Received: March 8, 2024. Revised: March 22, 2024. 445

International Journal of
Intelligent Engineering & Systems

http://www.inass.org/

Facial Expression Recognition Based on GSO Enhanced Deep Learning in IOT Environment

Rana H. AL-Abboodi[1]* Ayad A. AL-Ani[1]

[1]*AL-Nahrain University, Iraq*
* Corresponding author's Email: rana.alabboodi.st2022@nahrainuniv.edu.iq

Abstract: Facial expressions play an important role in human communication and integrating deep learning techniques into Internet of Things (IoT) scenarios enhances the understanding of this data, enabling applications in industries such as healthcare, security, and human-computer interaction. The existing methods suffer from lower accuracy and higher computational complexity compared to the proposed Deep Convolutional Neural Network using Galactic Swarm Optimization (DCNN-GSO) approach hinder their practical applicability in real-time image processing tasks. This paper proposes a comprehensive framework for facial expression analysis in IoT environments. The preprocessing phase uses a Gaussian filter to improve image quality and reduce noise. Feature extraction is performed using a spatial temporal interest point (STIP), which captures the spatial and temporal cue of facial expressions. The proposed method leverages Deep Convolutional Neural Networks (DCNN) to extract discriminative features from facial images captured by IoT devices. Galactic Swarm Optimization (GSO) is employed to optimize the hyperparameters of the DCNN model, thereby improving its performance in facial expression classification tasks. By integrating GSO with deep learning, the proposed approach aims to overcome the challenges of limited computational resources and energy constraints inherent in IoT environments. GSO optimizes the parameters of deep learning models for facial expression recognition, improving accuracy and robustness. The proposed framework provides a comprehensive approach for facial expression analysis in IoT environments, solving challenges such as noise, computational complexity, accuracy, etc. in IoT systems, and opening the way for humans and improved device performance and connectivity. The DCNN-GSO method outperforms the competition with remarkable results: 94% accuracy, 92.3% precision, and 91% recall. With a very low mean absolute error (MAE) of 3.46, it shows itself to be a reliable and accurate solution for practical uses.

Keywords: Facial expressions, Internet of things, Gaussian filter, Spatial temporal interest points, Galactic swarm optimization, Deep CNN.

1. Introduction

Human interaction depends on heavily on facial expressions, which may reveal a great deal about feelings, intentions, and responses. For a long time, people have relied heavily on the capacity to read facial emotions in order to efficiently navigate social relationships. An increasing demand has been shown in using deep learning and the Internet of Things (IoT) to computationally recognise and interpret facial movements in a variety of settings as a result of technological breakthroughs [1]. By allowing computers to recognise complex patterns and characteristics straight from unprocessed data, deep learning, a subset of machine learning, has completely changed the area of computer vision. Convolutional Neural Networks (CNNs) are a popular deep learning architecture that have demonstrated impressive results in applications such as facial recognition, object identification, and photo classification [2]. Also, the created refined models that can effectively identify emotions from visual signals by training CNNs on massive datasets of labelled facial expressions. Real-time facial expression analysis in a variety of settings is made possible by the combined use of deep learning with Internet of Things (IoT) technology [3]. Sensor-

equipped IoT devices may effortlessly collect and transfer data from several places. Real-time human emotion perception and response is made possible by the integration of facial recognition algorithms into Internet of Things devices. IoT-enabled face expression recognition systems [4], for example, can help with remote patient monitoring and mental health evaluation in the healthcare industry.

The analysis of minute variations in facial expressions, these systems can offer significant perceptions into patients' mental health and support medical practitioners in providing individualized therapy. Similar to this, by determining how students respond to the lesson plan, IoT devices with face recognition software may improve student engagement and tailor the educational experience [5]. Additionally, IoT-based facial recognition can be used to gauge customer sentiment and preferences, enable more focused marketing campaigns, and increase customer satisfaction levels all around So these technologies in security applications They can also be used For example, IoT devices can improve surveillance systems by detecting suspicious activity and notifying authorities immediately [6]. However, there are severe privacy and proper concerns with the use of facial recognition technology in Internet of Things framework. Collecting and analysing facial information raises concerns about data ownership, permissions, and privacy. In order to successfully implement and use new technologies, strict privacy, transparent data processing and the adoption of various ethics are required to overcome these challenges. The analysis of facial expressions in a variety of settings has exciting new possibilities because to the deep learning and IoT technology integration [7]. However, intelligent systems that can recognize and react to psychological states by utilising algorithms developed for deep learning and IoT infrastructure. This will develop many industries and improve human-machine relations.

- The Gaussian filter is applied to lowering the noise during pre-processing improves image clarity and creates a cleaner input for further analysis.
- Significant spatial and temporal elements are captured by Spatial Temporal Interest Point (STIP) extraction, allowing for a thorough representation of dynamic visual material that is necessary for a variety of tasks including motion analysis and action detection.
- Integration of Galactic Swarm Optimization (GSO) with Deep Convolutional Neural Networks (DCNN) for facial expression recognition in IoT networks.
- Accurate object localization and region-based analysis are made possible by the Histogram of Oriented Gradients (HOG) segmentation method, which uses gradient orientations to split pictures into semantically significant sections.
- Enhanced accuracy and efficiency in object detection and categorization tasks are achieved by the optimization of feature extraction and classification procedures through the integration of Speeded Up Robust Features (SURF) for selection and Galactic Swarm Optimization (GSO) for detection and classification.

The remaining of this work is divided into the following categories: Similar works are included in Section 2, along with a detailed analysis of each. Details on the problem statement are provided in Section 3. In Section 4, the suggested GSO designs are covered in depth. Section 5 presents and analyses the experiment results, along with a detailed comparison of the proposed work with the advanced methods. The paper is finished in section 6, which is the last section.

2. Literature review

Hossain et al. [8] work introduces a facial expression recognition system tailored for diverse applications, with a primary focus on predicting facial expressions from human face regions. The implementation consists of three key components. Initially, a tree-structured part model is deployed to predict landmark points on input images, aiding in facial region detection. These detected regions are subsequently normalized and down-sampled to various sizes to exploit the benefits of multi-resolution images. Following this, different convolutional neural network (CNN) architectures are proposed to analyze texture patterns within the facial regions. Advanced techniques such as data augmentation, progressive image resizing, transfer-learning, and parameter fine-tuning are then applied in the third component to extract more distinctive and discriminative features, thereby enhancing the performance of the CNN models. Evaluation is conducted using benchmark databases including the Karolinska-directed emotional faces (KDEF), GENKI-4k, Cohn-Kanade (CK+), and Static Facial Expressions in the Wild (SFEW), where the proposed system is compared against existing methods, demonstrating superior performance. However, despite the integration of these advanced techniques to bolster performance and tackle challenges in facial expression recognition, the system may still encounter limitations in scenarios featuring extreme facial expressions, occlusions, or variations in

lighting conditions. Further research and refinement are warranted to comprehensively address these challenges.

Recently, there has been a growing interest in analyzing human facial expressions using thermal images captured by Infrared Thermal Imaging (IRTI) cameras, as opposed to standard cameras utilizing visible spectrum light. Infrared cameras are advantageous in low-light conditions and capture thermal distribution, making them valuable for applications such as robot interaction systems, cognitive response quantification, and disease control. This paper introduces IRFacExNet (InfraRed Facial Expression Network), a deep learning model designed for facial expression recognition (FER) from infrared images. IRFacExNet incorporates Residual and Transformation units to extract expressive features specific to facial expressions, aiding accurate emotion detection. The model employs the Snapshot ensemble technique with a Cosine annealing learning rate scheduler to enhance performance. Evaluation on the IRDatabase from RWTH Aachen University, containing expressions like Fear, Anger, Contempt, Disgust, Happy, Neutral, Sad, and Surprise, demonstrates a recognition accuracy of 88.43%, surpassing several state-of-the-art methods. However, for improved efficiency, integrating feature selection algorithms may reduce computational overhead. Additionally, evaluating the model on larger databases would bolster its robustness [9].

Emotion recognition is crucial for affective computing systems to adapt to users' current moods. Electroencephalography (EEG) signal analysis has gained traction in studying human emotions due to its non-invasive nature. This paper introduces a two-stage deep learning model that correlates facial expressions and brain signals to recognize emotional states. Unlike traditional approaches analysing large signal segments, this model leverages facial expressions as markers to enhance recognition accuracy. A facial emotion recognition technique (FER) is employed to identify emotional responses, guiding the extraction of relevant EEG segments for analysis. The proposed model is evaluated using the DEAP dataset. While these techniques improve precision in identifying when emotions occur, advancements in analysis and machine learning methodologies could further enhance their effectiveness [10].

Kopaczka et al. [11] present a comprehensive system employing diverse image processing algorithms for automated thermal face analysis in various settings, including both controlled laboratory conditions and real-world environments. Our system integrates functionalities for face detection, facial landmark detection, face frontalization, and subsequent analysis into an adaptable and modular workflow. It allows for the incorporation of additional algorithms to enhance performance or cater to specific requirements. Our pipeline includes a histogram of oriented gradients support vector machine (HOG-SVM) based face detector, alongside multiple landmark detection methods utilizing feature-based active appearance models, deep alignment networks, and deep shape regression networks. Face frontalization is accomplished through piecewise affine transformations. For emotion recognition, we employ HOG features and a random forest classifier, complemented by a respiratory rate analysis module computing average temperatures from automatically detected regions of interest. Our system demonstrates comparable performance to current stand-alone state-of-the-art methods for thermal face and landmark detection, achieving a classification accuracy of 65.75% for four basic emotions. Future research endeavors will concentrate on applying these algorithms in real-world scenarios to detect emotions and stress, with a view to refining the robustness and precision of our approach.

Each of the mentioned conventional techniques exhibits certain drawbacks that warrant attention. Firstly, in the facial expression recognition system proposed by Hossain et al. [8], despite the integration of advanced techniques such as data augmentation and parameter fine-tuning, limitations persist in scenarios involving extreme facial expressions, occlusions, or variations in lighting conditions. Further refinement and research are necessary to comprehensively address these challenges. Secondly, while thermal imaging using Infrared Thermal Imaging (IRTI) cameras presents advantages over standard cameras, such as operating well in low-light conditions, the IRFacExNet model introduced by an unnamed source [9] may face challenges in terms of computational overhead and robustness. Integrating feature selection algorithms and evaluating the model on larger databases could enhance its efficiency and robustness. Additionally, the two-stage deep learning model proposed for emotion recognition by correlating facial expressions and EEG signals [10] offers improved precision in identifying emotions. However, there is room for advancements in analysis and machine learning methodologies to further enhance effectiveness. Lastly, in the system presented by Kopaczka et al. [11], which utilizes various image processing algorithms for automated thermal face analysis, challenges may arise in real-

world scenarios due to limitations in accurately detecting emotions and stress. Future research efforts will focus on refining the robustness and precision of the approach to address these limitations comprehensively.

3. Problem statement

Based on the identified drawbacks of conventional techniques, the proposed method aims to address these limitations comprehensively by refining the robustness and precision of automated thermal face analysis. By integrating advanced image processing algorithms and leveraging deep learning techniques, the scope of the proposed method extends towards improving accuracy in detecting facial expressions, particularly in challenging scenarios involving extreme facial expressions, occlusions, variations in lighting conditions, and real-world settings. Additionally, the method seeks to enhance computational efficiency and robustness, potentially through the integration of feature selection algorithms and the evaluation on larger databases. Through these efforts, the proposed method aims to advance the effectiveness and applicability of thermal face analysis for various applications, including emotion recognition and stress detection.

4. Proposed HOG-GSO method

In the proposed framework, Gaussian filter is used as the first step in pre-processing in this suggested image processing pipeline in order to smoothen and remove noise in the picture. Next, to identify relevant sites in both geographic and temporal dimensions, feature extraction is carried out using geographic Temporal Interest sites (STIP). The next step in the process is segmentation, which divides the picture into relevant parts based on gradient orientations using a Histogram of Oriented Gradients (HOG). Then, to emphasize unique local characteristics, feature selection is carried out using Speeded Up Robust characteristics (SURF). Galactic Swarm Optimization (GSO), a metaheuristic optimization algorithm influenced by the behavior of galactic swarms, is then used to carry out the detection and classification, allowing for the accurate and efficient identification and classification of objects or patterns within the processed picture data [12].

4.1 Data collection

The data collection involves using a webcam to capture images for facial expression recognition, as demonstrated in the Kaggle dataset Emotion Detection with Webcam [13]. This dataset likely contains images labelled with various facial expressions such as happiness, sadness, anger, etc., along with corresponding metadata for training machine learning models to recognize these emotions from webcam input.

4.2 Image pre-processing - gaussian filter

Reducing noise while preserving as many important picture components as possible, such edges and features, is the goal of image de-noising. For noise reduction in an additive noise scenario, linear filters that convolve the image with a constant matrix to get a linear combination of neighborhood values

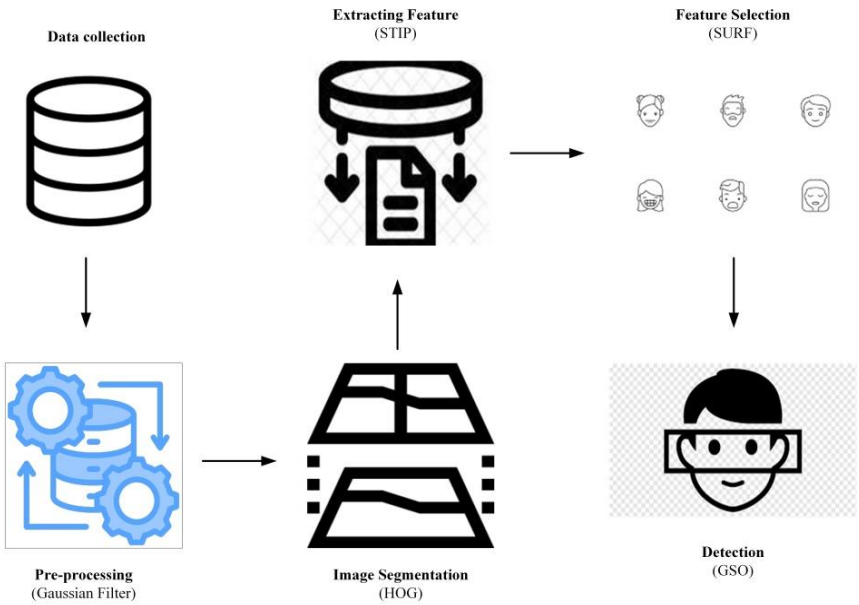

Figure. 1 Workflow of proposed HOG-GSO method

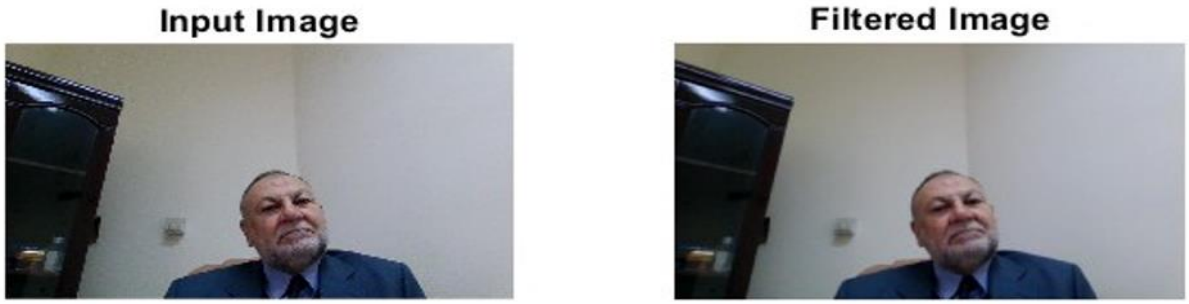

Figure. 2 Pre-processed image

have been widely used. This could still provide a fuzzy, smoothed-out image with inadequate feature localization and noise reduction. Filters based on Gaussian functions are particularly significant since the morphologies of Gaussian functions are easily established and the forward and inverse Fourier transformations of a Gaussian function are real Gaussian functions [14]. Moreover, a wider spatial domain filter produced by a smaller frequency domain filter attenuates low-frequency signals and enhances smoothing/blurring. Gaussian filters are the linear filters that are frequently used for image de-noising. The weight of the pixels in Gaussian filters decreases with separation from the filter center [15], as seen by Eq. (1).

$$G_\sigma(x,y) = \frac{1}{2\pi\sigma^2 e^{-(x^2+y^2)/2\sigma^2}} \tag{1}$$

The average pixel values within a local region, Gaussian filters decrease noise while maintaining picture characteristics. Gaussian filters presume for images to have smooth spatial fluctuations and that pixels within a neighbourhood have near values. This assumption, however, breaks down at the borders where the spatial fluctuations are not smooth and the edges become blurry when the Gaussian filter is applied. The pre-processed image is depicted in Fig. 2.

4.3 Feature extraction-spatial temporal interest point (STIP)

Interest Points in the Spatial Domain, the model image $f^{sp}: R^2 \to R$ by its linear scale-space representation $M^{sp}: R^2 * R + \to R$ in Eq. (2) [16]

$$M^{sp}(a,b;\sigma_m^2) = h^{sp}(a,b;\sigma_m^2) * f^{sp}(a,b) \tag{2}$$

According to the convolution of f^{sp} with Gausian kernel of variance σ_m^2 [16].

$$h^{sp}(a,b;\sigma_m^2) = \frac{1}{2\pi\sigma_m^2} exp\left[-\frac{(a^2+b^2)}{2\pi\sigma_m^2}\right] \tag{3}$$

Finding geographical areas where f^{sp} exhibits noticeable variations in both directions of the interest point detector. To find these spots, use a second moment matrix that has been integrated across a Gaussian frame of variance σ_i^2 for a particular scale of observation σ_m^2 is given as Eqs. (4) and (5) [16].

$$\mu^{sp}(.;\sigma_m^2,\sigma_i^2) = h^{sp}(.;\sigma_i^2) * \left[\left(\nabla M(.;\sigma_m^2)\right)\left(\nabla M(.;\sigma_m^2)\right)^T\right] \tag{4}$$

$$\mu^{sp}(.;\sigma_m^2,\sigma_i^2) = h^{sp}(.;\sigma_i^2) * \begin{bmatrix} \left(M_a^{sp}\right)^2 & M_a^{sp}M_b^{sp} \\ M_a^{sp}M_b^{sp} & \left(M_b^{sp}\right)^2 \end{bmatrix} \tag{5}$$

Where, the convolution operator is represented as $*$, M_a^{sp} and M_b^{sp} Gaussian derivatives computed at local scale σ_m^2 according to $M_a^{sp} = \partial a\,[h^{sp}(.;\sigma_m^2) * f^{sp}(\cdot)]$ and $M_b^{sp} = \partial b\,[h^{sp}(.;\sigma_m^2) * f^{sp}(\cdot)]$. Therefore, the eigenvalues λ_1, λ_2 of μ^{sp} serve as descriptors of fluctuations in f^{sp} along the two directions of the image. More specifically, the presence of an interest point is indicated by two considerably large values of λ_1 and λ_2 in Eqs. (6) and (7) [16].

$$l^{sp} = \det(\mu^{sp}) - K\,trace^2(\mu^{sp}) \tag{6}$$

$$l^{sp} = \lambda_1\lambda_2 - K(\lambda_1+\lambda_2)^2 \tag{7}$$

The ratio of the eigen values, $\alpha = {\lambda_2}/{\lambda_1}$, must be large at the locations of the interest spots. The ratio α must meet $k < {\alpha}/{(1+\alpha)^2}$ for positive local maxima of H^{sp}, as may be seen from (5). Higher values of α correlate to more elongated-shaped interest sites, which may be detected with lower values of k [16].

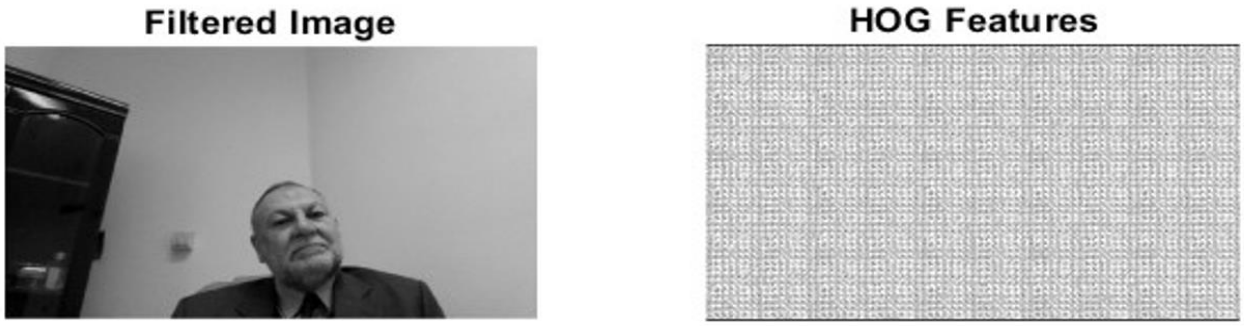

Figure. 3 HOG features

4.4 Image segmentation - histogram of oriented gradients (HOG) process

The density distribution of gradients may be used to characterize the local object look and form inside an image, according to the HOG features descriptor. By splitting the image into discrete areas known as cells, this description may be implemented [17]. For every pixel inside the cell, a gradient direction histogram is constructed. The object is segmented using the HOG technique in four phases. The first step is to compute the gradient values using 1-D centeredness to get the discrete derivative mask point in both the horizontal and vertical orders as shown below Eqs. (8) and (9) [18]:

$$D_x = [-1 \quad 0 \quad 1] \tag{8}$$

$$D_y = \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix} \tag{9}$$

If I is the image, use the convolution technique to get the derivatives of x and y in Eq. (10) [18]:

$$I_x = I_x * D_x \text{ and } I_y = I_y * D_y \tag{10}$$

The formula to determine the gradient's magnitude [18] is given as Eq. (11)

$$|G| = \sqrt{I_x^2 + I_y^2} \tag{11}$$

and for gradient orientation [18] is given by Eq. (12)

$$\theta = arc\tan {I_x}/{I_y} \tag{12}$$

Spatial orientation binning is the next phase. This stage provides a voting method for the cell histogram result. A weighted vote for orientation is cast by each pixel in the facial image, with each pixel corresponding to the nearest bin between 0 and 180^0 [18]. To normalize the cell and histogram from the whole block region into a vector form, the third step uses the HOG descriptor. In the last phase of the block normalization, the L2 norm is used as in the following manner of Eq. (13) and Fig. 3 describes the HOG Features [19].

$$b = \frac{b}{\sqrt{\|b\|^2 + \varepsilon^2}} \tag{13}$$

4.5 Feature selection - speeded up robust features (SURF)

SURF is a technique for selecting local features. The image's characteristic key points are selected using an area consistent fast crucial point detection. It picks the picture and features description utilizing a distinct description. In compared to the SIFT extraction of features strategy, it is a theoretically efficient and effective approach. A SURF method works essentially in the following system: The essential elements of a image are used to select its feature key points [20].

- The important points are then given the orientation. In relation to the intriguing focal point, the positioning is allocated in a rounded motion.
- The squared area is then adjusted in accordance with the chosen orientation.
- Finally, Haar wavelet replies are used to derive feature description. Typically, a description vector is mined from an 8D feature vector.

4.6 Detection and classification - galactic swarm optimization (GSO)

Once feature extraction from the Spatio-Temporal Interest Points (STIP) based activity recognition is completed, the further processing of images ensues. These processed images yield feature

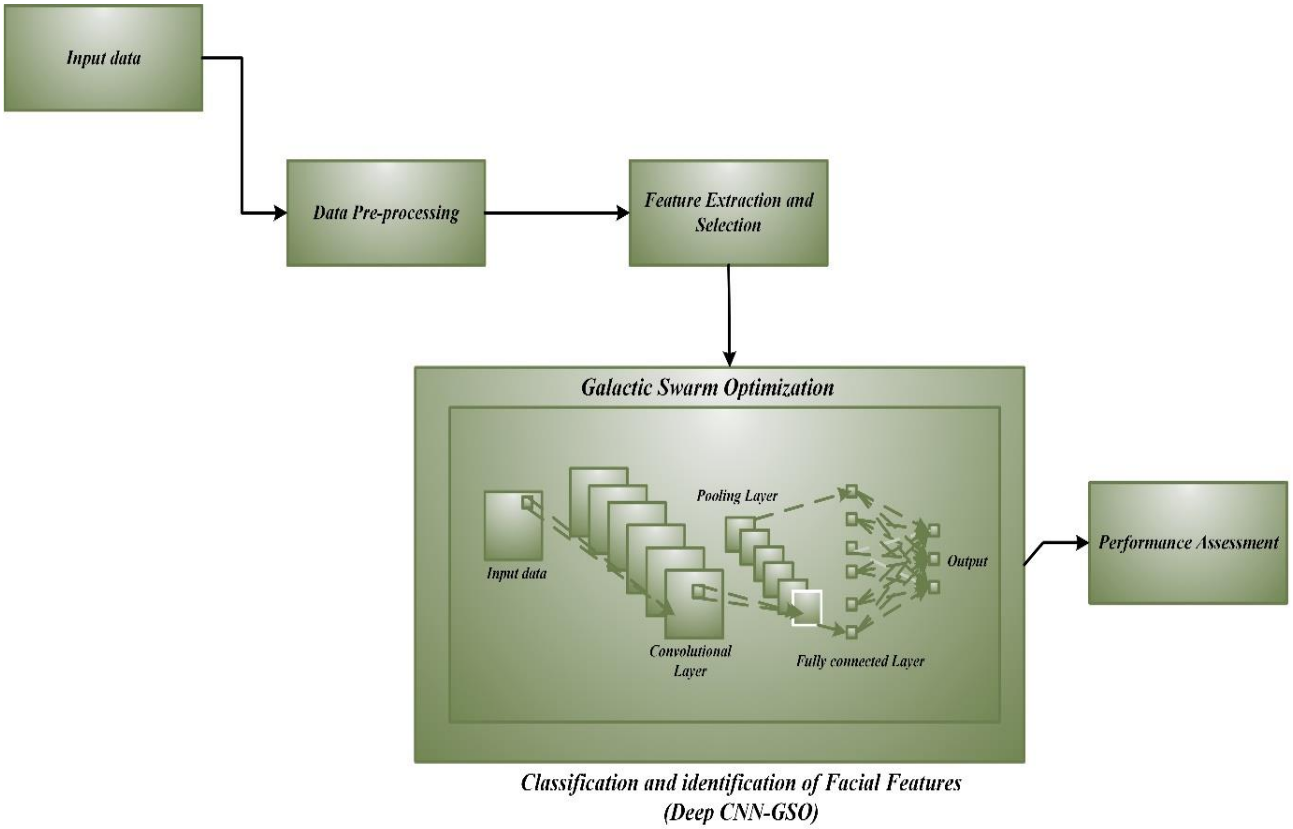

Figure. 4 Classification and identification using DCNN-GSO

vectors, which serve as input to Deep Convolutional Neural Networks (D-CNN) Fig. 4 shows the proposed DCNN-GSO architecture. D-CNN plays a pivotal role in this context by convolving images with kernels to generate feature maps. The weights associated with these kernels facilitate connections between every unit of the feature map and preceding layers [14]. During dataset training, these kernel weights are instrumental in enhancing the input characteristics, allowing the network to learn relevant patterns and features. Notably, the weights that necessitate training within the convolutional layers are fewer compared to fully connected layers. This discrepancy arises from the fact that the kernels are specific to each unit of the feature map, thus reducing the parameter space that requires optimization. Subsequently, the feature vectors extracted from each frame are fed into the CNN for training. This training process entails adjusting the network's parameters, including kernel weights, through iterative optimization techniques like backpropagation, thereby enabling the network to learn and recognize activity patterns effectively. Overall, this proposed methodology leverages the capabilities of D-CNN to extract discriminative features from image data, facilitating robust activity recognition in the spatio-temporal domain [21].

4.6.1 Input image

In the initial step of the process, the input image utilized is the image captured from the webcam for facial expression recognition. This image delineates the face of the individual, encapsulating the area where emotion detection is intended to be conducted. By isolating the relevant portion of the face, the captured image provides a focused and pertinent visual input for subsequent analysis. This focused input is crucial for ensuring that the facial expression recognition system directs its attention solely to the facial features, thereby enhancing the accuracy and efficiency of the recognition process. Additionally, the facial recognition process helps to eliminate extraneous information or background noise, enabling the system to concentrate exclusively on the pertinent facial expressions within the designated region. Overall, the utilization of the captured image as the input facilitates targeted and effective facial expression recognition analysis, laying the groundwork for accurate emotion detection.

4.6.2 CNN

The convolution layer is a pivotal component within convolutional neural networks (CNNs),

responsible for applying a series of kernel functions to input images. These kernels convolve across the images, extracting diverse features crucial for subsequent analysis. Initially, the layer focuses on identifying low-level features like color variations, edges, and gradient orientations, aiding in the foundational understanding of the visual data. With each subsequent layer, the convolutional operations become more intricate, enabling the network to discern increasingly complex features. This hierarchical feature extraction process contributes to a comprehensive understanding of the image dataset, capturing both elementary visual elements and more nuanced patterns. Ultimately, the convolution layer's role is fundamental in facilitating the network's ability to interpret and learn from image data, forming the backbone of CNN architectures and enhancing their capacity for sophisticated pattern recognition tasks [22].

4.6.3 Max-pooling layer

The max-pooling layer serves as a critical component in convolutional neural networks (CNNs), primarily focusing on reducing the spatial dimensions of convoluted images through dimensionality reduction. During this process, max pooling selectively chooses the maximum value from each kernel, while alternative methods such as average pooling calculate the average value. By effectively down sampling the feature maps, the layer helps to mitigate the influence of noise present in low-level features, thereby enhancing the robustness and efficiency of the feature extraction process. Through this selective pooling operation, the layer aids in preserving important information while discarding redundant or less significant details, contributing to improved model performance and generalization capabilities. Ultimately, the max-pooling layer plays a crucial role in optimizing the computational efficiency and interpretability of CNN architectures, facilitating more effective and reliable image analysis tasks [12].

4.6.4 Activation unit

The activation unit serves as a pivotal component within convolutional neural networks (CNNs), wielding significant influence over classification outputs. While various activation functions exist, the Rectified Linear Unit (ReLU) stands as a prominent choice in CNN architectures due to its simplicity and effectiveness. Alongside ReLU, the binary step function has been introduced to streamline computational efforts in the classification process. Operating on a threshold-based principle, the binary step function activates a neuron if its input value surpasses or falls below a predefined threshold, sending an unaltered signal to the subsequent layer. This approach aids in simplifying computations by discretizing the activation process, efficiently distinguishing between activated and deactivated neurons without the need for complex calculations. By employing the binary step function in conjunction with ReLU, CNNs can strike a balance between computational efficiency and classification accuracy, thereby optimizing performance in various image recognition and classification tasks [23].

4.6.5 Fully connected layer

The fully connected 2D layer is a crucial component in neural network architectures, typically following the convolutional layers. This layer operates by transforming the processed image into a column matrix format, allowing for the consideration of all possible combinations of high-level features extracted by preceding layers, particularly the convolutional layer. Through this transformation, the fully connected layer facilitates the integration of spatial information across the image, enabling comprehensive feature representation. Furthermore, the layer incorporates activation units, such as Rectified Linear Units (ReLU) or the binary step function, to introduce non-linearity into the model and enhance its expressive power. By binding the high-level features with these activation functions, the fully connected 2D layer contributes to the development of a robust and discriminative model capable of accurately classifying input images. Overall, this layer plays a pivotal role in consolidating extracted features and refining the model's representation of the input data, ultimately leading to improved performance in various image recognition and classification tasks [24].

After the Deep Convolutional Neural Network (DCNN) process, Galactic Swarm Optimization (GSO) optimization is employed to fine-tune the network parameters and improve its performance. GSO is a nature-inspired optimization algorithm based on the behavior of galaxies, which aims to efficiently search the parameter space and enhance the convergence and accuracy of the DCNN model.

GSO is used for selecting optimum weightiness parameter. A fuzzy method is used in AGSO to vigorously alter the situations. The movement of galaxies and stars caused by gravity pull served as the inspiration for this GSO algorithm. Phases one and two are carried out by GSO: exploration and exploitation. In the exploration phase, the subpopulation particles search the vector space in an

attempt to discover the best approach. Throughout the exploitation phase, the appropriate solution from each subpopulation moves toward the global best. Moreover, on a vast enough scale, these galaxies are regarded as point masses. As soon as a substantial mass of these acquired galaxies is collected to create the super cluster of galaxies. The whole particle population in AGSO is originally split up into M sub-swarms [25]. PSO is carried out for every sub-swarm. Eqs. (14) and (15) provide the formula that is used to update the location and velocity [26]:

$$V_j(k) \leftarrow I_{w1}V(k) + C_1R_1\left[P_{bj}(k) - e_j^k\right] + C_2R_2\left[G_b(k) - e_j^k\right] \quad (14)$$

$$e_j^k \leftarrow e_j^k + V_j(k) \quad (15)$$

Where, $P_{bj}(k)$ is the personal best for the particle e_j^k is denoted as $P_{bj}(k)$, $G_b(k)$ is the best solution of the global sub-swarm, and $V_j(k)$ represents the present velocity. The direction of the best global and local solutions is represented by the constants X^k, C_1 and C_2. Inertia weight is represented by I_{w1}, and R_1 and R_2 are found by using the following formulas in Eqs. (16) and (17) [27].

$$I_{w1} = 1 - {}^{m}/_{s_1} + 1 \quad (16)$$

$$R_i = \cup(-1,1) \quad (17)$$

Where, R_1, R_2 stand for the random numbers, the number of current iterations is given in m, which goes from 0 to s_1, and the inertia weight is represented as I_{w1}. To create super-clusters, the global bests that are reached in the next phases are grouped together. Eqs. (18) and (19) represents the super-swarm Z, which is formed by combining the global bests from X^k subswarms [27]:

$$Z^k \in Z: k = 1,2,3, \dots, N \quad (18)$$

$$Z^k = G_bk \quad (19)$$

Similar to the first level, another level practises the PSO foundation to compute the spot and velocity of particles. However, the second level slightly modifies the equations, as seen below in Eqs. (20), and (21) [27].

$$V(k) \leftarrow I_{w2}V(k) + C_3R_3P_b(k) - Z^k + C_4R_4(G_b - Z^k) \quad (20)$$

$$Z^k \leftarrow Z^k + V(k) \quad (21)$$

P_bk is the level of personal best; C_3 and C_4 are the acceleration constants; G_bk is the level of global best solution. The estimation of I_{w2}, R_3, and R_4 is done using the same equations that were used for level 1. Because the super swarm Concentrations on the finest worldwide from the sub-swarm, which may increase utilization. The super-swarm uses the data that was generated to apply the optimal solution found by the sub-swarms. However, in order to get better results, dynamic adjustments to the C_3 and C_4 parameters are needed throughout the execution process. Here, the settings are dynamically adjusted using fuzzy logic to improve GSO performance [27]. Using fuzzy logic [27], the adaptive parameters C_3 and C_4 are defined as Eq. (22)

$$C_3 = \frac{\sum_{j=1}^{R^{C_3}} \mu_j^{C_3}(C_{3j})}{\sum_{j=1}^{R^{C_3}} \mu_j^{C_3}}; C_4 = \frac{\sum_{j=1}^{R^{C_4}} \mu_j^{C_4}(C_{4j})}{\sum_{j=1}^{R^{C_4}} \mu_j^{C_4}} \quad (22)$$

Where, the output result for rule j is described as C_{3j} and C_{4j}, and IC_{3j}, IC_{4j} denotes the membership function associated with rule j. The full rule of this fuzzy system is indicated as R^{C_3}, R^{C_4}. Detection and classification from photos of nature scenes is widely used in many different disciplines. Thus, in order to mine text from images, are including the deep machine learning technique here.

5. Results and discussion

Facial expression detection involves using algorithms to identify and analyze facial features to determine emotions. It reads facial landmarks and interprets them into specific emotional states as mentioned in Figs. 5 and 6. In Fig. 7, a green rectangle highlights the detected face, accompanied by text indicating the presence of one detected face. On the right, Fig. 8 represents the Happy Face a yellow rectangle outlines the person's face, potentially signifying emotion recognition, with the label stating Happy Face suggesting detection of a happy emotion. In Fig. 9, a green rectangle highlights the detected face, with accompanying text indicating the detection of one face. On the right side, Fig. 10 represents the Angry Face, and a yellow rectangle outlines the person's face, potentially indicating emotion recognition. Fig. 11 shows the Detected Face, and a green rectangle delineates the detected face, accompanied by text indicating the detection of one face. On the right side, Fig. 12 shows the Angry Face, and a yellow rectangle outlines the individual's face, potentially indicating emotion recognition.

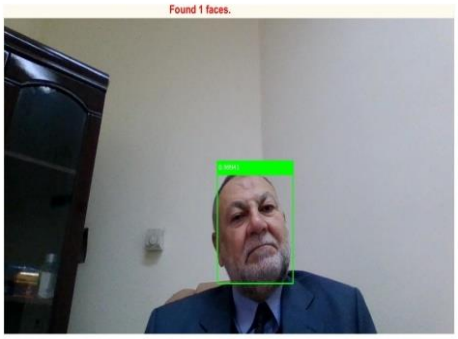

Figure. 5 Face detected

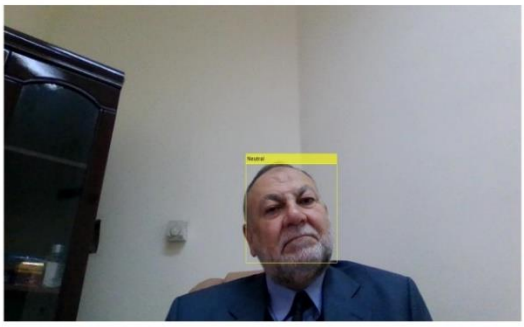

Figure. 6 Neutral face

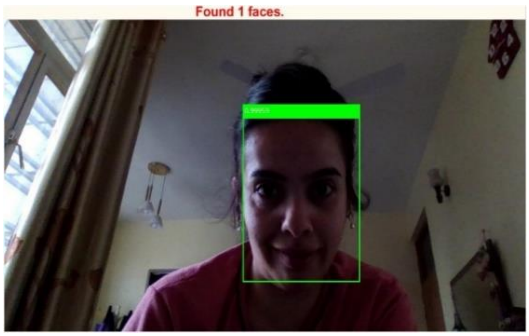

Figure. 7 Face detected

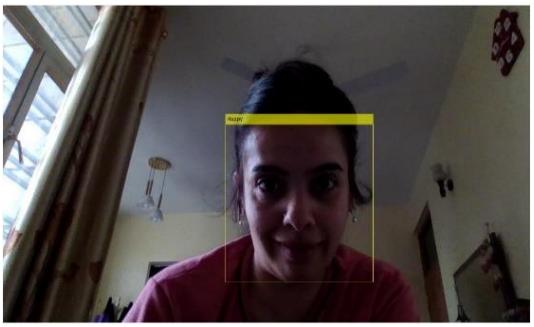

Figure. 8 Happy face

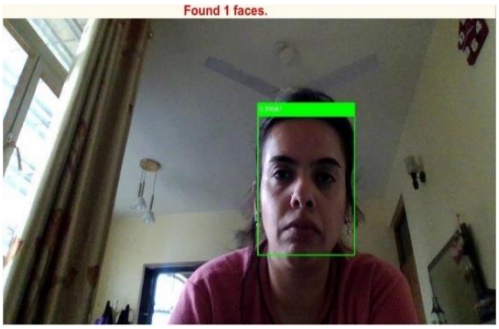

Figure. 9 Detected face

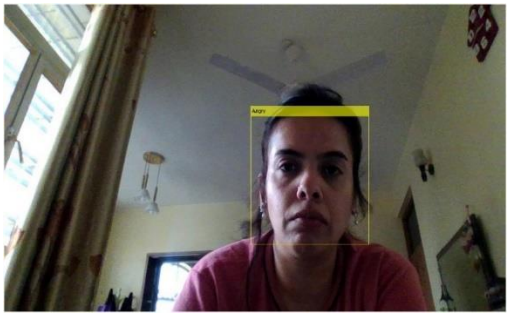

Figure. 10 Angry face

Figure. 11 Detected face

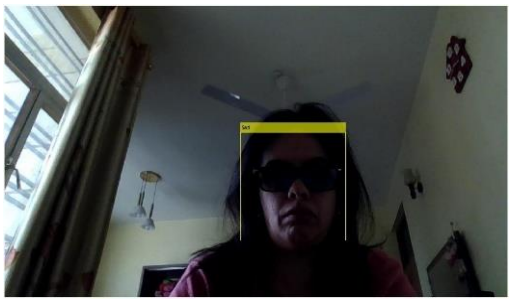

Figure. 12 Sad face

5.1 Performance Evaluation

Accuracy: It computes the percentage of genuine results, includes true positives and true negatives, throughout all instances analyzed [28]. This is stated in the Eq. (23),

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (23)$$

Fig. 13 depicts the accuracy of a deep learning model over iterations. As the number of iterations increases, accuracy improves. The x-axis denotes the number of iterations (ranging from 0 to 100), while the y-axis

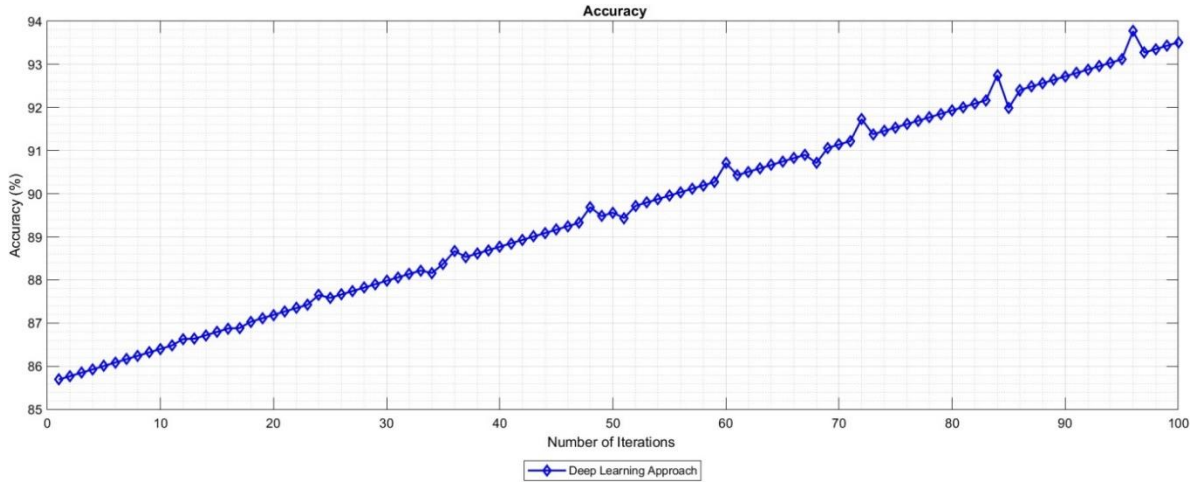

Figure. 13 Accuracy of a Deep Learning Model Over Iterations

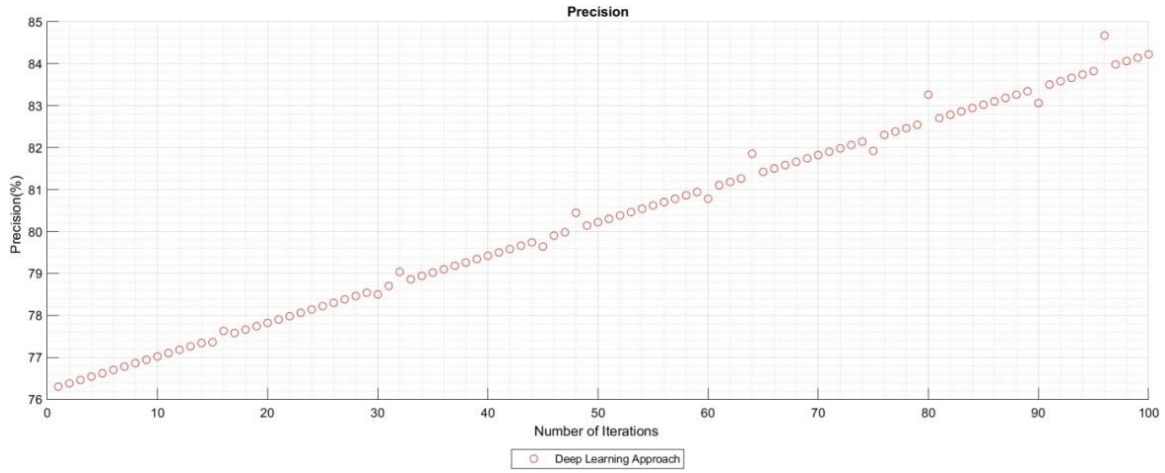

Figure. 14 Precision of a Deep Learning Model Over Iterations

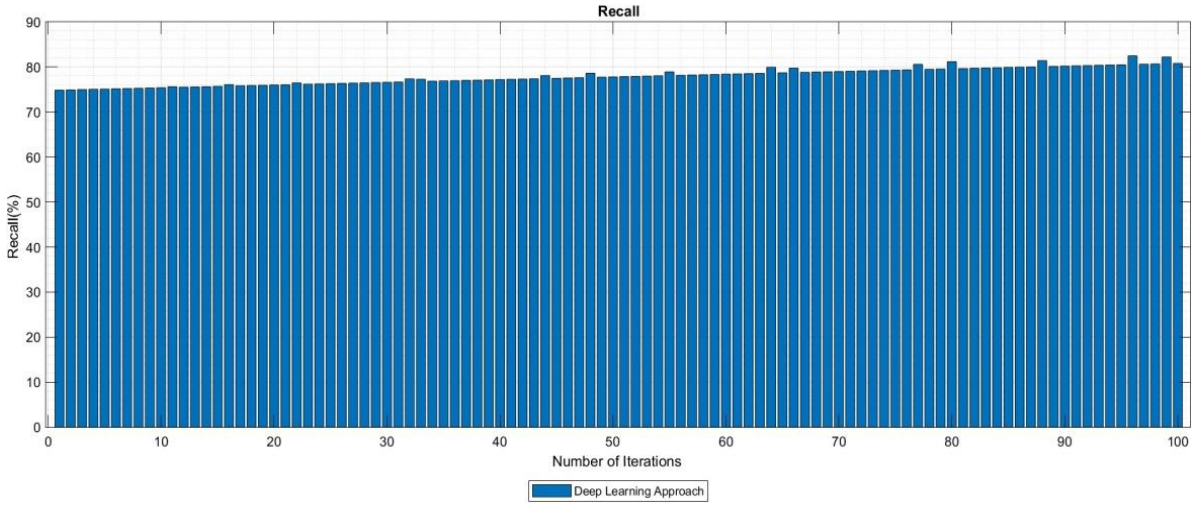

Figure. 15 Recall of a Deep Learning Model with Iterations

shows accuracy percentage (85% to 94%). The blue line with diamond markers indicates a consistent upward trend in accuracy as the model iterates.

Precision: Precision can be described as the ratio of accurately expected favorable results to total projected favourable instances [28]. The precision is obtained using Eq. (24). and is depicted in Fig. 14.

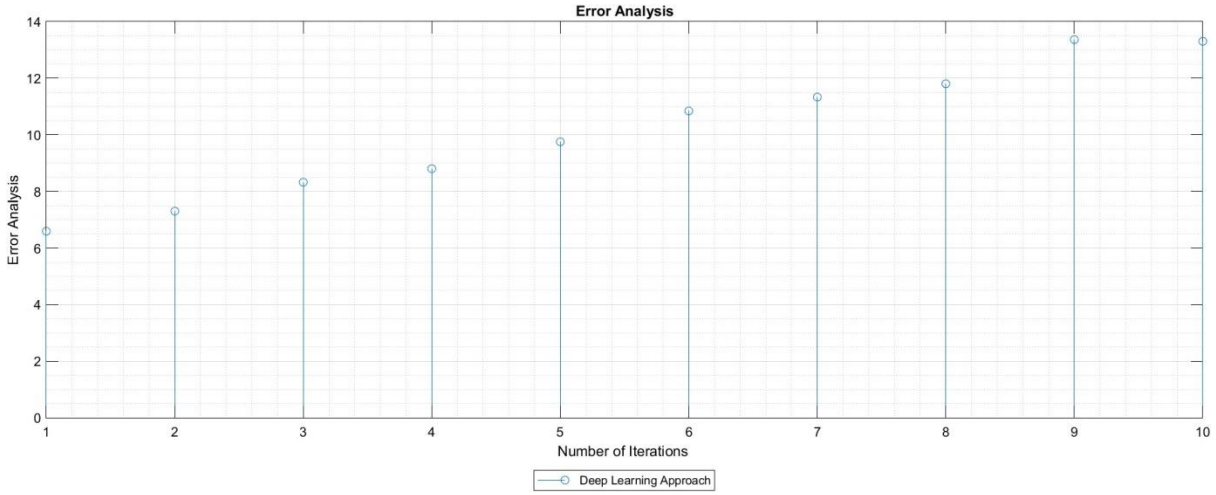

Figure. 16 Error Analysis

$$Precision = \frac{TP}{TP+FP} \quad (24)$$

Recall: The percentage of actual positive samples that were predicted to be positive is measured by the recall [28]. Compute the value recall using Eq. (25), as shown in Fig. 15.

$$Recall = \frac{TP}{TP+FN} \quad (25)$$

F1-score: In classification tasks, recall and accuracy are related. Whereas a large percentage of them is ideal, the truth is that outstanding precision is often accompanied with low recall, or vice versa. In order to adjust for both remembrance and accuracy, the F1-score, and this is the average of recall and preciseness, may be applied [28]. Eq. (26) displays the meaning of F1-score.

$$F1 - score = 2 * \frac{Precision*Recall}{Precision+Recall} \quad (26)$$

Error Analysis: Error analysis is a method used to assess the accuracy of a model's predictions by measuring the disparity between predicted values and actual observations. It is typically calculated utilizing a suitable error metric, Mean Absolute Error (MAE) [29], represented by the formula of Eq. (27):

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \quad (27)$$

Where, y_i signifies the authentic value, $\hat{y}_i$ represents the projected value, and n is the total amount of samples.

Fig. 16 shows the error analysis involves evaluating and interpreting the mistakes made by a model. In the graph, error rates fluctuate across different iterations of a deep learning approach, with errors peaking at certain iterations and decreasing at others.

Table 1. Performance Metrics of Proposed HOG-GSO Vs Traditional Methods

Methods	Accuracy (%)	Precision (%)	Recall (%)	MAE
CNN	80	78	82	13.42
RCNN	85	83.4	78.3	12.67
Proposed DCNN-GSO	94	92.3	91	3.46

Table 1 and Fig. 16 describes the three methods CNN, RCNN, and the proposed HOG-GSO—were evaluated for facial expression detection and is depicted in Fig. 16. The CNN: Achieved an accuracy of 80%, with precision at 78% and recall at 82%. The mean absolute error (MAE) was 13.42. RCNN: Demonstrated higher accuracy (85%) and precision (83.4%), but recall was slightly lower at 78.3%. The MAE improved to 12.67. The proposed HOG-GSO: Outperformed both with an impressive accuracy of 94%, precision at 92.3%, and recall at 91%. The lowest MAE of 3.46 indicates superior performance. The HOG-GSO method shows remarkable promise for accurate facial expression detection, surpassing existing approaches. Its robustness and precision make it a compelling choice for real-world applications.

Fig. 17 illustrates the performance of two algorithms, with "PSO Fitness" represented in blue and "GSO Fitness" in orange. Initially, the PSO algorithm exhibits higher fitness levels, followed by a gradual decline over the x-axis. In contrast, the GSO algorithm starts with superior fitness levels and

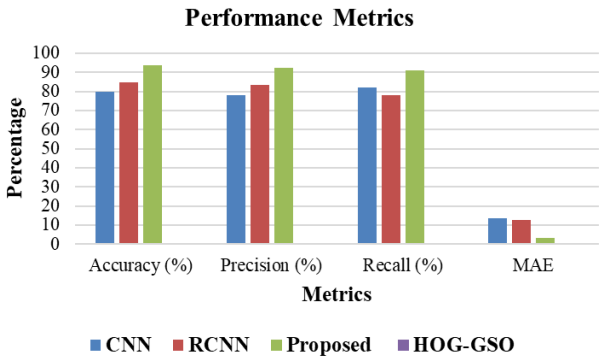

Figure. 17 Performance Metrics

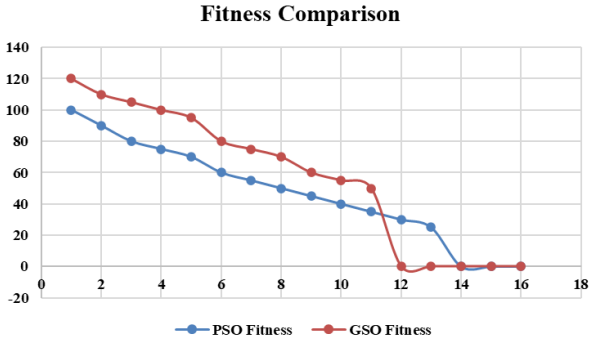

Figure. 18 Comparison of PSO and GSO

Table 2. Performance Comparison

Method	Dataset	Accuracy (%)
[8]	KDEF	82%
[9]	IRFacEx Net	88%
[10]	DEAP	92%
[11]	DEAP	64%
Proposed method	Kaggle dataset Emotion Detection with Webcam	94%

demonstrates a steeper decline. However, both algorithms converge around point "10," with their fitness levels becoming approximately equal. Overall, the GSO algorithm outperforms PSO, starting with higher fitness levels and showing a faster convergence rate.

Table 2 presents a performance comparison of various facial expression recognition methods across different datasets. Hossain et al.'s method [8] achieves an accuracy of 82% on the KDEF dataset, while the IRFacExNet model introduced by an unnamed source [9] attains 88% accuracy. Another approach [10] achieves a higher accuracy of 92% on the DEAP dataset, whereas Kopaczka et al.'s method [11] yields a lower accuracy of 64% on the same dataset. The proposed method, evaluated on the Kaggle dataset "Emotion Detection with Webcam," outperforms all other methods with an accuracy of 94%. These results highlight the effectiveness of the proposed method in achieving superior performance across different datasets.

6. Conclusion

Research presents a significant contribution to the field of facial expression analysis in IoT environments by proposing a comprehensive framework that integrates deep learning techniques with Galactic Swarm Optimization (GSO). The proposed Deep Convolutional Neural Network using Galactic Swarm Optimization (DCNN-GSO) approach addresses the challenges of limited computational resources and energy constraints inherent in IoT systems, offering a practical solution for real-time image processing tasks. By leveraging Gaussian filtering in the preprocessing phase to enhance image quality and reduce noise, and utilizing spatial temporal interest points (STIP) for feature extraction, the proposed method effectively captures spatial and temporal cues of facial expressions. The integration of DCNN with GSO optimizes the hyperparameters of the model, resulting in improved accuracy and robustness in facial expression classification tasks.

The scientific contribution of this work is evident in the remarkable results achieved by the DCNN-GSO method, outperforming existing methods with an impressive 94% accuracy, 92.3% precision, and 91% recall. Additionally, the very low mean absolute error (MAE) of 3.46 demonstrates the reliability and accuracy of the proposed solution for practical applications. By overcoming challenges such as noise, computational complexity, and accuracy limitations in IoT systems, this framework significantly enhances device performance and connectivity, opening new avenues for human-computer interaction, healthcare, security, and other industries. Furthermore, the successful integration of GSO with deep learning models for facial expression recognition showcases the potential of hybrid optimization techniques in overcoming resource constraints and improving the efficiency of IoT applications. The proposed framework provides a holistic approach to facial expression analysis in IoT environments, offering a scalable and robust solution for real-world deployment. Overall, this research advances the state-of-the-art in facial expression analysis and demonstrates the feasibility of employing deep learning techniques in IoT scenarios to enhance understanding and communication between humans and devices.

Conflicts of Interest

The authors declare that they have no relevant conflicts of interest.

Author Contributions

Conceptualization, Rana.H.AL-Abboodi and Ayad..A.AL-An; methodology, Rana.H.AL-Abboodi; software, Rana.H.AL-Abboodi; validation, Rana.H.AL-Abboodi, and Ayad..A.AL-An; formal analysis, Rana.H.AL-Abboodi; investigation, Rana.H.AL-Abboodi; resources, Rana.H.AL-Abboodi; data curation, Rana.H.AL-Abboodi; writing—original draft preparation, Rana.H.AL-Abboodi; writing—review and editing, Rana.H.AL-Abboodi; visualization, Aayd A.AL-Ani; supervision, Rana.H.AL-Abboodi; project administration, Ayas A.AL-Ani.

References

[1] A. F. Klaib, N. O. Alsrehin, W. Y. Melhem, H. O. Bashtawi, and A. A. Magableh, "Eye tracking algorithms, techniques, tools, and applications with an emphasis on machine learning and Internet of Things technologies", *Expert Syst. Appl.*, Vol. 166, p. 114037, 2021.

[2] N. Singh and H. Sabrol, "Convolutional neural networks-an extensive arena of deep learning. A comprehensive study", *Arch. Comput. Methods Eng.*, Vol. 28, No. 7, pp. 4755-4780, 2021.

[3] A. Hassouneh, A. Mutawa, and M. Murugappan, "Development of a real-time emotion recognition system using facial expressions and EEG based on machine learning and deep neural network methods", *Inform. Med. Unlocked*, Vol. 20, p. 100372, 2020.

[4] M. Alshamrani, "IoT and artificial intelligence implementations for remote healthcare monitoring systems: A survey", *J. King Saud Univ.-Comput. Inf. Sci.*, Vol. 34, No. 8, pp. 4687-4701, 2022.

[5] R. Abdulkader *et al.*, "Optimizing student engagement in edge-based online learning with advanced analytics", *Array*, Vol. 19, p. 100301, 2023.

[6] P. W. Khan, Y.-C. Byun, and N. Park, "A data verification system for CCTV surveillance cameras using blockchain technology in smart cities", *Electronics*, Vol. 9, No. 3, p. 484, 2020.

[7] S. Hossain, S. Umer, V. Asari, and R. K. Rout, "A unified framework of deep learning-based facial expression recognition system for diversified applications", *Appl. Sci.*, Vol. 11, No. 19, p. 9174, 2021.

[8] S. Hossain, S. Umer, V. Asari, and R. K. Rout, "A Unified Framework of Deep Learning-Based Facial Expression Recognition System for Diversified Applications", *Appl. Sci.*, Vol. 11, No. 19, p. 9174, 2021, doi: 10.3390/app11199174.

[9] A. Bhattacharyya, S. Chatterjee, S. Sen, A. Sinitca, D. Kaplun, and R. Sarkar, "A deep learning model for classifying human facial expressions from infrared thermal images", *Sci. Rep.*, Vol. 11, No. 1, p. 20696, Oct. 2021, doi: 10.1038/s41598-021-99998-z.

[10] A. R. Aguiñaga, D. E. Hernandez, A. Quezada, and A. Calvillo Téllez, "Emotion Recognition by Correlating Facial Expressions and EEG Analysis", *Appl. Sci.*, Vol. 11, No. 15, p. 6987, Jul. 2021, doi: 10.3390/app11156987.

[11] M. Kopaczka, L. Breuer, J. Schock, and D. Merhof, "A Modular System for Detection, Tracking and Analysis of Human Faces in Thermal Infrared Recordings", *Sensors*, Vol. 19, No. 19, p. 4135, Sep. 2019, doi: 10.3390/s19194135.

[12] S. Hossain, S. Umer, V. Asari, and R. K. Rout, "A unified framework of deep learning-based facial expression recognition system for diversified applications", *Appl. Sci.*, Vol. 11, No. 19, p. 9174, 2021.

[13] "Emotion Detection with Webcam.", Accessed: Feb. 17, 2024. [Online]. Available: https://kaggle.com/code/khandelwalmitesh3/emotion-detection-with-webcam

[14] M. Singh *et al.*, "A facial and vocal expression based comprehensive framework for real-time student stress monitoring in an IoT-Fog-Cloud environment", *IEEE Access*, Vol. 10, pp. 63177-63188, 2022.

[15] C. Tomasi and R. Manduchi, "Bilateral Filtering for Gray and Color Images".

[16] I. Laptev, "On space-time interest points", *Int. J. Comput. Vis.*, Vol. 64, pp. 107-123, 2005.

[17] H. Wang, S. Wei, and B. Fang, "Facial expression recognition using iterative fusion of MO-HOG and deep features", *J. Supercomput.*, Vol. 76, pp. 3211-3221, 2020.

[18] B. Sugiarto *et al.*, "Wood identification based on histogram of oriented gradient (HOG) feature and support vector machine (SVM) classifier", In: *Proc. of 2017 2nd International conferences on Information Technology, Information Systems and Electrical Engineering (ICITISEE)*, IEEE, pp. 337-341, 2017.

[19] Y. Yaddaden, M. Adda, and A. Bouzouane, "Facial expression recognition using locally linear embedding with lbp and hog descriptors",

In: *Proc. of 2020 2nd International Workshop on Human-Centric Smart Environments for Health and Well-being (IHSH)*, IEEE, pp. 221-226, 2021.

[20] J. Noor, M. Daud, R. Rashid, H. Mir, S. Nazir, and S. A. Velastin, "Facial expression recognition using hand-crafted features and supervised feature encoding", In: *Proc. of 2020 International Conference on Electrical, Communication, and Computer Engineering (ICECCE)*, IEEE, 2020, pp. 1-5.

[21] Pankaj, P. Bharti, and B. Kumar, "A New Design of Occlusion-Invariant Face Recognition Using Optimal Pattern Extraction and CNN with GRU-Based Architecture", *Int. J. Image Graph.*, Vol. 23, No. 04, p. 2350029, 2023.

[22] A. S. Alphonse, S. Abinaya, and K. Arikumar, "A novel monogenic Sobel directional pattern (MSDP) and enhanced bat algorithm-based optimization (BAO) with Pearson mutation (PM) for facial emotion recognition", *Electronics*, Vol. 12, No. 4, p. 836, 2023.

[23] M. Jiang *et al.*, "IoT-based remote facial expression monitoring system with sEMG signal", In: *Proc. of 2016 IEEE sensors applications symposium (SAS)*, IEEE, pp. 1-6, 2016.

[24] M. Z. Uddin, M. M. Hassan, A. Almogren, M. Zuair, G. Fortino, and J. Torresen, "A facial expression recognition system using robust face features from depth videos and deep learning", *Comput. Electr. Eng.*, Vol. 63, pp. 114-125, 2017.

[25] M. Masud, G. Muhammad, H. Alhumyani, S. S. Alshamrani, O. Cheikhrouhou, S. Ibrahim, and M. Shamim Hossain, "Deep learning-based intelligent face recognition in IoT-cloud environment", *Comput. Commun.*, Vol. 152, pp. 215-222, 2020.

[26] S. Talukder, "Mathematicle modelling and applications of particle swarm optimization", *Master's Thesis Mathematical Modelling and Simulation Thesis*, No. 2010:8, 2011.

[27] E. Kaya, İ. Babaoğlu, and H. Kodaz, "Galactic swarm optimization using artificial bee colony algorithm", In: *Proc. of 2017 15th International Conference on ICT and Knowledge Engineering (ICT&KE)*, IEEE, pp. 1-6, 2017.

[28] A. R. Khan, "Facial emotion recognition using conventional machine learning and deep learning methods: current achievements, analysis and remaining challenges", *Information*, Vol. 13, No. 6, p. 268, 2022.

[29] D. Chaves, E. Fidalgo, E. Alegre, R. Alaiz-Rodríguez, F. Jáñez-Martino, and G. Azzopardi, "Assessment and estimation of face detection performance based on deep learning for forensic applications", *Sensors*, Vol. 20, No. 16, p. 4491, 2020.