Automatika

Journal for Control, Measurement, Electronics, Computing and Communications

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/taut20

Deep neural network-based emotion recognition using facial landmark features and particle swarm optimization

S. Vaijayanthi & J. Arunnehru

To cite this article: S. Vaijayanthi & J. Arunnehru (2024) Deep neural network-based emotion recognition using facial landmark features and particle swarm optimization, Automatika, 65:3, 1088-1099, DOI: 10.1080/00051144.2024.2343964

To link to this article: https://doi.org/10.1080/00051144.2024.2343964

Published online: 22 Apr 2024.

AUTOMATIKA
2024, VOL. 65, NO. 3, 1088–1099
https://doi.org/10.1080/00051144.2024.2343964

Taylor & Francis
Taylor & Francis Group

OPEN ACCESS Check for updates

Deep neural network-based emotion recognition using facial landmark features and particle swarm optimization

S. Vaijayanthi and J. Arunnehru

Department of Computer Science and Engineering, SRM Institute of Science and Technology, Vadapalani Campus, Chennai, India

ABSTRACT

Relating specifically to human–computer interaction (HCI), computer vision research has placed a substantial emphasis on intelligent emotion recognition in recent years. The primary emphasis lies in investigating speech aspects and bodily motions, while the knowledge of recognizing emotions from facial expressions remains relatively unexplored. Automated facial emotion detection allows a machine to assess and understand a person's emotional state, allowing the system to predict intent by analyzing facial expressions. Therefore, this research provides a novel parameter selection strategy using swarm intelligence and a fitness function for intelligent recognition of micro emotions. This paper presents a novel method based on geometric visual representation obtained from facial landmark points. We employ the Deep Neural Networks (DNN) model to analyze the input features from the normalized angle and distance values derived from these landmarks. The results of the experiments show that Particle Swarm Optimization (PSO) worked very well by using only a few carefully chosen features. The method achieved a recognition success rate of 98.76% on the MUG dataset and 97.79% on the GEMEP datasets.

ARTICLE HISTORY
Received 19 December 2023
Accepted 11 April 2024

KEYWORDS
Deep neural network; particle swarm optimization; facial motion recognition; geometrical features

1. Introduction

Emotion is one of the primary reliable modes for identifying a person's mental state in daily communication. Recent research on facial studies indicates that emotional cues can accurately predict non-vocal communicational inference. The imperative way of conveying human emotions is through eyes, vocal speech, facial gestures and body postures [1]. The state of one's eyes and a collection of interrelated articulations can reveal and express people's feelings in depth. For instance, in daily communication, vocal tone conveys 38% of the data [2], whereas facial expressions transmit 55% of the total data. As a result, facial expression acts as an essential non-verbal communication and a versatile instrument in revealing the person's cognitive state, including physical form, discomfort, anxiety and degree of attentiveness in expressing the emotions of humans in today's modern world. Suspicious human action recognition to detect and alarm security guards, universal health care services and autistic patient assistance [3] are just a few application areas for automatic emotion recognition using facial gestures [4].

A cross-cultural study shows that people experience certain fundamental emotions uniformly, irrespective of their cultural dependence. Automatic facial expression and gesture recognition [5] and [6] have recently undergone rigorous study by cognitive scientists and computer vision researchers. The implementation of the Facial Action Coding System (FACS) began after observing facial intensity early in the twentieth century. The research showed that fear, happiness, anger, sadness, surprise and disgust are uniformly linked to certain emotions, providing a biological foundation for facial expressions. FACS [7] represents the facial definition and animation parameters and describes the variations in facial muscle movements in 44 different Action Units (AUs). Influencing FACS from image or video sequences is very laborious work with geometric feature approaches by determining the facial landmarks. Recent improvements in the human emotion recognition system include annoyance, anxiety and mixed compound emotions [8].

The Facial Emotion Recognition (FER) system typically includes facial action units and facial behaviours for displaying deep expressive emotions. Facial action units [9] encircle the eyebrow raiser, nose wrinkler, chin raiser, dimpler, cheek puffer and lip corner. In contrast, facial behaviour predicts facial muscle movements with various facial deformations identified by the multiple viewpoints of changes in facial appearance. Different positions and trends of face parts will cause different changes in how the face moves, showing different emotional looks with different affine texture warpings, depending on the face's appearance. Recently, researchers have undergone many levels in extracting spatiotemporal information from expressive

CONTACT S. Vaijayanthi vaijayanthisekar@gmail.com Department of Computer Science and Engineering, SRM Institute of Science and Technology, Vadapalani Campus, Chennai, Tamil Nadu 600026, India

features hidden inside the face component detection. This information can be static or dynamic, highlighting the geometric and transient appearance features[10]. The following are our research's novel contributions:

- The objective of the research lies in identifying the micro emotions that express subtle movement changes in facial action units in less than a second, focusing on temporary duration and minimal motion by calculating the distance and angle measured at facial feature points.
- We presented a 66-dimensional face point using a deep neural network model that was constructed by optimizing parameter selection using Particle Swarm Optimization (PSO) to predict the micro emotions.
- Corresponding to previous cutting-edge results, with the selected set of features, our proposed approach attains 98.7% accuracy on the MUG dataset and 97.7% accuracy on the GEMEP dataset, respectively.

The present work is summarized as follows: the next section elaborates on the previous views on the geometric extraction of facial features. Section 3 discusses the major work of two different datasets in detail, while the proposed feature extraction process and the DNN framework are discussed in Section 4. Section 5 describes the interpretation of results and performance metrics on the MUG and GEMEP datasets, with a comparative study, and lastly, Section 6 presents the conclusion.

2. Prior work on geometric face features

This segment provides an overview of the latest research methodologies used for recognizing facial emotions based on visual cues. In general, the process of facial emotion recognition includes identifying the presence of a face within the frame, extracting distinct characteristics and categorizing the displayed emotions [11]. The first step is preprocessing, which involves enhancing the facial expressions by applying contrast adjustment, image scaling and additional enhancements. Previous studies primarily concentrated on detecting the peak expression occurrence in the Action Units (AUs), measuring the intensity of AUs and identifying the facial characteristic points for better classification of emotions. The preceding approaches involve visual pattern approaches, mainly composed of recognizing facial images based on geometric traits like appearance, texture and hybrid methods.

Aifanti et al. [12] proposed distance manifolds representing the minimum parameters as the 70 landmark points in the MUG dataset. They recognized the linear subspaces of various feature points regarding variations in the six basic emotions. The subject-dependent and independent accuracy performance comparison is made with the CK + dataset using SVM (RBF kernel) classification. Chickerur et al. [13] introduced a Parallel-Scale Invariant Feature Transform algorithm (PSIFT); it sets a key point detection to determine the expression identified in the input Indian and Japanese database containing 2500 image samples and features extraction done via Euclidean distance matching by dividing the task into subtasks with multiple processors. Finally, the active appearance model classifies the images for emotion recognition.

Fathallah et al. [14] recognize facial emotions using convNet by adding VGG Net to the architecture, enhancing the accuracy of various pooling and fully connected layers with CK+, MUG and RAFD datasets. The author does not provide any information about the preprocessing of the image samples. Zadeh et al. [15] propose a Convolution framework for classifying seven emotional types in the JAFFE database. He improved the system learning rate with 2 Gabor filters in the feature extraction phase and presented an accuracy of 87.5 with a sample size of 213 images. Verma et al. [16] introduce two-level hybrid features using a CNN to capture edge variational expressions in the facial appearance. He analyzed the discriminative features from eye positions, nose, lips and mouth regions, comparing the four datasets: MUG, CK+, OULU and AFEW. The paper also highlights the visual response of the specific features with multi-convolution.

Ravi et al. [17] presented a paper with Local Binary Patterns and CNN. This paper deals with thresholding the eight neighbourhood pixels to their binary equivalent. The SVM classifier helps recognize the expression of an image sequence with three different datasets named CK+, JAFFE and YALE FACE. The detailed summary of various evaluation metrics was missing in the paper and provided that the accuracy of JAFFE is 73.81%, significantly less than the other two datasets. Jude Hemanth et al. [18] recognize six different facial emotions using the CK + dataset utilizing the initial pre-processing for image resizing for the pre-trained feature extraction process of Convets are carried with different transfer learning architectures like VGG-19, Resnet 50, Mobile Net and Inception V3 and achieved an accuracy of 98.5%. Ali I. Siam et al. [19] propose keypoint generation with a Media pipe face mesh algorithm for the selection of facial key points using angular encoding modules and the final decomposed features are passed through different algorithms like KNN, LR, SVM, NB, RF and the model used MLP to boost the accuracy upto 97%. Fatma M. Talaat et al. [20] propose an attention-based learning model for early diagnosis of Autism spectrum disorder in developing children with the help of distinctive patterns in the CIFAR dataset. The model is trained with six different facial emotions in Deep CNN with autoencoder and attained an accuracy of 95%.

Table 1. Prior work on Facial emotion recognition.

Year	Dataset		Features	Classification	Merits	Demerits
	Name	Samples				
2014 [12]	MUG & CK+	458	70 Landmark Points	SVM	An individual set of images for training and testing	Pre-processing is not done
2015 [13]	Indian & JAFFE	2500	Sub Task	P-SIFT	Addressed all approaches	Less accuracy compared to other approaches
2018 [14]	MUG & CK + & RAFD	3276	Face	CNN	Addressed CNN cons	Not efficient for real-time applications
2019 [15]	JAFFE	213	Entire face	CNN	Increased system performance	Less no of samples
2019 [16]	CK+, MUG, OULU, RAFD	N/A	Facial Features	HI-Net	Visual saliency	NIL
2020 [17]	CK + & JAFFE and YALE FACE	N/A	Face	CNN	3 Different dataset	Low accuracy
2021 [18]	CK+	148	Face	CNN	NIL	Less sample size
2022 [19]	CK+, JAFFE, RAFD	239	Facial Landmark	MLP + DNN	Increased Accuracy	NIL
2023 [20]	CIFAR	72	Entire Face	DCNN	NIL	Sample size can be increased

Basedon the literature findings, a comprehensive work on previous recognition techniques in discussed in Table 1. the research focuses only on the primary six facial expressions, using artificial learning algorithms, but there seems to be a missing element in inheriting the micro emotions. This study aims to tackle the issue by introducing a new method that optimizes the hyper-parameters of the deep neural network model using particle swarm optimization. The strategy effectively emphasizes the most subtle facial regions of interest for achieving better model performance. The concept of picking the PSO algorithm in emotion recognition requires unconstrained optimization; thus, by examining a variety of optimization techniques, we concluded that.

- An exhaustive search mechanism can perform parameter tuning. However, it takes much time to converge, whereas, in evolutionary algorithms like genetic, ant colony and swarm intelligence, the algorithm can search the optical parameter quickly.
- The algorithm for Ant colonies converges to a locally optimal solution. The accuracy and convergence speed of the approach are not comparatively higher than those of other evolutionary algorithms.
- Particle Swarm Optimization has the highest computational efficiency compared to the genetic algorithm, especially when dealing with non-linear and unconstraint problems. In contrast, the Genetic algorithm performs well in constraint problems. So, leveraging the advantage of both algorithms, a particle swarm algorithm which uses a genetic algorithm operator is developed to solve our constraint problem.
- The Novelty of the proposed Particle Swarm algorithm utilizes genetic operators to gain the advantage of the genetic algorithm.

3. Emotional databases

This work analyzes two facial expression databases: (a) MUG and (b) GEMEP, wherein each image or video sequence exclusively depicts a single face.

3.1. The MUG database

MUG is a laboratory-controlled facial expression dataset, known as the Multimedia Understanding Group [21], containing seven fundamental emotion prototypes as specified in the FACS manual. The dataset includes 1462 facial sequences with different action units from 86 subjects expressed by 35 females and 51 males aged 20–35 at 19 frames per second. In total, 11,758 images were collected from the preprocessed sequences.

Figure 1 discloses the seven fundamental instances of facial expressions used for geometric feature point extraction. The authentic face cues in this database outperform the drawbacks of previous comparable FER datasets, such as the partially occluded face, illumination variations and multiple takes per person.

3.2. The GEMEP Corpus

The Geneva Multimodal Expressional Corpus [22] is a multimodal framework designed based on expressive modalities by Klaus Scherer and Tanja Banziger. Ten actors, each leveraging a unique modality concerning facial features, speech and bodily posture video

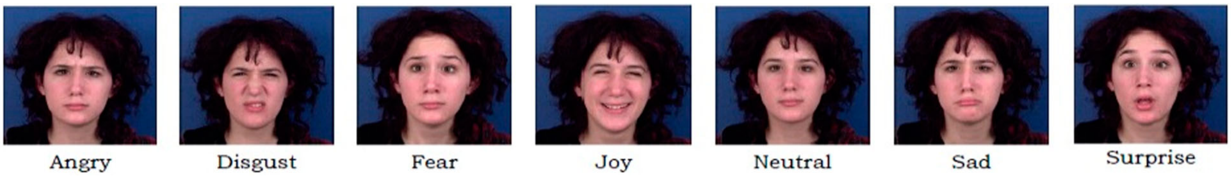


Figure 1. MUG database basic expressions with one sample image.

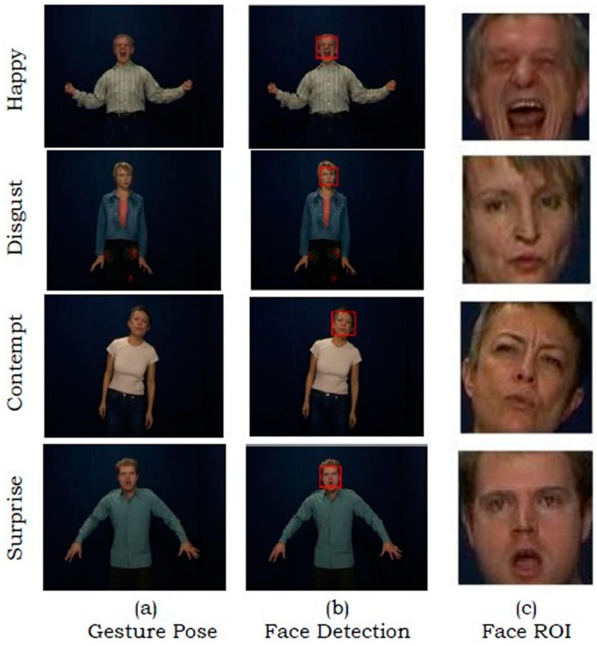


Figure 2. GEMEP Corpus with four sample frames.

sequences for the GEMEP dataset, It has 17 micro-coded emotional states like admiration, anxiety, amusement, anger, contempt, disgust, joy, despair, fear, irritation, interest, pleasure, pride, sadness, relief, surprise and tenderness.

The initial preprocessing was performed on 1823 instances for our experimental studies, wherein each video depiction was converted into still images. Thereby, with the cascade function, face detection is achieved and the face region of interest (ROI) is cropped for our research. Figure 2 (a) depicts the body gesture pose, (b) shows the face detection in the sample frames and (c) finally depicts the trimmed visage.

4. Building distance features

The preliminary stage in the process of discerning emotional intentions involves the precise and meaningful extraction of distinguishing features from facial landmarks, such as the forehead, eyebrows, nostrils and cheekbones. These landmarks are essential in many face analysis activities to express certain facial behaviours based on facial muscle movements. Figure 3 gives an overview of our proposed cognitive system, which is composed of pre-processing, keypoint extraction, parameter selection and recognition of emotions via a DNN model.

In the pre-processing stage, by taking the main facial components, facial regions are cropped and the geometric displacement values are taken to generate a feature matrix from the database images. We use a unique PSO-based feature selection method that uses a fitness function to turn the data into a non-linear model that makes it easier for the DNN model to recognize the micro-coded emotions. This makes the extracted features more relevant.

4.1. Facial landmark mapping

Primarily, we detect the face and the facial region from the input video sequence. To facilitate better recognition, we utilized OpenCV's Haar cascades to trace the subject's face in the given image [23]. The Viola-Jones (VJ) technique is employed to improve the trade-off between speed and accuracy in object detection. This algorithm utilizes a cascade function to detect objects more effectively. Secondly, the Dlib frontal face detector toolkit [24] helps to obtain temporal information on the face region by identifying the 68 landmark points for feature mapping.

Figure 4 (a) sample image presenting surprise emotion from the MUG dataset; (b) depicts the spotted facial ROI in the rectangular box using a Haar cascade; (c) annotations of 68 landmark points; and (d) selected landmark face points in green colour to analyze the temporal information of the face.

4.2. Geometrical key point extraction

This article aims to offer a completely intuitive approach to identifying facial expressions that rely entirely on the geometrical features of the face. It mainly extracts the minute fluctuations in facial shape and movements by tracking the prominent points in the facial region, such as the head movement, blink, eyebrow corners, nostril dilator, tongue bulge, lip tightness and jaw drop. The facial traits, distances and angle measurements that have a negligible effect on the emotions classifier's decision-making disappear, enabling just the most serious ones to be tracked and preserved in the database.

The obtained features from the forehead, lips and nose region play an essential part in supplying enough information to recognize any fundamental and microexpression changes from the MUG and GEMEP datasets. From the above-pre-processed image concerning 68 landmark points, we provide a geometrical model depicted in Figure 5, which consists of manually annotated 66-dimension face points with 47 unique Euclidean distances and 19 angle points for mapping the landmark points within a frame. Moreover, it provides the displacement information of landmark points for each input feature vector in the x and y directions.

Using Pythagoras's theorem the distance and angle measure is identified with the help of facial characteristic points. The (x_i, y_i) cartesian coordinates indicate the distance between key landmark points in the facial regions. These indices of the points (p) were computed with the help of 66-dimensional feature points that exhibited the least pronounced variations in facial muscle movements. The feature vector for FER can be obtained by calculating the pairwise coordinates of the two landmark key points. Table 2. presents the 47 unique distance and 19 angle features with an identifier assigned to each.

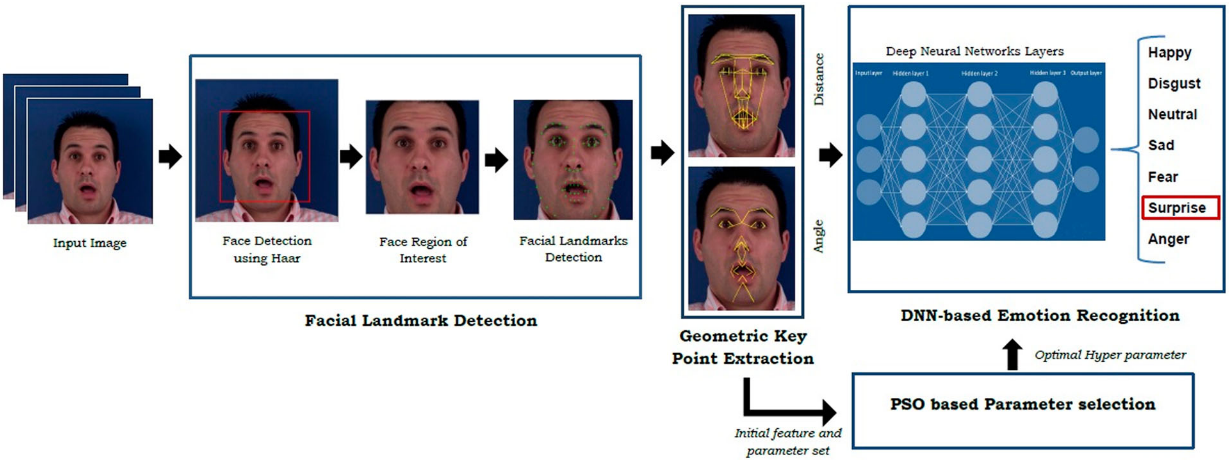

Input Image
Face Detection using Haar
Face Region of Interest
Facial Landmarks Detection
Facial Landmark Detection
Distance
Angle
Geometric Key Point Extraction
Deep Neural Networks Layers
Input layer
Hidden layer 1
Hidden layer 2
Hidden layer 3
Output layer
Happy
Disgust
Neutral
Sad
Fear
Surprise
Anger
DNN-based Emotion Recognition
Optimal Hyper parameter
PSO based Parameter selection
Initial feature and parameter set

Figure 3. Proposed approach for emotion recognition.

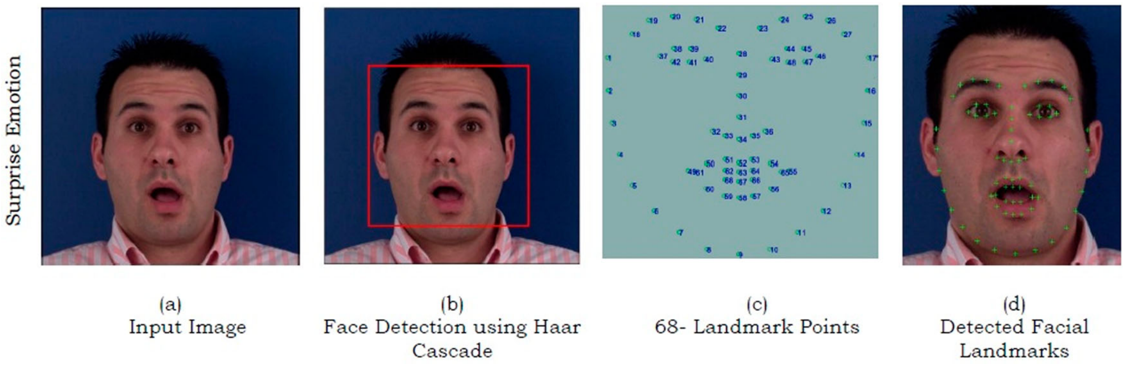


Figure 4. Face Localization and Landmark Detection.

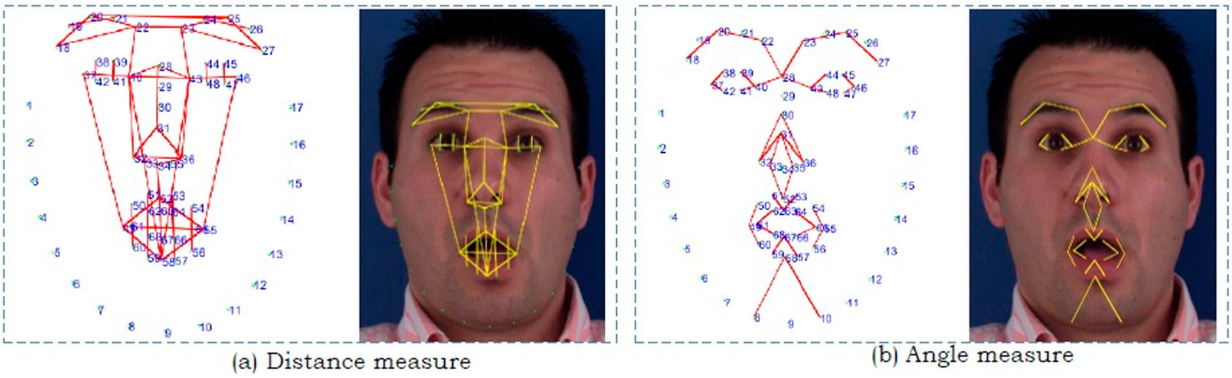


Figure 5. Manually annotated 66 dimensions Geometrical Face Points.

Step 1: The Euclidean distance$D(p_1, p_2)$ is computed for each pair of cartesian coordinate facial landmark points$p_1(x_1, y_1)$ and$p_2(x_2, y_2)$ as shown in Equation (1)

$$D_{(p_1,p_2)} = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2} \qquad (1)$$

Step 2: The cosine angle obtained by two non-zero vectors$A(p_1, p_2, p_3)$, which is from three landmark feature points$p_1(x_1, y_1)$, $p_2(x_2, y_2)$ and $p_3(x_3, y_3)$ is shown in Equation (2)

$$A_{(p_1,p_2,p_3)} = \cos^{-1}\left(\frac{(p_3 - p_2).(p_1 - p_2)}{p_3 - p_2 \quad p_1 - p_2}\right) \qquad (2)$$

We, first analyze, the geometric distance between each pair of landmark coordinates starting from $D1_{(18,22)}$ up to the final points pair $D47_{(43,58)}$ within the frame is calculated, followed by an estimation of the angle $A1_{(18,20,22)}$ between the three pairs of points. The feature is then processed via a unique PSO algorithm for parameter selection before being trained with the DNN for further identifying the micro-coded emotions.

4.3. Novel particle swarm optimization algorithm: an overview

Kennedy and Eberhart first designed a robust stochastic algorithm for selecting multiple features called the

Table 2. Facial Landmark coordinates for feature extraction.

Facial Geometric Features					
Distance				Angle	
#ID	Points pair	#ID	Points pair	#ID	θ
D1	18, 22	D25	34, 52	A1	18, 20, 22
D2	19, 21	D26	49, 55	A2	23, 25, 27
D3	18, 20	D27	52, 58	A3	38, 37, 42
D4	20, 22	D28	49, 52	A4	39, 40, 41
D5	23, 27	D29	52, 55	A5	44, 43, 48
D6	24, 26	D30	49, 58	A6	45, 46, 47
D7	23, 25	D31	58, 55	A7	32, 31, 36
D8	25, 27	D32	37, 49	A8	33, 31, 35
D9	20, 25	D33	46, 55	A9	40, 28, 43
D10	22, 23	D34	40, 28	A10	32, 30, 36
D11	22, 40	D35	43, 28	A11	51, 63, 53
D12	23, 43	D36	31, 28	A12	59, 67, 57
D13	37, 40	D37	32, 34	A13	50, 49, 60
D14	38, 42	D38	36, 34	A14	62, 61, 68
D15	39, 41	D39	62, 68	A15	54, 55, 56
D16	43, 46	D40	63, 67	A16	64, 65, 66
D17	44, 48	D41	64, 66	A17	33, 52, 35
D18	45, 47	D42	50, 60	A18	8, 58, 10
D19	40, 43	D43	54, 56	A19	22, 28, 23
D20	40, 32	D44	53, 57		
D21	43, 36	D45	51, 59		
D22	32, 36	D46	40, 58		
D23	31, 32	D47	43, 58		
D24	31, 36				

Particle Swarm optimization algorithm [24]. The technique is chosen based on the inspiration of movement and intelligence of natural swarms. The social conduct of animal groupings, bird flocks and fish schools is a collective effort to find nourishment. In the original PSO [25], throughout the iterations, each particle in the swarm maintains track of its optimum personal solution (pbest) and the optimum solution in the global swarm (gbest) so that each particle can dynamically modify its velocity and position according to its flying acquaintance.

4.3.1. Determination of velocity and position

The population with a collection of particles where the position of each particle is given as $\mathrm{P_i}$, $[i = 1, \ldots, M]$ in a multi-dimensional space with a position $\mathrm{X_i}$. The velocity function $\mathrm{k_i}$, which considers a particle's individual best location $\mathrm{p_i}$ (i.e. cognitive component) and the best part of that particle within a population $\mathrm{p_g}$ (i.e. social component), causes all the particles to shift their roles in their appropriate space [26]. All the particles swarm to a new spot throughout each iteration until they find the optimum position, and the particle velocity changes each time. For every population, the position vector of each particle updates frequently using the below Equation (3)

$$X_i(t+1) = X_i(t) + V_i(t+1), \tag{3}$$

Here the particle speed estimation concerning its velocity in Equation (4) is as follows.

$$\begin{aligned} V_i(t+1) = w.V_i(t) + c_1 r_1.(pbest - X_i) \\ + c_2 r_2.(gbest - X_i) \end{aligned} \tag{4}$$

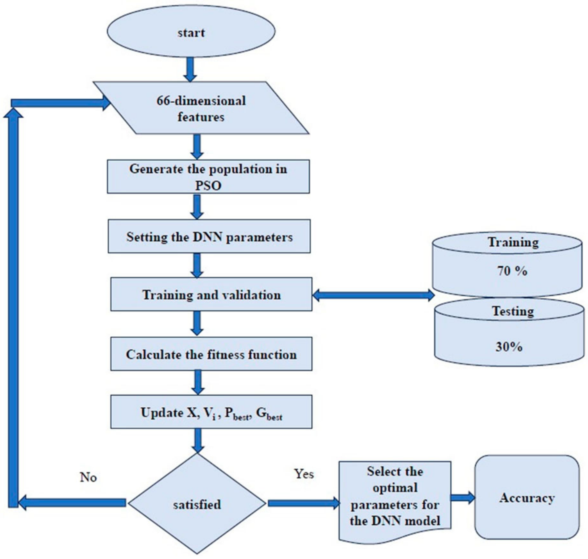

start
66-dimensional features
Generate the population in PSO
Setting the DNN parameters
Training and validation
Calculate the fitness function
Update X, V_i, P_{best}, G_{best}
satisfied
No
Yes
Select the optimal parameters for the DNN model
Accuracy
Training
70 %
Testing
30%

Figure 6. Flowchart for optimizing the DNN Hyperparameters using PSO.

Algorithm: Novel Particle Swarm Optimization algorithm for optimal parameter selection

```
Input: MUG and GEMEP Dataset, 66-dimensional feature vector, label and
   Initial parameter range
Output: Hyper-parameters: Hl, $Hl_n$, $O_n$, Af, $N_e$, Lr, Bs, $D_o$, $O_f$, $O_c$#Define the
   accuracy of the DNN as a fitness function $Acc = \left[\frac{tp+tn}{tp+fp+tn+fn}\right]$
#Define initial parameter values as random positions within the bound
   defined in the input range
Initialize the parameters, search space (X), and velocity ($V_i$).
Initialize the best positions $P_{best}$ for each particle.
while continuing until convergence is achieved up to the max iterations
   reached, then do
   for each particle do
      update the velocity resting in the population
      revise the particle position
         determine the fitness (Acc) of each individual using DNN
      if the current fitness is better than $P_{best}$ then
            modify the $P_{best}$ and fitness
       end if
               for each surrounding, do
                  modify the $g_{best}$ and fitness
               end for
   end for
end while
#Use the $g_{best}$positionas Optimal Hyperparameters.
```

Where $\mathrm{V_i}$ represents the particle velocity at iteration t, w is the inertia of weight that controls particle momentum. $\mathrm{C_1}$ and $\mathrm{C_2}$ are the social and cognitive best position coefficients, that accelerate towards the swarm optimization with added random numbers $\mathrm{r_1}$ and $\mathrm{r_2}$ [0,1]. The particles move through the parameter space representing a possible candidate solution, and the particle's movement is given by a fitness function in terms of $\mathrm{p_{best}}$ and $\mathrm{g_{best}}$ by providing the expected fitness in future trials. Figure 6 gives the detailed flow of optimizing the deep neural network model hyperparameters using the particle swarm intelligence algorithm.

4.4. Statistical feature analysis

The box-and-whisker plot is a statistical method, that visualizes the dataset's distribution graphically.

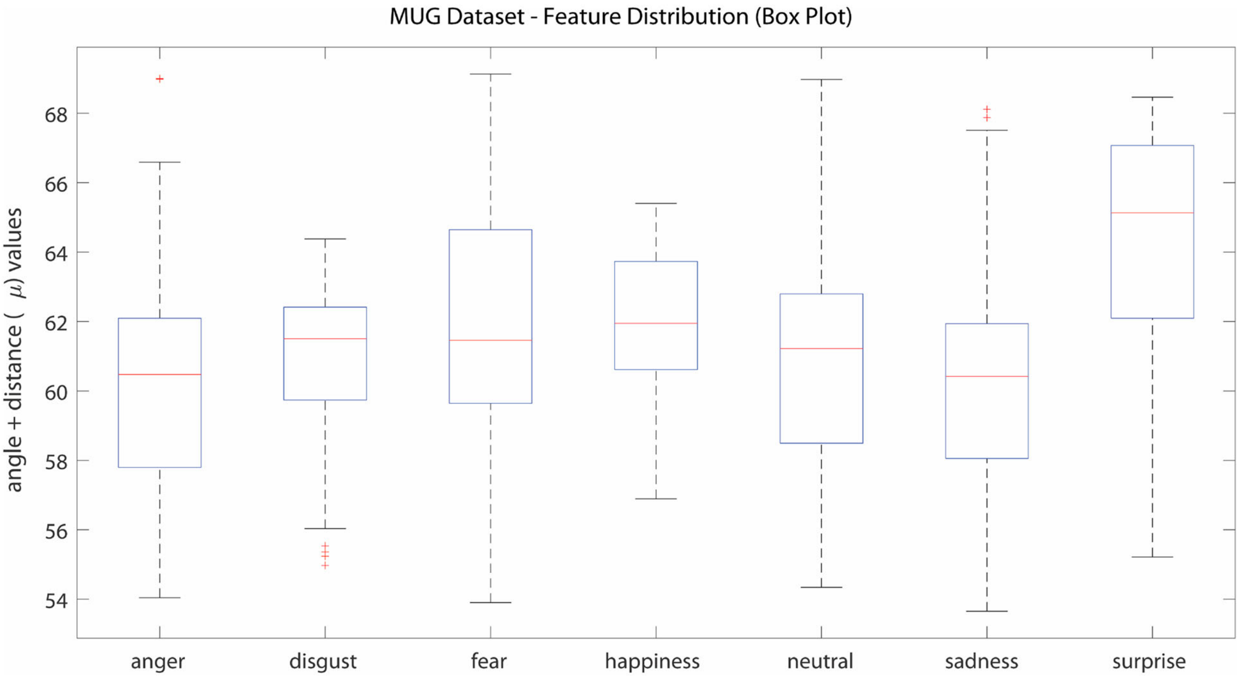


Figure 7. Feature Distribution of emotions in the MUG dataset.

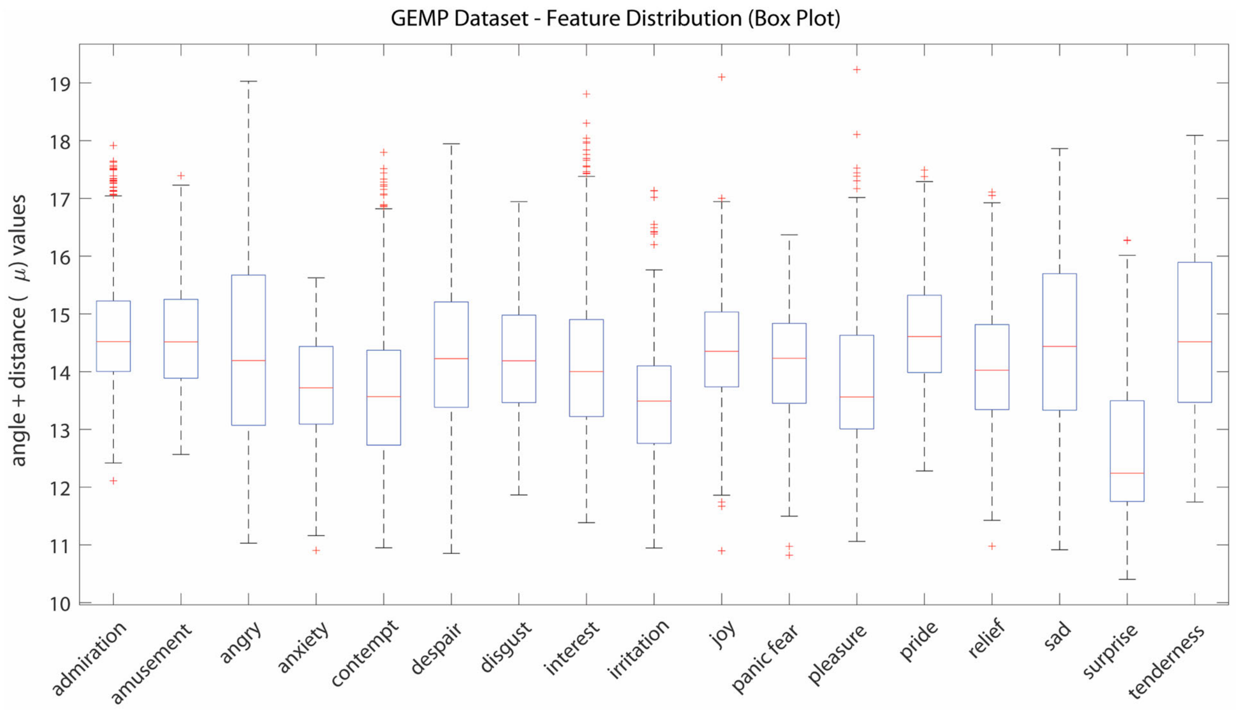


Figure 8. Feature Distribution of emotions in the GEMEP dataset.

The interface represents the individual class instance with essential statistical metrics like median, quartiles and potential outliers. Figures 7 and 8 illustrate the angle + distance feature distribution of data.

The individual box in the box plot indicates the instance of the emotional class, and the central line inside the rectangular box denotes the median of the sample data. The whiskers, which are placed at the upper and lower extrema of the box, represent data points that are devoid of extreme outliers. The plus signs, which are positioned outside the box, represent attributes that define outliers. As a result, these plots illustrate the effective differentiation of emotions in the MUG and GEMEP datasets by the variant features.

4.5. Deep neural networks (DNN)

The present work predicts the latent facial emotions in the emotional datasets through the implementation of a deep neural network model [27]. Often, a DNN primarily acts as an Artificial neural network (ANN) widely used recently in many generic tasks in emotion

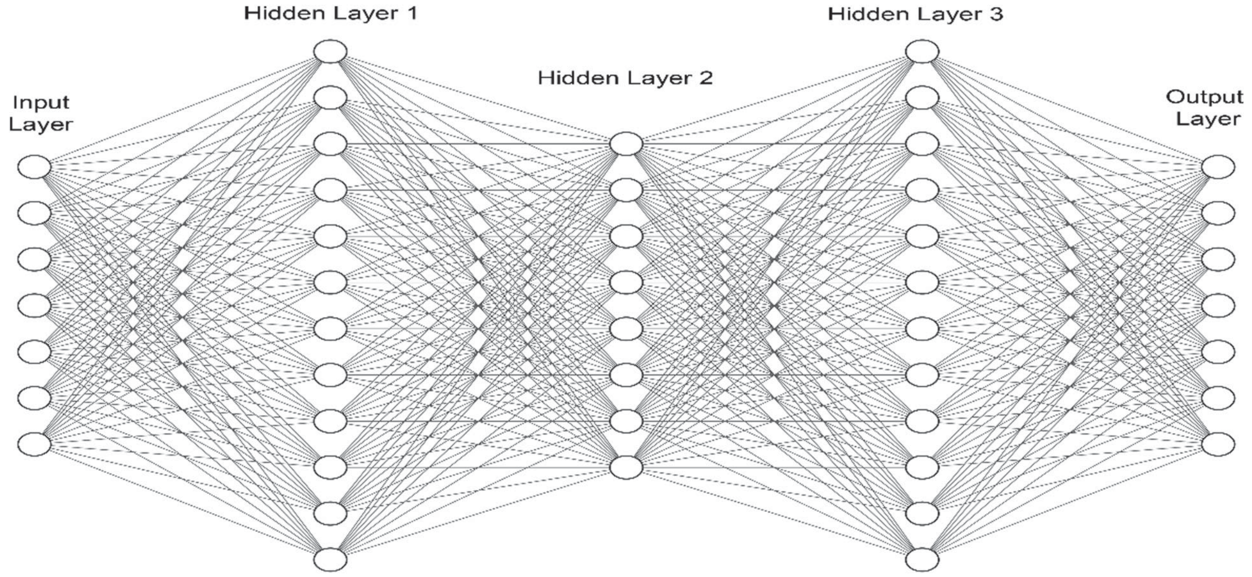


Figure 9. Deep Neural Networks Architecture.

Table 3. Hyperparameter setting for training the DNN using the PSO Algorithm.

Hyper-parameters	Range of values searched for Optimization	The optimal value obtained for MUG Dataset	The optimal value obtained from the GEMEP Dataset
Number of hidden layers (Hl)	[3–10]	3	3
Number of units per hidden layer (Hl_n)	[128–512]	128	256
Optimizer (O_n)	[Adam, Adamax, SGD]	Adamax	Adamax
Activation functions (Af)	[ReLu, Leaky ReLu, Sigmoid, Tanh]	ReLu	ReLu
Epochs number (N_e)	[50–200]	100	100
Learning Rate (Lr)	[0.0001–0.01]	0.001	0.001
Batch Size (Bs)	[4–128]	16	16
Drop out (D_o)	[10–50%] in hidden layers	30%	10%
Output function (O_f)	[Softmax, Sigmoid]	Sigmoid	Sigmoid

recognition and problem-solving. The DNN functions as a feed-forward network with an input and hidden layer involving multiple hidden neurones. Each neurone in the network has its weight and bias and receives connections from the preceding layer to synthesize the perceptron's extracted features. The activation functions max out, relu, leaky relu, sigmoid and tanh assist the network in discovering the intricate connection between layers.

Figure 9 represents a five-layered DNN model consisting of a single input layer and three hidden layers with drop out, followed by the output layer. Each layer comprises several perceptrons, and the connections between neurones in adjacent layers have a feature matrix. The input layer extracts relevant information from an input feature vector via neurones with automatically learnable parameters (weights).

5. Experimental setup

This section discusses the environmental parameters selected by the PSO algorithm for better classification of emotions used in our proposed work. The experimental setup utilizes 32 GB of RAM in Windows 10, a 2.10 GHz Intel i7 processor. The software prerequisites for the DNN prototype consist of Python (3.10.10), the TensorFlow (2.3.1) libraries and the Keras (2.4.3) framework. A variety of hyper-parameter ranges were utilized in the training and testing outcomes, as detailed in Table 3.

Building deep neural network models can be quite challenging, especially when it comes to selecting the right combinations of hyperparameters. The selection of hyperparameters plays a vital role by implicitly enhancing the trade-off performance and model accuracy. Swarm-based deep learning facilitates the efficient selection of the most complex model prediction parameters and prevents overfitting. Therefore, we choose the architecture with fewer layers, resulting in faster inference and better suitability for real-time applications.

5.1. Performance measures

This section assesses the DNN model performance on the validation set in the training phase and enables us to grasp the model's convergence speed better. The proposed extracted geometrical features use 70% for training and 30% for testing. Statistical metrics, accuracy $Acc = \left[\frac{(TP+TN)}{(TP+FP+TN+FN)}\right]$, assesses the proportion of accurately identified samples with the overall sample

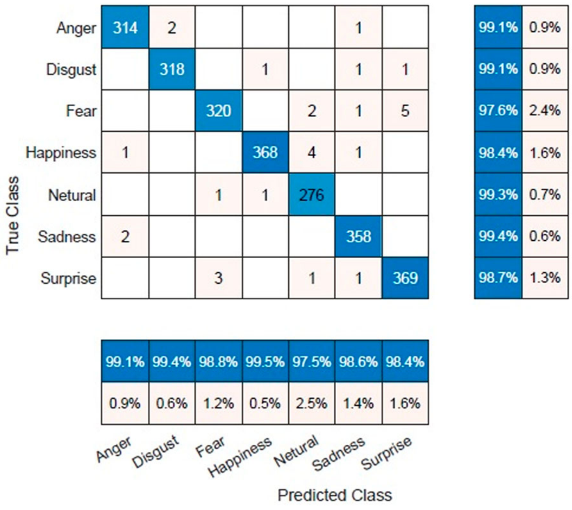


Figure 10. Confusion Matrix for the MUG Dataset with 66 features.

count. From the total number of positive samples, precision (P) estimates the expected positive samples. $P = \left[\frac{TP}{(TP+FP)}\right]$, Recall $R = \left[\frac{TP}{(TP+FN)}\right]$ measures the rates at which emotions are accurately recognized and $F1 - score = 2\left[\frac{(Precision*Recall)}{(Precision+Recall)}\right]$ seeks suitable performance metrics by utilizing the mean of recall and precision. In this context, tp and fp symbolize true and false positives, while tn and fn denote true and false negatives.

5.2. Experimental results on the MUG dataset

The final obtained features from the MUG dataset were trained for 100 epochs with the optimizer Adamax specifying the Lr 0.001. To boost the model's accuracy, we used the non-linear activation function ReLu in the hidden layer with the batch size 16 and to finally boost the output performance "sigmoid" is used.

The confusion matrix derived from the 66 features of the proposed angle and distance measure on the MUG dataset reveals that the emotions "sadness", "neutral", "anger" and "disgust" have been accurately classified with the highest level of precision. However, the emotion "fear" has a lower accuracy due to the misclassification of certain emotions such as happiness, which are incorrectly labelled as surprise, as depicted in Figure 10.

Table 4 presents the performance metrics for the DNN model in the fundamental emotions, with all four statistical metrics. The performance accuracy produced outstanding results with remarkably high consistency, and the system recognized the emotion with 98.76% accuracy on the MUG dataset.

5.3. Experimental results on the GEMEP dataset

The selected optimal features of distance and angle from the GEMEP frames are trained and tested in a deep neural networks model for 100 epochs. We use the ReLu as the non-linear function for each hidden neurone in the layer and a "sigmoid" as the activation function to optimize the output performance. Training models from inception in small datasets such as GEMEP is facilitated by dropout and regularization, resulting in the best validation set performance.

Table 4. Performance measure of seven basic emotions.

Emotions	Accuracy	Precision	Recall	F1-score
angry	99.05	99.05	99.05	99.74
disgust	99.38	99.07	99.22	99.79
fear	98.77	97.56	98.16	99.49
happy	99.46	98.4	98.92	99.66
sadness	97.53	99.28	98.40	99.62
neutral	98.62	99.44	99.03	99.70
surprise	98.40	98.66	98.53	99.53

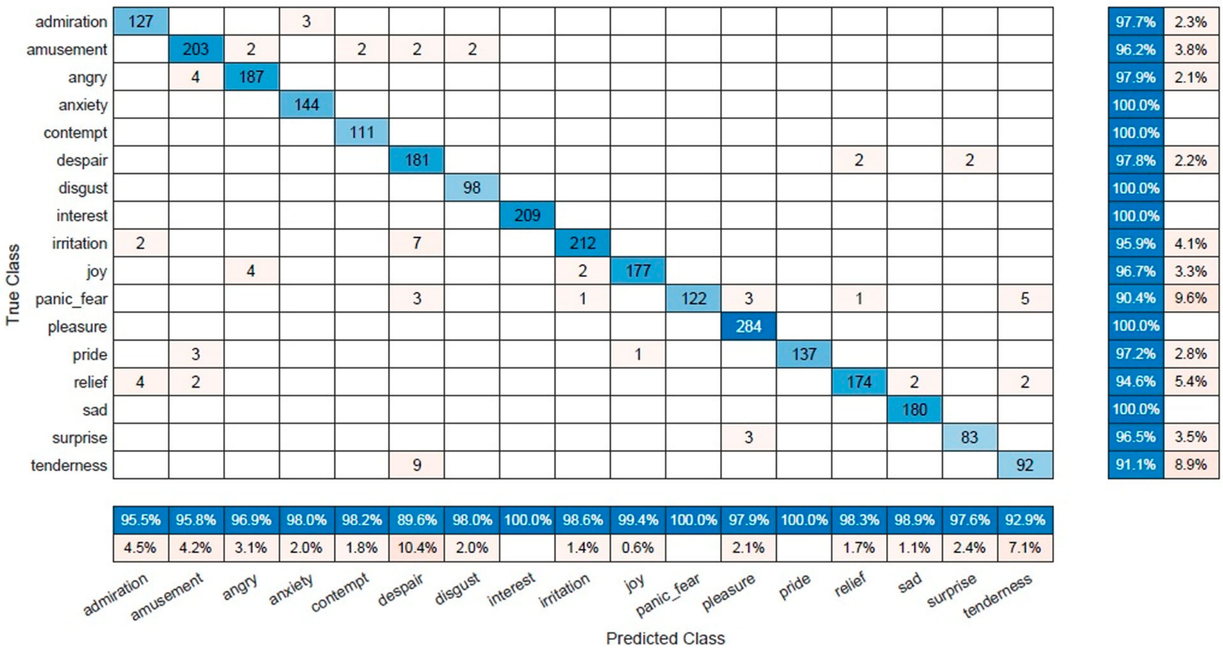
Table 5. Performance measure of micro-coded emotions.

Emotions	Accuracy	Precision	Recall	F1-score
admiration	95.49	97.69	96.58	99.68
amusement	95.75	96.21	95.98	99.39
angry	96.89	97.91	97.40	99.64
anxiety	97.96	100	98.97	99.89
contempt	98.23	100	99.11	99.93
despair	89.6	97.84	93.54	99.11
disgust	98	100	98.99	99.93
interest	100	100	100	100
irritation	98.6	95.93	97.25	99.57
joy	99.44	96.72	98.06	99.75
panic fear	100	90.37	94.94	99.53
pleasure	97.93	100	98.95	99.79
pride	100	97.16	98.56	99.86
relief	98.31	94.57	96.4	99.53
sad	98.9	100	99.45	99.93
surprise	97.65	96.51	97.08	99.82
tenderness	92.93	91.09	92	99.43

Figure 11, illustrates the confusion matrix of the presented features on the dataset and indicates that the categories "anxiety," "contempt," "disgust," "interest," "sad," and "pleasure" have been correctly identified with a high level of accuracy. However, the categories "relief," and "tenderness" have the lowest level of precision.

Table 5 performs the various micro-expression emotional sequences validated with the confusion matrix. In Table 5, the recommended model demonstrates reduced computational time while achieving enhanced performance metrics for the given input size.

Overall, the validation accuracy is 97.79% in combining distance and angle features for the individual emotion instance in the GEMEP dataset. The results reveal that the MUG Dataset is more accurate than the GEMEP Dataset. The MUG Dataset investigates only 7 facial emotions, whereas the GEMEP Dataset deals with 17 facial micro-expressions. Additionally, the MUG dataset is used only for facial emotion action sequences, whereas GEMEP contains the frontal pose of full-body gesture emotions added to it. The sample of 66 features indicates that deep neural networks are excellent at predicting accurate results and recognition of facial micro emotions.



	admiration	amusement	angry	anxiety	contempt	despair	disgust	interest	irritation	joy	panic_fear	pleasure	pride	relief	sad	surprise	tenderness		
admiration	127			3														97.7%	2.3%
amusement		203	2		2	2	2											96.2%	3.8%
angry		4	187															97.9%	2.1%
anxiety				144														100.0%	
contempt					111													100.0%	
despair						181								2		2		97.8%	2.2%
disgust							98											100.0%	
interest								209										100.0%	
irritation	2					7			212									95.9%	4.1%
joy			4						2	177								96.7%	3.3%
panic_fear						3			1		122	3		1			5	90.4%	9.6%
pleasure												284						100.0%	
pride		3								1			137					97.2%	2.8%
relief	4	2												174	2		2	94.6%	5.4%
sad															180			100.0%	
surprise												3				83		96.5%	3.5%
tenderness						9											92	91.1%	8.9%
	95.5%	95.8%	96.9%	98.0%	98.2%	89.6%	98.0%	100.0%	98.6%	99.4%	100.0%	97.9%	100.0%	98.3%	98.9%	97.6%	92.9%		
	4.5%	4.2%	3.1%	2.0%	1.8%	10.4%	2.0%		1.4%	0.6%		2.1%		1.7%	1.1%	2.4%	7.1%		

Figure 11. GEMEP Dataset Confusion matrix outcomes.

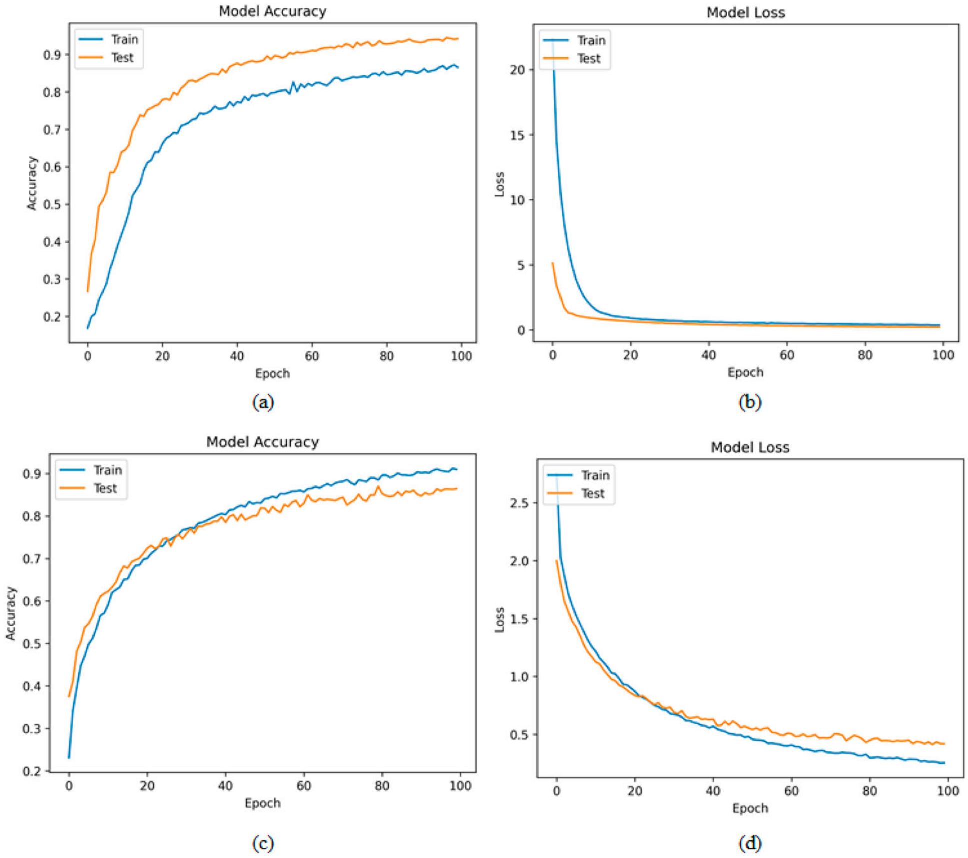


Figure 12. (a) and (b) Model Performance for the MUG Dataset, (c) and (d) Model Performance for the GEMEP Dataset.

5.4. Discussions

The major concern with facial emotion recognition is identifying the micro emotions in facial expressions like pleasant, pride, tenderness, anxiety and calm, Panic_fear. Our research contributes by recognizing the minute subtle changes that happened in a face point less than a second by building a DNN architecture from scratch. Here are the key findings from existing experiments:

- The proposed PSO-DNN model proved superior performance compared to other models in terms of accuracy. This indicates that particle swarm optimization (PSO) effectively optimized new hyper-parameters, resulting in improved performance.

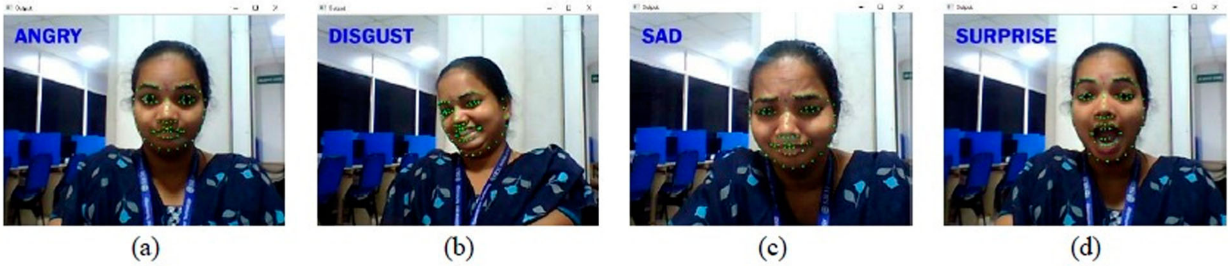


Figure 13. Emotion Recognition from real-time Dataset.

Specifically, the PSO-DNN model achieved 98.76% accuracy for the MUG dataset and 97.7% accuracy for the GEMEP dataset.

- The performance accuracy and loss graph generated exceptional outcomes with a high degree of consistency. As seen in Figure 12 (a,b,c,d) The system accurately classifies emotions in the MUG dataset, and starting from the 60th epoch, the testing and training data converged and remained entirely stable.
- As observed, Figure 13 describes the GEMEP dataset model performance with the minimized training time and error rate. Overall, the validation accuracy is upward with 86.45%; however, significant fluctuations occur at certain epochs. Examining the results reveals that the MUG Dataset is more accurate than the GEMEP Dataset. The reason is that the MUG Dataset investigates only 7 facial emotions whereas the GEMEP Dataset deals with 17 different facial micro expressions. Additionally, the MUG dataset is used only for facial emotion action sequences whereas GEMEP contains the frontal pose of full-body gesture emotions added to it. The sample of 66 features indicates that deep neural networks are excellent at predicting accurate results in exploiting emotion recognition.

Figure 13 The sample images from our real-time Dataset were downloaded and tested in our indoor lab session. (a) Angry, (b) Disgust, (c) Sad, (D) surprise. We have deployed seven basic emotions consisting of six subjects of both male and female characteristics. Random testing went on to assess the efficacy of the facial emotion model. The model is validated with our proposed approach and gives an accuracy of 96.4% in real time.

Table 6. demonstrates the efficacy of the proposed geometrical feature extraction strategy by comparing the anticipated angle and distance measurements to the state-of-the-art outcomes.

Table 6. Cutting-edge results achieved in the emotional datasets.

Reference	Method	Dataset	Accuracy in (%)
Proposed Work	Geometric Features + DNN	GEMEP	90.3
[6]	Dense optical flow + CNN		96.6
Proposed work	**Geometric features using PSO + DNN**		**97.7**
[16]	HiNet using visual saliency	MUG	87.8
[28]	Dense SIFT + SVM		91.8
[12]	Distance Manifolds using SVM		92.7
Proposed work	Geometric Features + DNN		93.1
	Geometric Features + PSO + DNN		**98.7**

6. Conclusion

This study uses deep neural networks to propose a novel framework for facial emotion identification. Concentrating on certain facial features in specific regions of the face aids in accurately identifying subtle facial microexpressions. In this case, we utilized Euclidean distance and angle metrics to extract facial feature information, specifically focusing on 66 unique facial action units. Constructing a Deep Neural Network facilitates the process of training the network from its initial state and making explicit predictions for producing input attributes. The model's efficiency was assessed by individual experiments conducted in both datasets. The findings indicate that the MUG dataset achieves a recognition accuracy of 98.76%, while the GEMEP dataset achieves 97.79% accuracy using PSO. The proposed model is tested in a real-time environment and has a precision of 96.4%. Ultimately, the techniques outlined in different methodologies suggested for facial emotion identification were contrasted. The findings determined that the algorithm was unable to effectively differentiate between surprise and pride. The suggested work aims to decrease computing time and enhance performance metrics for the specified input size. Our additional investigation uses real-time information to identify the emotional states expressed by merging facial traits with visual body motions.

Acknowledgement

The authors of the article would like to express their heartfelt gratitude to the editors and each anonymous reviewer for their thoughtful comments.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Availability of data and materials

The information supporting the current outcomes of the research can be accessed at https://mug.ee.auth.gr/datasets/ and https://www.unige.ch/cisa/gemep. Data used under license for this work are restricted.

ORCID

S. Vaijayanthi http://orcid.org/0000-0002-1959-7906

J. Arunnehru http://orcid.org/0000-0002-2245-5001

References

[1] Vaijayanthi S, Arunnehru J. Human emotion recognition from body posture with machine learning techniques. Commun Comput Inf Sci. 2022;1613:231–242. doi:10.1007/978-3-031-12638-3_20

[2] Vaijayanthi S, Arunnehru J. Synthesis approach for emotion recognition from cepstral and pitch coefficients using machine learning. Singapore: Springer; 2021; p. 515–528.

[3] Kuusikko S, Haapsamo H, Jansson-Verkasalo E, et al. Emotion recognition in children and adolescents with autism spectrum disorders. J Autism Dev Disord 2009;39(6):938–945. doi:10.1007/s10803-009-0700-0

[4] Hung JC, Lin KC, Lai NX. Recognizing learning emotion based on convolutional neural networks and transfer learning. Appl Soft Comput J. Nov 2019;84:105724. doi:10.1016/j.asoc.2019.105724

[5] Arunnehru J, Geetha MK. Motion intensity code for action recognition in video using PCA and SVM; 2013. doi:10.1007/978-3-319-03844-5_8

[6] Sekar V, Jawaharlalnehru A. Semantic-based visual emotion recognition in videos-a transfer learning approach. Int J Electr Comput Eng. 2022;12(4): 3674–3683. doi:10.11591/ijece.v12i4.pp3674-3683

[7] Clark EA, Kessinger J, Duncan SE., et al. The facial action coding system for characterization of human affective response to consumer product-based stimuli: a systematic review. Front Psychol 2020;11:1–21. doi:10.3389/fpsyg.2020.00920

[8] Arunnehru J, Kumar A, Verma JP. Early prediction of brain tumor classification using convolution neural networks. Commun Comput Inf Sci. 2020;1192:16–25. doi:10.1007/978-981-15-3666-3_2

[9] Gunes H, Piccardi M. Bi-modal emotion recognition from expressive face and body gestures. J Netw Comput Appl Nov. 2007;30(4):1334–1345. doi:10.1016/j.jnca.2006.09.007

[10] Fan X, Tjahjadi T. A spatial-temporal framework based on histogram of gradients and optical flow for facial expression recognition in video sequences. Pattern Recognit 2015;48(11):3407–3416. doi:10.1016/j.patcog.2015.04.025

[11] Ghimire D, Lee J. Geometric feature-based facial expression recognition in image sequences using multiclass AdaBoost and support vector machines. Sensors (Switzerland). 2013;13(6):7714–7734. doi:10.3390/s130607714

[12] Aifanti N, Delopoulos A. Linear subspaces for facial expression recognition. Signal Process Image Commun. 2014;29(1):177–188. doi:10.1016/j.image.2013.10.004

[13] Chickerur S, Reddy T, Shabalina O. Parallel scale invariant feature transform based approach for facial expression recognition. Commun. Comput. Inf. Sci. 2015;535:621–636.

[14] Fathallah A, Abdi L, Douik A. Facial expression recognition via deep learning. In: Proc. IEEE/ACS Int. Conf. Comput. Syst. Appl. AICCSA, vol. 2017; Oct 2018. p. 745–750. doi:10.1109/AICCSA.2017.124

[15] Taghi Zadeh MM, Imani M, Majidi B. Fast facial emotion recognition using convolutional neural networks and gabor filters. In: 2019 IEEE 5th Conf. Knowl. Based Eng. Innov. KBEI 2019; 2019. p. 577–581. doi:10.1109/KBEI.2019.8734943

[16] Verma M, Vipparthi SK, Singh G. Hinet: hybrid inherited feature learning network for facial expression recognition. IEEE Lett. Comput. Soc. 2019;2(4):36–39. doi:10.1109/LOCS.2019.2927959

[17] Ravi R, Yadhukrishna SV, Prithviraj R. A face expression recognition using CNN LBP. In: Proc. 4th Int. Conf. Comput. Methodol. Commun. ICCMC 2020; 2020. p. 684–689. doi:10.1109/ICCMC48092.2020.ICCMC-000127

[18] Chowdary MK, Nguyen TN, Hemanth DJ. Deep learning-based facial emotion recognition for human–computer interaction applications. Neural Comput Appl 2023;35(32):23311–23328. doi:10.1007/s00521-021-06012-8

[19] Siam AI, Soliman NF, Algarni AD, et al. Deploying machine learning techniques for human emotion detection. Comput Intell Neurosci. 2022;2022:8032673. doi:10.1155/2022/8032673

[20] Talaat FM, Ali ZH, Mostafa RR, et al. Real-time facial emotion recognition model based on kernel autoencoder and convolutional neural network for autism children. Soft Comput 2024; 1–14. doi:10.1007/s00500-023-09477-y

[21] Aifanti N, Papachristou C, Delopoulos A. The mug facial expression database. In: 11th Int. Work. Image Anal. Multimed. Interact. Serv. WIAMIS 10, January 2016; 2010.

[22] Dael N, Mortillaro M, Scherer KR. Emotion expression in body action and posture. Emotion. 2012;12(5): 1085–1101. doi:10.1037/a0025737

[23] Vaijayanthi S, Arunnehru J. Facial Expression Recognition Using Hyper-Complex Wavelet Scattering and Machine Learning Techniques. In: Proceedings of the 6th International Conference on Advance Computing and Intelligent Engineering: ICACIE 2021. Singapore: Springer Nature Singapore; 2002. p. 411–421.

[24] Mohd Yamin MN, Aziz KA, Siang TG, et al. Particle swarm optimisation for emotion recognition systems: a decade review of the literature. Appl Sci. 2023;13(12), 7054. doi:10.3390/app13127054

[25] Zhang Y, Yan L. Face recognition algorithm based on particle swarm optimization and image feature compensation. SoftwareX. 2023;22:101305. doi:10.1016/j.softx.2023.101305

[26] Donuk K, Ari A, Özdemir MF, et al. Deep feature selection for facial emotion recognition based on BPSO and SVM. Politek Derg. 2023;26(1):131–142. doi:10.2339/politeknik.992720

[27] Krizhevsky A, Sutskever I, Hinton GE. Imagenet classification with deep convolutional neural networks. Commun ACM. 2017;60(6):84–90. doi:10.1145/3065386

[28] Vaijayanthi S, Arunnehru J. Dense SIFT-based facial expression recognition using machine learning techniques. In: Proceedings of the 6th International Conference on Advance Computing and Intelligent Engineering; 2023. p. 301–310.