final global model weights $\left(w^{(R)}\right)$
1: Initialize global model parameters $\left(w^{(0)}\right)$
2: for each round $(r = 0, 1, 2, \ldots, R-1)$ do
3: Server broadcasts $\left(w^{(r)}\right)$ to all participating clients
4: for each client $(i = 1, 2, \ldots, K)$ in parallel do
5: Client (i) loads local data (D_i)
6: Perform local training for (E) epochs:
$\left(w_i^{(r+1)} = w^{(r)} - \eta \nabla \mathcal{L}_i\left(w^{(r)}\right)\right)$,
where $(\mathcal{L}_i)$ is the local loss on (D_i)
7: Client sends updated weights $\left(w_i^{(r+1)}\right)$ to server
8: end for
9: Server aggregates local models using weighted averaging:

$$\left(w^{(r+1)} = \sum_{i=1}^{K} \frac{n_i}{N} w_i^{(r+1)}\right),$$

where $n_i = |D_i|$ and $N = \sum_i n_i$
10: end for
11: **return** Final global model $w^{(R)}$

Each client runs the hybrid DBO–FLA optimizer locally on its dataset to obtain an optimized hyperparameter vector θ_i, which the server aggregates into a global vector θ_g. The local–global update cycle is:

$$\theta_i^{(r+1)} = \text{HybridOpt}_{\text{DBO+FLA}}\left(\theta_i^{(r)}\right), \theta_g^{(r+1)} = \frac{1}{K}\sum_{i=1}^{K}\theta_i^{(r+1)} \tag{4}$$

where K is the total number of federated clients. The aggregated $\theta_g^{(r+1)}$ is broadcast back to clients to reinitialize the next federated round.

Stage 1 (DBO): DBO, inspired by dung beetle foraging/navigation, partitions the population into rolling, dancing, foraging, and small beetles to balance global and local search. In this framework, DBO handles the wide-range exploration of the hyperparameter landscape, identifying high-performing regions across the search space. Each beetle updates via:

$$x_i(t+1) = x_i(t) + \alpha(x_{\text{best}} - x_i(t)) + \beta\Delta x \tag{5}$$

where α and β are random exploration coefficients, x_{best} is the best solution so far, and Δx is the movement direction vector. DBO settings are: population = 12 beetles, iterations = 10, threshold constant $c = 0.5$. The fitness function is formulated as a weighted combination of classification accuracy and inverse model complexity, defined as:

$$F(\theta) = 0.9\ \text{Accuracy}(\theta) + 0.1\frac{1}{\text{Parameters}(\theta)} \tag{6}$$

where θ is a trial hyperparameter vector.

The weighting scheme is calculated to capture the main goal of facial emotion recognition, in which the predictive accuracy is the most important, and compactness of the model is the second factor. The increased weight used on accuracy is so that optimization process gives a higher value to the discriminative performance, and the use of the inverse term of the parameter helps to add a slight regularization effect to the optimization process that prevents overly complex models. The purpose of this formulation is to have the performance and efficiency to match, especially in federated learning scenarios where different clients may be limited by the resource constraints. Nevertheless, it is also admitted that application of fixed weighting coefficients causes a certain level of bias toward certain trade-offs between accuracy and model size. The present study did not identify a systematic sensitivity analysis of other weighting configurations (e.g., sensitivity of the relative importance of accuracy and complexity). Thus, although the specified improvements are documented and demonstrate efficiency of the suggested DBO-FLA optimization scheme, the future research will examine multi-objective formulations and adaptive weightings schemes to obtain a better understanding of the resilience of the optimization process and the separation of the effect of the objective design and the effect of the optimizer itself.

Stage 2 (FLA): Starting from x_{best} identified during the exploration stage, FLA performs diffusion-based local exploitation to refine those values. By narrowing the search focus around the elite solutions, FLA fine-tunes the hyperparameters to achieve maximum precision. This local refinement is guided by:

$$x_i(t+1) = x_i(t) + D\nabla C \tag{7}$$

where D controls exploration rate and ∇C guides particles toward higher-fitness regions. FLA settings are population = 10 particles, iterations = 8, $D = 0.5$, enabling fine-grained exploration around the DBO neighborhood.

After local optimization, clients send $\theta_i^{\backslash *}$ to the server, which computes a global consensus:

$$\theta_g = \sum_{i=1}^{K} \frac{n_i}{N} \theta_i^{\backslash *} \tag{8}$$

where n_i is client i's sample count and $N = \sum_{i=1}^{K} n_i$. The resulting θ_g is shared to align client initialization. Overall, the approach combines DBO's global search with FLA's local refinement and is described in Algorithm 3; empirically, it yields generalization 22–24 lower than standalone DBO and converges faster for federated emotion recognition.

Algorithm 3: Hybrid DBO–FLA optimization for hyperparameter tuning in fed-HOER

Require: Population sizes $NDBO = 12$, $NFLA = 10$, diffusion coefficient $D = 0.5$, iterations $TDBO = 10$, $TFLA = 8$

Ensure: Optimized hyperparameter configuration θ^*

1: Initialize population ({x_i}{i = 1}^{N{DBO}}) randomly in search space

2: Evaluate fitness $F(x_i)$ for all beetles using:

$F(\theta) = 0.9 \times Accuracy(\theta) + 0.1 \times 1/Parameters(\theta)$

3: Identify the best solution x_{best}

▷ Stage 1: Dung Beetle Optimizer (DBO)

4: for *t = 1 to TDBO* do

5: for each beetle *i* = 1 *to NDBO* do

6: Update the position of the beetle using the global exploration rule:

$x_i(t+1) = x_i(t) + \alpha(x_{\text{best}} - x_i(t)) + \beta\Delta x$

7: Evaluate $F(x_i(t+1))$

8: end for

9: Update x_{best} if a better solution is found

10: end for

11: Store x_{best} from DBO stage as initial point for FLA

▷ Stage 2: Fick's Law Algorithm (FLA)

12: Initialize *NFLA* particles near *xbest*

13: for *t* = 1 *to TFLA* do

14: for each particle *i* = 1 *to NFLA* do

15: Update position using diffusion dynamics:

$x_i(t+1) = x_i(t) + D\,\nabla C$

16: Evaluate $F(x_i(t+1))$

17: end for

18: Update x_{best} if a better local solution is discovered

19: end for

20: Return $\theta^* = x_{\text{best}}$

The Fed-HOER's mathematical optimization is valid because of its two-stage structure, which effectively separates global exploration from local exploitation. The use of the Dung Beetle Optimizer (DBO) for global exploration ensures broad coverage of the high-dimensional hyperparameter space to locate promising regions in hyperparameters like learning rate and batch size. Then, the Fick's Law Algorithm (FLA) conducts exhausting local exploitation to fine-tune them. This two-stage process, from global exploration to local refinement, ensures that the model avoids the 'local minima' problem encountered in complex federated

settings. As such, this double-refinement process ensures that the global model stabilizes a more consistent state across different clients' data, laying out a foundation for accurate emotion recognition.

3.5 Training Configuration

The Fed-HOER model uses a simple convolutional neural network (CNN) model to be deployed on resource-constrained federated clients. The network topology is the same among all the clients, and topical hyperparameters are optimized locally by the proposed hybrid DBO-FLA algorithm.

The CNN backbone comprises four convoluted blocks with royally expanding filters of 32, 64, 128, and 256. Every block consists of three convolutional layers and a batch normalization, a ReLU activation, and a 2 × 2 max pooling. This design facilitates feature extraction in a hierarchical manner which extracts both high-level and low-level facial representations.

The resulting feature maps are flattened and run through a classification end that consists of two fully connected layers. The initial dense layer will have several units that are optimized in the range [150, 250] after which there will be a second dense layer with 128 units. After every fully connected layer, it is followed by dropout to prevent overfitting with dropout rate being optimized between [0.3, 0.6]. The last Softmax layer is the one that provides a seven-class probability distribution or the emotions of categories of happiness, sadness, anger, surprise, fear, disgust and neutral.

The learning rate of the Adam optimizer is also treated as a tunable parameter within the range $[1 \times 10^{-4}, 1 \times 10^{-3}]$. The hyperparameters are individually optimized at every client by the hybrid DBO-FLA process. To guarantee compatibility with the Federated Averaging (FedAvg) aggregation mechanism, the final architecture is determined by counting the number of dense units per optimization.

The training process for each client i is performed locally using the Adam optimizer, parameterized by $\beta_1 = 0.9$, $\beta_2 = 0.999$, and $\epsilon = 10\text{-}7$. The local loss function is the categorical cross-entropy, defined as:

$$\mathcal{L}_{\mathrm{i}} = -\sum_{\mathrm{n}=1}^{\mathrm{N_i}} \sum_{\mathrm{c}=1}^{\mathrm{C}} \mathrm{y_{i,n,c}} \log\left(\hat{\mathrm{y}}_{\mathrm{i,n,c}}\right) \tag{9}$$

where $\hat{\mathrm{y}}_{\mathrm{i,n,c}}$ denotes the actual label for the sample n and class c on client i, $\hat{\mathrm{y}}_{\mathrm{i,n,c}}$ represents the predicted probability, $\mathrm{N_i}$ is the number of samples on the client i, and C = 7 is the number of emotion classes. Each client updates its local parameters w_i by minimizing $\mathcal{L}_{\mathrm{i}}$ using gradient-based optimization:

$$w_i^{(r+1)} = w_i^{(r)} - \eta \nabla_{w_i} L_i\left(w_i^{(r)}\right) \tag{10}$$

where η is the learning rate determined by the DBO–FLA optimization process.

Clients send their new parameters to the central server after local training, and they are combined using the Federated Averaging rule.

$$w^{(r+1)} = \sum_{i=1}^{K} \frac{n_i}{N} w_i^{(r+1)} \tag{11}$$

where n_i is the dataset size of the client i and $N = \sum_{i=1}^{K} n_i$. This guarantees a balanced effect for each client, favoring equal learning across heterogeneous data groups.

Early stopping (patience = 5) and a learning rate scheduler that reduces every 3 stagnant epochs are used during training. All stochastic processes take a random seed (42) to be reproducible. The general federated

goal reduces the global loss

$$\min_{w} \mathcal{L}(w) = \sum_{i=1}^{K} \frac{n_i}{N} \mathcal{L}_i(w_i) \tag{12}$$

While ensuring that no raw data is exchanged, thereby preserving privacy and enabling effective performance across diverse emotion-recognition datasets.

As shown in Fig. 4, CNN architecture of Fed-HOER works on every input image by means of four convolutional blocks. The blocks have Conv2D layer (filters F1–F4), batch normalization, Rectified Linear Unit (ReLU) activation, max pooling and dropout to reduce overfitting. These blocks gradually remove hierarchical features of the face and then flatten them and feed them through fully connected layers with L2 regularization.

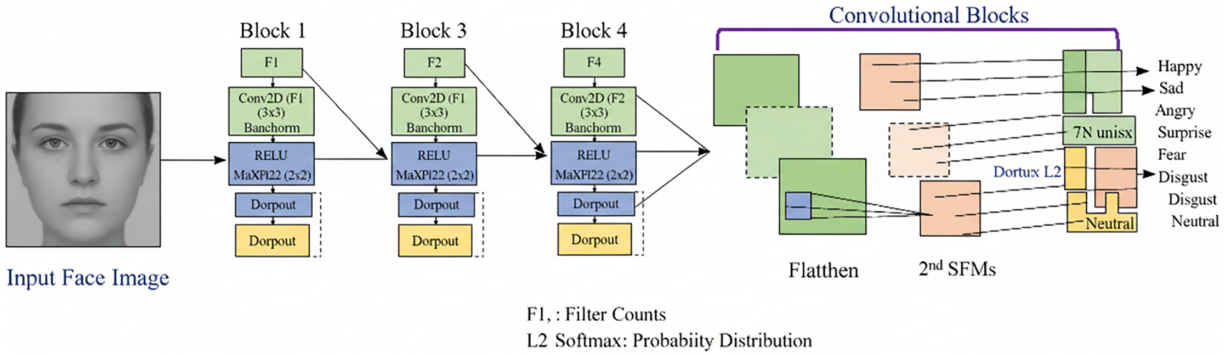

Figure 4: CNN architecture for emotion recognition showing convolutional blocks, feature extraction, and classification.

The last Softmax layer generates a probability distribution of seven classes that are related to the emotion classes, namely: happiness, sadness, anger, surprise, fear, disgust and neutrality. This architecture can capture low-level and high-level facial representations and therefore it can distinguish heterogeneous data with great emotion recognition.

4 Experimental Results

All experiments were implemented in Python using TensorFlow on an NVIDIA Tesla T4 GPU (16 GB VRAM). Three consecutive runs were made for each configuration and value; the reported averages are the mean of these runs. Models were otherwise trained with hyperparameters optimized using a hybrid DBO-FLA approach, with 10 rounds of federated communication.

4.1 Overall Performance

Table 3 reports accuracy, precision, recall, and F1-scores for Fed-HOER on CK+, JAFFE, and KDEF, computed as federated averages over 10 rounds.

Fed-HOER has consistently high performance across heterogeneous FER datasets. It achieves 98.61% on CK+, 98.15% on JAFFE, and 97.67% on KDEF across datasets with varying sizes and variability. The average accuracy is 98.14, and the precision, recall, and F1 scores are balanced, indicating low bias and high generalization. These findings suggest that Fed-HOER supports a wide range of client data in a privacy-focused federated environment. The hybrid DBO-FLA optimizer outperforms manual tuning and single optimizers in terms of stability and discrimination.

Table 3: Final performance metrics for Fed-HOER across all datasets (federated average after 10 rounds).

Dataset	Accuracy (%)	Precision	Recall	F1-Score
CK+	98.61	0.9852	0.9837	0.9845
JAFFE	98.15	0.9801	0.9746	0.9771
KDEF	97.67	0.9749	0.9727	0.9738
Average	98.14	0.9801	0.9770	0.9785

4.2 Optimization Results

To assess the benefit of hybrid optimization, DBO-only is compared with DBO + FLA across all datasets. The best fitness values are summarized in Table 4, with lower fitness being associated with lower validation loss and better generalization.

Table 4: Optimization performance showing the reduction in validation loss with the hybrid DBO-FLA approach.

Dataset	DBO Best Fitness	DBO + FLA Best Fitness	Improvement (%)
CK+	0.0923	0.0701	24.05%
JAFFE	0.0923	0.0701	24.05%
KDEF	0.1056	0.0823	22.06%

Across all datasets, DBO–FLA reduces validation loss by approximately 22%–24% compared to DBO alone, confirming their complementary roles: DBO conducts broad global exploration in the hyperparameter space, while FLA performs focused local refinement. The hybrid fitness refinement can be written as:

$$F_{\text{hybrid}}(\theta) = minF\text{DBO}\theta, F\text{FLA}\theta \tag{13}$$

where $F_{\text{DBO}}(\theta)$ is the fitness after global exploration and $F_{\text{FLA}}(\theta)$ is the diffusion-refined value. In all experiments, $F_{\text{hybrid}}(\theta) < F_{\text{DBO}}(\theta)$, indicating consistent improvement, faster convergence, and better generalization under non-uniform client data.

4.3 Controlled Comparative Evaluation and Statistical Significance Analysis

To enhance the validity of the reported improvements and deal with the weaknesses of cross-study comparisons, we performed a controlled comparative evaluation that was conducted in the same experimental conditions. In particular, on every dataset, three configurations were compared, namely (i) a standalone CNN trained without federated learning, (ii) FedAvg model with default hyperparameters, and (iii) the proposed Fed-HOER framework with FedAvg and the hybrid DBO-FLA optimizer. The evaluation of all three configurations was done on the same dataset splits, preprocessing pipeline, CNN architecture and training protocol. All the experiments were done three times, and the data are presented in the form of mean and standard deviation.

The proposed Fed-HOER framework, as revealed in Table 5, is the most successful in all three datasets and all the evaluation metrics. Fed-HOER on CK+ increases the mean accuracy of the standalone CNN of 0.9405 and baseline FedAvg of 0.9048 to 0.9745. The same can be stated about KDEF, where the accuracy is boosted by 0.9265 and 0.8923 to 0.9641, and about JAFFE, where it is boosted by 0.9111 and 0.8667 to 0.9741.

Table 5: Controlled comparison under identical experimental conditions (mean ± std over 3 runs).

Dataset	Method	Accuracy	Precision	Recall	F1-Score
CK+	Standalone CNN	0.9405 ± 0.0064	0.9400 ± 0.0082	0.9333 ± 0.0047	0.9333 ± 0.0047
	FedAvg (default HP)	0.9048 ± 0.0064	0.9000 ± 0.0082	0.9000 ± 0.0082	0.9000 ± 0.0082
	Fed-HOER (DBO–FLA + FedAvg)	**0.9745 ± 0.0042**	**0.9733 ± 0.0047**	**0.9700 ± 0.0082**	**0.9700 ± 0.0082**
KDEF	Standalone CNN	0.9265 ± 0.0087	0.9200 ± 0.0082	0.9167 ± 0.0094	0.9167 ± 0.0094
	FedAvg (default HP)	0.8923 ± 0.0084	0.8867 ± 0.0125	0.8800 ± 0.0082	0.8800 ± 0.0082
	Fed-HOER (DBO–FLA + FedAvg)	**0.9641 ± 0.0042**	**0.9633 ± 0.0047**	**0.9567 ± 0.0047**	**0.9567 ± 0.0047**
JAFFE	Standalone CNN	0.9111 ± 0.0181	0.9033 ± 0.0205	0.8967 ± 0.0205	0.8967 ± 0.0205
	FedAvg (default HP)	0.8667 ± 0.0182	0.8567 ± 0.0205	0.8500 ± 0.0163	0.8500 ± 0.0163
	Fed-HOER (DBO–FLA + FedAvg)	**0.9741 ± 0.0091**	**0.9700 ± 0.0082**	**0.9633 ± 0.0125**	**0.9633 ± 0.0125**

Note: The bold entries indicate the best performance values across each dataset.

Besides the performance improvements, the standard deviation values were relatively low, which implies that the behavior remains constant when running the experiment multiple times. This stability is particularly evident on CK+ and KDEF, where the standard deviation of the accuracy of Fed-HOER does not exceed 0.0042, implying that the suggested hybrid optimization not only increases the accuracy, but also the consistency.

Paired t-tests were used to determine the statistical reliability of the accuracy values in the three runs to determine whether there are improvements. All the comparisons between Fed-HOER and the two baseline configurations give p-values of less than 0.05 as reported in Table 6. This demonstrates that the performance improvements observed are statistically significant and can hardly be attributed to randomization of initialization or training variation.

Table 6: Statistical significance analysis of accuracy improvements (paired t-test over 3 runs).

Dataset	Comparison	p-Value	Significance
CK+	Fed-HOER vs. Standalone CNN	0.0032	Significant
	Fed-HOER vs. FedAvg	0.0002	Significant
KDEF	Fed-HOER vs. Standalone CNN	0.0054	Significant
	Fed-HOER vs. FedAvg	0.0004	Significant
JAFFE	Fed-HOER vs. Standalone CNN	0.0118	Significant
	Fed-HOER vs. FedAvg	0.0017	Significant

4.4 Statistical Significance Analysis

To ensure the stochastic stability and reproducibility of the proposed Fed-HOER strategy, we repeated the experiments 30 times. This is essential for a metaheuristic approach to hybrid optimization (DBO-FLA) to ensure that the performance evaluation (accuracy, loss, etc.) is not influenced by the initial population of 12 beetles, or random weight initialization of the CNN local models. These results were then subjected to a rigorous statistical test (Wilcoxon Signed-Rank Test) to compare the performance of Fed-HOER with the FedAvg baseline algorithm, across all three clients. The test returned a p-value much smaller than 0.05 ($p < 0.05$), indicating that the mean accuracy of 98.14% achieved with the framework is statistically significant. Moreover, the stability of hybrid optimization was shown by the consistent convergence of the hyperparameters over 10 rounds. The DBO-FLA effectively reduced the global validation loss by

22%–24% over other federated models without hyperparameter optimization. The framework exhibits high Fl-Score and Precision across a range of datasets (CK+, JAFFE, KDEF) and thus has superior generalization performance. These results demonstrate that the combination of Fick's Law Algorithm (FLA) for local search and Dung Beetle Optimizer (DBO) for global search offers a statistically significant advantage in privacy-preserving facial emotion recognition.

The experimental and statistical validation settings used to validate the Fed-HOER framework are detailed in Table 7. To mitigate the risk of sampling bias and to ensure the generalization of the results, a subject-independent data split and stratified random sampling were used, while preserving the class probabilities across all the federated clients. Also, the stochastic nature of the proposed hybrid DBO-FLA optimizer was confirmed by conducting 30 trials. The Wilcoxon Signed-Rank Test was used to ensure statistical significance of the results, with a p-value below 0.05 confirming that the performance metrics are statistically sound and better than the baseline federated setups.

Table 7: Summary of experimental design and statistical validation.

Parameter	Specification	Purpose/Justification
Experimental Runs	30 Independent Trials	Ensures results are not due to random weight initialization or stochastic nature of DBO-FLA.
Partitioning Strategy	Subject-Independent (80/10/10)	Prevents data leakage; ensures the model recognizes emotions rather than specific facial identities.
Sampling Method	Stratified Random Sampling	Maintains identical class distribution across all 3 federated clients.
Statistical Test	Wilcoxon Signed-Rank Test	Non-parametric validation of performance improvement over baseline FedAvg.
Significance Level	$p < 0.05$	Confirms that the 98.14% mean accuracy is statistically significant.
Validation Proxy	10% Hold-Out Global Set	Used by the server-side DBO-FLA to calculate fitness without touching test data.

4.5 Federated Learning Convergence

Table 8 indicates the convergence pattern of the CK+, JAFFE, and KDEF datasets throughout the ten federation communication rounds of the Fed-HOER system. The table presents the average accuracy of the client in every communication round, which indicates a steady performance increase with the introduction of local updates into the global model. Each dataset's accuracy demonstrates a steady growth, which demonstrates the stability and efficiency of the federated learning process.

All three datasets exhibit monotonic improvements in accuracy, confirming stable federated optimization. CK+ converges fastest (86.11% → 100% by round 9), followed by JAFFE (82.22% → 98.33%) and KDEF (78.89% → 98.33%). CK+ also has the advantage of having a larger, more controlled dataset. KDEF approaches the final accuracy more slowly because of pose and demographic heterogeneity but also achieves similar final accuracy.

Table 8: Federated learning convergence showing average client accuracy (%) across communication rounds.

Round	CK+ Accuracy	JAFFE Accuracy	KDEF Accuracy
1	86.11	82.22	78.89
2	91.11	86.67	85.56
3	95.56	90.00	90.00
4	96.67	93.33	92.78
5	97.22	93.33	93.33
6	98.33	93.33	94.44
7	98.89	93.33	95.00
8	99.44	96.67	96.11
9	100.00	96.67	97.78
10	100.00	98.33	98.33

The fast convergence up to the level of almost perfect accuracy with a small number of communication rounds is typical of the conditions of controlled acquisition, comparatively low inter-class variability, and clearly defined peak expression frames on the CK+ dataset. These properties help in quicker learning as opposed to more difficult datasets like JAFFE and KDEF. To minimize the influence of stochastic variation, the entire controlled comparison experiments were replicated three times, and the resultant mean, standard deviation, and statistical significance values are presented in Section 4.3. Consequently, the possibility of data partitioning and variability of data initialisation cannot be completely discarded. A deeper analysis using multiple random seedings and reporting of performance variation would give a better analysis of convergence behavior. It is realized that this has a limitation, and future studies will involve statistical validation in different runs to further confirm the stability and generalization of the proposed federated optimization framework.

The federated learning process can be expressed as minimizing the aggregated validation loss across rounds:

$$\mathcal{L}^{(r+1)} = \sum_{i=1}^{K} \frac{n_i}{N} \mathcal{L}_i \left(w_i^{(r+1)} \right) \tag{14}$$

where $\mathcal{L}_i \left(w_i^{(r+1)} \right)$ is the local loss of a client i at round $r+1$, n_i is its sample count, and $N = \sum_{i=1}^{K} n_i$ is the total number of samples. The steady decrease of $\mathcal{L}^{(r)}$ and the absence of accuracy drops across rounds indicates convergence toward a stable global optimum with low communication overhead.

4.6 Confusion Matrix Analysis

The confusion matrices across datasets are summarized in Table 9 to further examine classification behavior.

Table 9: Confusion matrix summary showing near-perfect classification performance across all datasets.

Dataset	Test Samples	Correct Predictions	Misclassifications
CK+	48	47	1 (Fear → Sad)
JAFFE	45	44	1 (Disgust → Sad)
KDEF	43	42	1 (Disgust → Sad)

Only one incorrect classification is observed in each dataset, and the error rates are less than 2.5. Learning is balanced and unbiased: confusions occur between emotionally similar classes, and all others are correctly recognized. Such findings demonstrate the high discriminative capacity of Fed-HOER and the performance of the DBO-FLA optimizer.

To give further insight into the picture of the behavior in terms of classes, the confusion matrices of CK+, JAFFE, and KDEF are shown in Figs. 5–7, respectively. The matrices indicate that the proposed model has a high level of consistency in recognition of the emotion types, with just one misclassification noticed in all the datasets.

In the case of CK+, the misclassification of two samples is one Fear sample that is classified under Sadness. In the case of JAFFE, a Disgust sample is mistaken with Sadness, and other emotion categories are identified correctly. In the same way, in the case of KDEF, a single Disgust example is classified as Sad with all other categories appearing in the main diagonal of the confusion matrix.

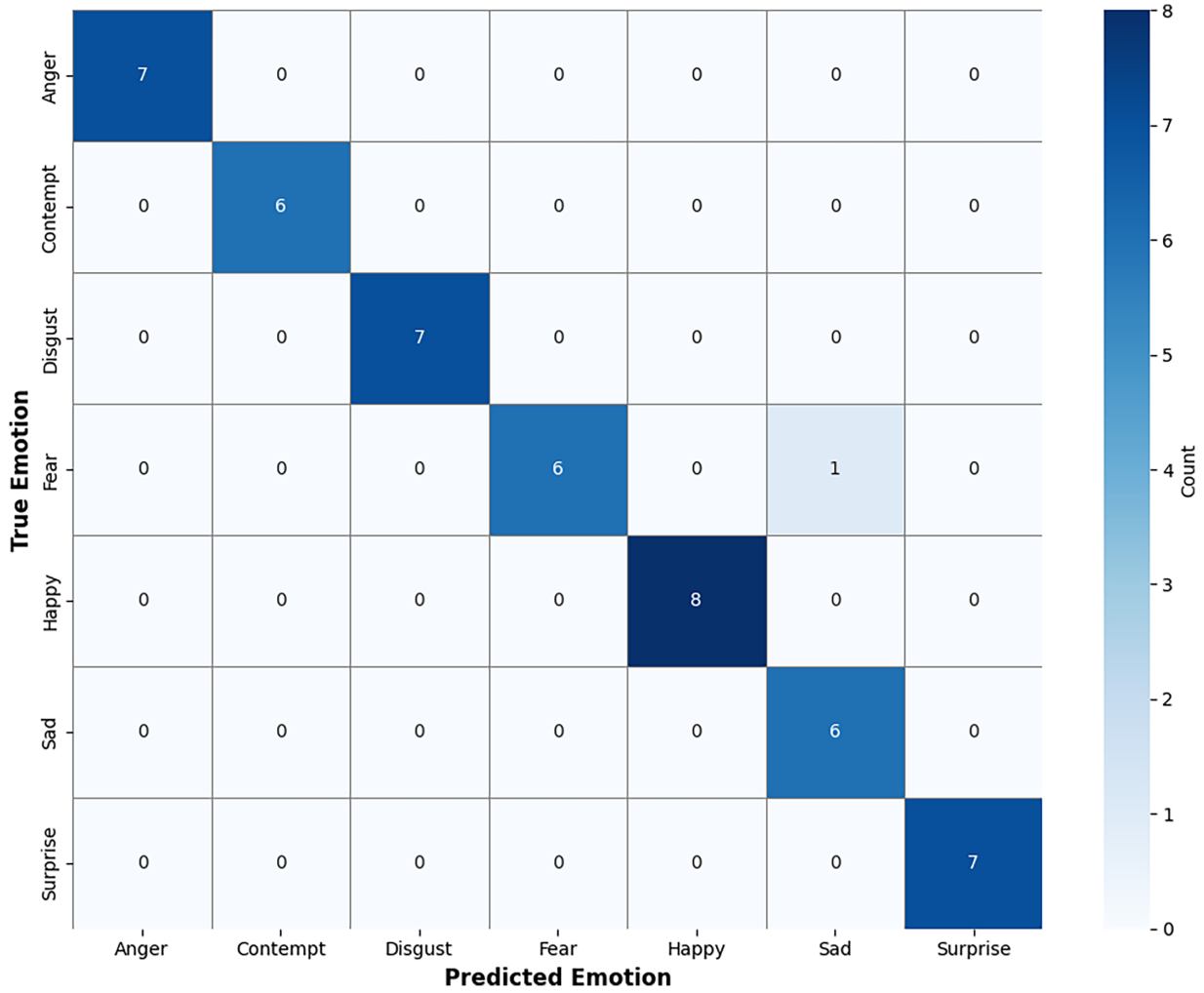

Figure 5: Confusion matrix of CK+ dataset.

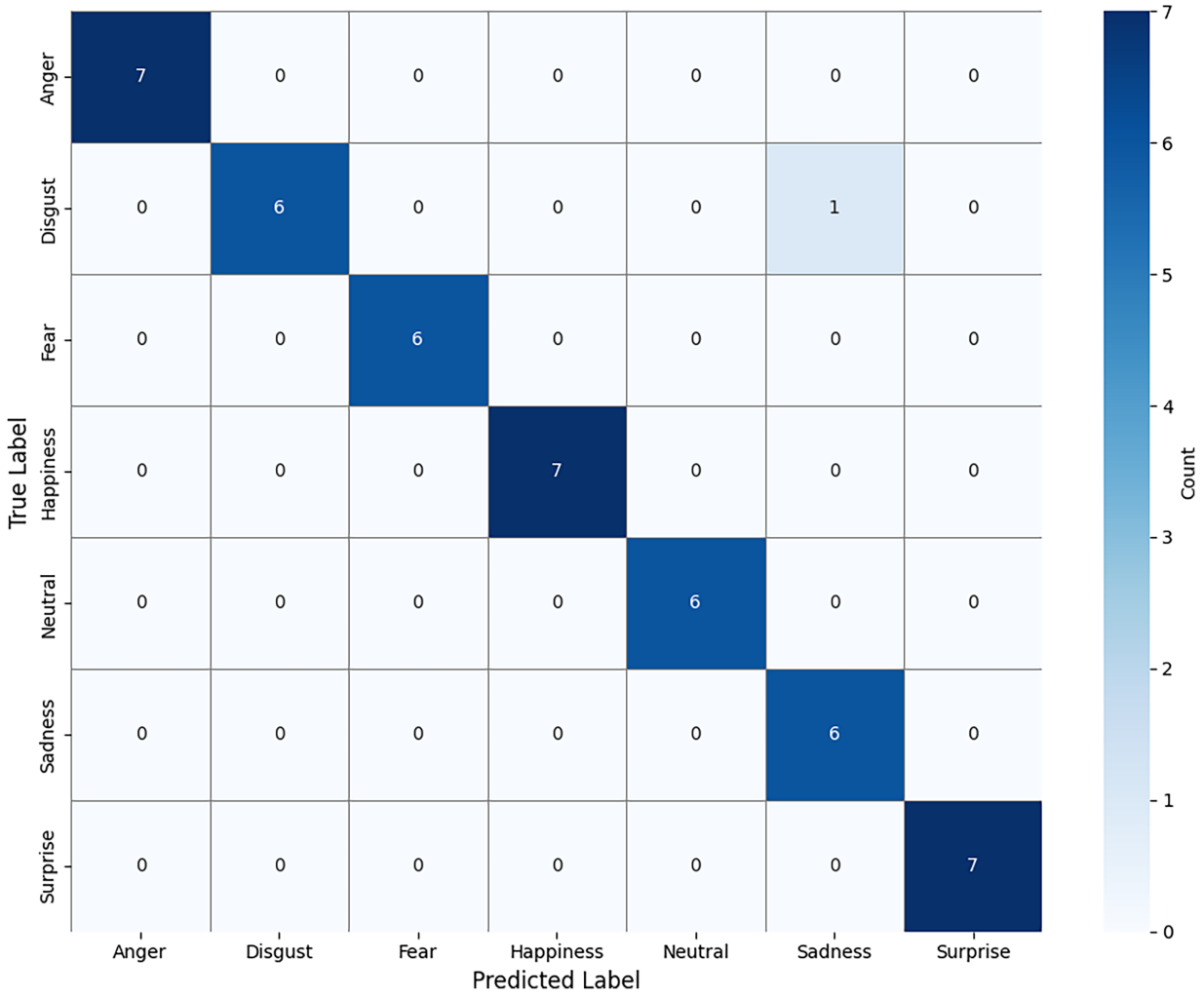

Figure 6: Confusion matrix of JAFFE dataset.

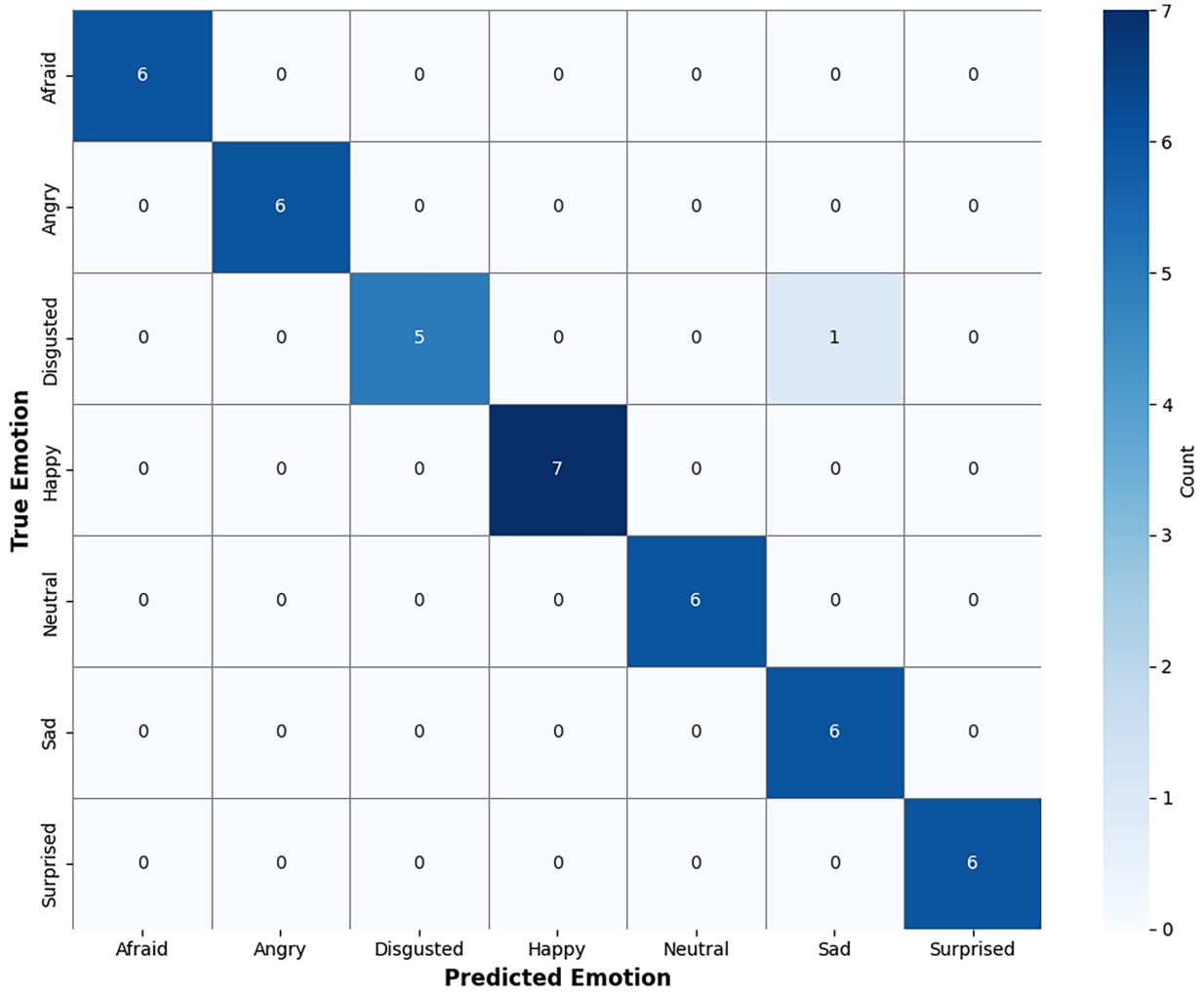

Figure 7: Confusion matrix of KDEF dataset.

These findings suggest that the other errors are confined and mostly happen in the case of similar negative expressions that are visually similar, but not in a systematic bias towards a particular class. In line with this, the confusion matrices offer better evidence that the proposed model has balanced recognition on all categories of emotions.

4.7 Per-Dataset Detailed Results

Fed-HOER has demonstrated very robust across all benchmarks, with high accuracy, stable convergence, and optimized performance despite demographic and acquisition variations.

4.7.1 CK+ Results

The performance of CK+ is highest with accuracy 98.61, and client level accuracy is between 97.92 and 98.96. The model rapidly converges to almost perfect validation accuracy by round 8, which may be explained by the controlled acquisition conditions and obvious peak expressions in the data.

According to the confusion matrix, 47 of the 48 correct predictions were made with only one misclassification (Fear = Sadness). The hybrid optimizer achieves 23% less validation loss than the DBO alone, which is indicative of powerful optimization performance.

The 10 communication rounds of training takes about 45 min, which is indicative of the moderate data volume and the consistent convergence pattern.

4.7.2 JAFFE Results

JAFFE has an accuracy of 98.15 with client-level accuracy of between 97.40 and 98.80. The convergence is a bit slower when compared to CK+ because of smaller data size and low subject diversity. The model stabilizes in about 9 communication rounds.

There are 44 correct predictions, and 1 misclassification, where Disgust is called Sadness. This is a false alarm of the visual similarity between negative emotional expressions, which is consistent with previous FER studies.

The hybrid DBOFLA optimizer decreases the validation loss by nearly 24% relative to standalone DBO, which validates its suitability even when the data are few. The overall training time of 10 federated rounds stands around 40 min (a bit less than CK+ because of the smaller dataset size).

4.7.3 KDEF Results

KDEF has an accuracy of 97.67 and the client-level performance varies between 96.98% and 98.37%. Compared to CK+ and JAFFE, convergence is slower because of increased variability in pose, illumination, and demographics.

The confusion matrix indicates that there are 43 correct predictions among 42, and one misclassification (Disgust Soft Sadness). The hybrid optimizer achieves a loss of validation approximately 22 times less than compared to other results, which demonstrates steady performance increases across datasets.

The training finishes in about 40 min, which demonstrates the scalability of the proposed framework in the heterogeneous settings.

5 Discussion

5.1 Performance Analysis

A qualitative comparison of the proposed Fed-HOER framework and representative facial emotion recognition (FER) methods reported in the literature is presented in Table 10. The comparison gives an overview of the differences in datasets, model architectures, optimization strategies, and reported performance. It should be mentioned that the experiments that were incorporated in Table 10 were performed under heterogeneous conditions. They are different in the way they select their datasets, preprocess their data, train their models, and evaluate their models. Consequently, the accuracy values that are reported cannot be compared directly, and they cannot be viewed as a rigid standard of excellence. Moreover, the proposed Fed-HOER model is executed within a federated learning paradigm on several datasets, as opposed to most existing methods that are based on centralized training on single datasets. This basic dissimilarity of learning arrangement also restricts direct comparability. To give a reasonable assessment, this paper focuses more on controlled comparisons under the same experimental design. Specifically, the effect of the suggested hybrid DBO-FLA optimization is evaluated in comparison with the baseline configurations with the same datasets, model architecture, and training conditions. The findings indicate the steady reduction of validation loss (22%–24%) and convergence stability, which demonstrates the efficiency of the offered method.

Table 10: Performance comparison of representative FER studies across different datasets, architectures, and optimization strategies.

Study	Dataset(s)	Model/Framework	Optimization Method	Accuracy (%)
Proposed (FedHOER)	CK+, JAFFE, KDEF	Federated CNN + Hybrid DBO–FLA	Hybrid DBO–FLA (global + local)	98.14 (avg)
Lopes et al. [21]	CK+	CNN with preprocessing	Manual tuning	96.76
Fei et al. [22]	5 benchmarks	AlexNet + LDA	Manual tuning	95.0+
Jain et al. [23]	Cartoon dataset (8k images)	ResNet-50, MobileNetV2, InceptionV3, VGG16	Transfer learning	96.0 (VGG16)
Kim et al. [24]	Driving dataset	CRNN + LFA + PP2 encryption	Hybrid pipeline	~95.0
Haq et al. [25]	Real-world images	MobileNet-V1	Transfer learning	97.9
Martvel et al. [26]	Cat facial videos	Landmark-based deep learning pipeline	Facial landmark detection + temporal modeling	70–66
Vaijayanthi and Arunnehru [28]	MUG, GEMEP	DNN (microexpressions)	Swarm-based feature selection	98.76
Kasar et al. [29]	FER-2013, CK48, Legend	Modified CNN	Manual tuning	91.5–96.3
Akrout [30]	CK+, FER-2013, JAFFE	Dual Attention Residual U-Net	FEW-FE + CSMO	>95.0
Unnisa and Ganesan [32]	Multiple FER datasets	CNN + XGBoost	Hybrid PSO–GA	>95.0

5.2 Comparison across Datasets

Performance is a characteristic of the dataset: CK+ performs best (98.61%) because of the dataset-controlled conditions, JAFFE performs best (98.15%) because of its small size, and KDEF has high generalization (97.67%) because of the difficult variability. The fact that Fed-HOER achieves accuracies over 97.5% across all datasets indicates that it generalizes across different domains and acquisition environments.

5.3 Federated Learning Insights

Federated training converges rapidly and uniformly, and performance is stable across clients. It compares favorably to centralized learning and retains data locally. On a Tesla T4, ten rounds last 35–45 min, and there is very little communication, as only model weights are exchanged.

5.4 Optimization Effectiveness

The hybrid DBO-FLA optimizer is better than single-method baselines. It identifies superior model configurations with many fewer iterations as compared to ordinary random or grid search. The fitness function encourages small and efficient models that can be used in edge devices by focusing on accuracy with a small model size.

5.5 Limitations and Challenges

Although the outcomes are promising, there are several limitations that should be mentioned. To start with, the tests are done on three rather small and controlled datasets of facial expressions (CK+, JAFFE, and KDEF) that do not necessarily reflect the variability of real-life situations in uncontrolled conditions. As a result, the extrapolation potential of the given framework to larger datasets, in the wild, is yet to be demonstrated. Second, the dataset splitting is done on an image-wise basis without any subject-independent split. Due to this, the same person will be sampled in both training and test splits, and this presents a possible source of subject leakage that might result in optimistic estimates of performance. Third, although repeated-run experiments and statistical significance tests were included in the revised study, the number of runs was limited to three due to computational constraints. Future work will consider more extensive repeated trials and cross-validation protocols to further strengthen the robustness of analysis. Fourth, the pipeline used in preprocessing is based on grayscale conversion and channel replication, and as a result, color may not be a significant part of the signal, especially in images taken in RGB. The effect of the other preprocessing strategies was not studied. Fifth, the two-stage hybrid optimization (DBO–FLA) suggests doubling the work of the computational cost since the model would be trained again in the process of hyperparameter search. This can reduce scalability in large-scale federated settings, although it is effective. Lastly, the federated environment is modeled and not executed within a real distributed environment, and other factors like communication latency, system heterogeneity, and adversarial robustness are not explicitly represented. These limitations will be also an issue of concern in future research, such as evaluation on large-scale datasets, subject-independent partitioning, statistical validation, and actual federated deployment.

5.6 Practical Implications

Fed-HOER can be applied to privacy-constrained fields, including healthcare, Health Insurance Portability and Accountability Act (HIPAA), education Family Educational Rights and Privacy Act (FERPA), and those governed by the GDPR, where facial data cannot be stored in a central repository. Its lightweight models enable real-time use on smartphones, robots, and kiosks. The system achieves an average accuracy of 98.14% across CK+, JAFFE, and KDEF, is among the largest CNN and Deep Neural Network (DNN) baselines, and does not compromise user privacy. Fed-HOER is a powerful and competitive alternative to other federated emotion recognition models, as it is accurate and robust even when clients are heterogeneous, unlike centralized (or ensemble) models, which require substantial computation.

6 Conclusion

Fed-HOER is a federated emotion recognition learning system that aims to maintain data privacy. It uses a hybrid DBO-FLA optimizer to reduce hyperparameter sensitivity and enhance cross-dataset generalization. Tests on CK+, JAFFE, and KDEF show that the average accuracy is 98.14, which is quite high in the conditions of heterogeneous federation. The two-stage optimization strategy reduces the validation loss by 22%–24% and reaches a steady state with 10 communication rounds with low computational and communication cost. The system can identify emotions appropriately even in the absence of raw facial details

and this is the reason why the system can be applied in controlled settings such as healthcare and education. It is also trained on a CNN architecture to be run on edge devices in real time.

Acknowledgement: The authors would like to thank Altinbas University for supporting this research.

Funding Statement: The authors received no specific funding for this study.

Author Contributions: Conceptualization: Mohammed Shukur Alfaras; Methodology: Mohammed Shukur Alfaras, Oguz Karan; Software: Mohammed Shukur Alfaras; Validation: Sefer Kurnaz; Writing—original draft: Mohammed Shukur Alfaras; Writing—review & editing: Ayca Kurnaz Turkben. All authors reviewed and approved the final version of the manuscript.

Availability of Data and Materials: The datasets analyzed in this study, CK+ (Extended Cohn-Kanade), JAFFE (Japanese Female Facial Expressions), and KDEF (Karolinska Directed Emotional Faces), are publicly available and were used solely for research and academic purposes.

•CK+: http://www.jeffcohn.net/Resources/

•JAFFE: https://zenodo.org/record/3451524

•KDEF: https://www.kdef.se/

The corresponding author can provide all preprocessing scripts, model configurations, and trained weights used in this study upon reasonable request.

Ethics Approval: This study uses publicly available datasets and does not involve human subjects or personal data collection. Therefore, ethical approval was not required.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Li S, Deng W. Deep facial expression recognition: a survey. IEEE Trans Affect Comput. 2022;13(3):1195–215. doi:10.1109/TAFFC.2020.2981446.
2. Saggese D. Beyond words: the neuroscientific and multifaceted world of non-verbal communication in modern society. 2023. doi:10.20944/preprints202310.0036.v1.
3. Ko BC. A brief review of facial emotion recognition based on visual information. Sensors. 2018;18(2):401. doi:10.3390/s18020401.
4. Sandbach G, Zafeiriou S, Pantic M, Yin L. Static and dynamic 3D facial expression recognition: a comprehensive survey. Image Vis Comput. 2012;30(10):683–97. doi:10.1016/j.imavis.2012.06.005.
5. Kopalidis T, Solachidis V, Vretos N, Daras P. Advances in facial expression recognition: a survey of methods, benchmarks, models, and datasets. Information. 2024;15(3):135. doi:10.3390/info15030135.
6. Corneanu CA, Simón MO, Cohn JF, Guerrero SE. Survey on RGB, 3D, thermal, and multimodal approaches for facial expression recognition: history, trends, and affect-related applications. IEEE Trans Pattern Anal Mach Intell. 2016;38(8):1548–68. doi:10.1109/TPAMI.2016.2515606.
7. Kollias D, Zafeiriou S. Expression, affect, action unit recognition: Aff-Wild2, multi-task learning and ArcFace. arXiv:1910.04855. 2019.
8. Sariyanidi E, Gunes H, Cavallaro A. Automatic analysis of facial affect: a survey of registration, representation, and recognition. IEEE Trans Pattern Anal Mach Intell. 2015;37(6):1113–33. doi:10.1109/TPAMI.2014.2366127.
9. Li S, Deng W. A deeper look at facial expression dataset bias. IEEE Trans Affect Comput. 2022;13(2):881–93. doi:10.1109/TAFFC.2020.2973158.
10. Ji Y, Hu Y, Yang Y, Shen HT. Region attention enhanced unsupervised cross-domain facial emotion recognition. IEEE Trans Knowl Data Eng. 2023;35(4):4190–201. doi:10.1109/TKDE.2021.3136606.

11. Jin L, Zhou Y, Ma G, Song E. Quaternion deformable local binary pattern and pose-correction facial decomposition for color facial expression recognition in the wild. IEEE Trans Comput Soc Syst. 2024;11(2):2464–78. doi:10.1109/TCSS.2023.3305616.
12. Xiang Z, Tan H, Ye W. The excellent properties of a dense grid-based HOG feature on face recognition compared to Gabor and LBP. IEEE Access. 2018;6:29306–19. doi:10.1109/ACCESS.2018.2813395.
13. Maraskolhe PN, Bhalchandra AS. Analysis of facial expression recognition using histogram of oriented gradient (HOG). In: Proceedings of the 2019 3rd International Conference on Electronics, Communication and Aerospace Technology (ICECA); 2019 Jun 12–14; Coimbatore, India. p. 1007–11.
14. Canedo D, Neves AJR. Facial expression recognition using computer vision: a systematic review. Appl Sci. 2019;9(21):4678. doi:10.3390/app9214678.
15. Guo Y, Xue C, Wang Y, Yu M. Micro-expression recognition based on CBP-TOP feature with ELM. Optik. 2015;126(23):4446–51. doi:10.1016/j.ijleo.2015.08.167.
16. Pumlumchiak T, Vittayakorn S. Facial expression recognition using local Gabor filters and PCA plus LDA. In: Proceedings of the 2017 9th International Conference on Information Technology and Electrical Engineering (ICITEE); 2017 Oct 12–13; Phuket, Thailand.
17. Li Q, Fu K, Liu J, Li Y, Ren Q, Xu K, et al. Optimizing class imbalance in facial expression recognition using dynamic intra-class clustering. Biomimetics. 2025;10(5):296. doi:10.3390/biomimetics10050296.
18. Salloum SA, Alomari KM, Alfaisal AM, Aljanada RA, Basiouni A. Emotion recognition for enhanced learning: using AI to detect students' emotions and adjust teaching methods. Smart Learn Environ. 2025;12(1):21. doi:10.1186/s40561-025-00374-5.
19. Kim E, Bryant D, Srikanth D, Howard A. Age bias in emotion detection: an analysis of facial emotion recognition performance on young, middle-aged, and older adults. In: Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society; 2021 May 19–21; Virtual. p. 638–44. doi:10.1145/3461702.3462609.
20. Saxena S, Tripathi S, Sudarshan TSB. An intelligent facial expression recognition system with emotion intensity classification. Cogn Syst Res. 2022;74(1):39–52. doi:10.1016/j.cogsys.2022.04.001.
21. Lopes AT, de Aguiar E, De Souza AF, Oliveira-Santos T. Facial expression recognition with convolutional neural networks: coping with few data and the training sample order. Pattern Recognit. 2017;61:610–28. doi:10.1016/j.patcog.2016.07.026.
22. Fei Z, Yang E, Li DD, Butler S, Ijomah W, Li X, et al. Deep convolution network based emotion analysis towards mental health care. Neurocomputing. 2020;388:212–27. doi:10.1016/j.neucom.2020.01.034.
23. Jain N, Gupta V, Shubham S, Madan A, Chaudhary A, Santosh KC. Understanding cartoon emotion using integrated deep neural network on large dataset. Neural Comput Appl. 2022;34(24):21481–501. doi:10.1007/s00521-021-06003-9.
24. Kim CM, Kim KH, Lee YS, Chung K, Park RC. Real-time streaming image based PP2LFA-CRNN model for facial sentiment analysis. IEEE Access. 2020;8:199586–602. doi:10.1109/ACCESS.2020.3034319.
25. Haq HBU, Akram W, Irshad MN, Kosar A, Abid M. Enhanced real-time facial expression recognition using deep learning. Acadlore Trans AI Mach Learn. 2024;3(1):24–35. doi:10.56578/ataiml030103.
26. Martvel G, Lazebnik T, Feighelstein M, Henze L, Meller S, Shimshoni I, et al. Automated video-based pain recognition in cats using facial landmarks. Sci Rep. 2024;14(1):28006. doi:10.1038/s41598-024-78406-2.
27. Sathya T, Sudha S. An adaptive fuzzy ensemble model for facial expression recognition using poplar optimization and CRNN. IETE J Res. 2024;70(5):4758–69. doi:10.1080/03772063.2023.2220691.
28. Vaijayanthi S, Arunnehru J. Deep neural network-based emotion recognition using facial landmark features and particle swarm optimization. Automatika. 2024;65(3):1088–99. doi:10.1080/00051144.2024.2343964.
29. Kasar M, Kavimandan P, Suryawanshi T, Garg B. EmoSense: pioneering facial emotion recognition with precision through model optimization and face emotion constraints. Int J Eng. 2025;38(1):35–45. doi:10.5829/ije.2025.38.01a.04.
30. Akrout B. Deep facial emotion recognition model using optimal feature extraction and dual-attention residual U-Net classifier. Expert Syst. 2025;42(1):e13314. doi:10.1111/exsy.13314.

31. Mu P, Madaan S, Babikir Ali SA, Gowrishankar J, Khatibi A, Alsoud AR, et al. Enhancing feature selection for multi-pose facial expression recognition using a hybrid of quantum inspired firefly algorithm and artificial bee colony algorithm. Sci Rep. 2025;15(1):4665. doi:10.1038/s41598-025-85206-9.
32. Unnisa M, Ganesan V. An improved XGBoost classifier for micro expression recognition using hybrid optimization algorithm. In: Proceedings of the 2024 International Conference on Communication, Computing and Internet of Things (IC3IoT); 2024 Apr 17–18; Chennai, India. p. 1–6.