Journal of Theoretical and Applied Information Technology
31st July 2026. Vol.104. No.14

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

DYNAMIC SOCIAL SKI-DRIVER OPTIMIZATION WITH DEEP CAPSULENET ARCHITECTURE FOR THE DETECTION AND CLASSIFICATION OF EMOTION USING FACIAL FEATURES

[1]S.Srinivas, [2]Dr. Mercy Paul Selvan

Research Scholar, Sathyabama Institute of Science and Technology, Chennai,India

Email: srinivas.sbu@gmail.com

Professor,School of Computing, Sathyabama Institute of Science and Technology Chennai,India

Email: mercypaulselvan.cse@sathyabama.ac.in

ABSTRACT

Facial expression is the universal language for humans to express their feelings and detect emotions via the computer vision approach and is a challenging process. Emotion detection by the machine is arduous with the presence of accessories, pose variation, non-uniform illumination, etc. Several works have been carried out in this field, however, the mutual optimization of feature extraction and detection of emotion has not been achieved. To tackle this issue, we propose an innovative technique known as optimized deep learning. To begin the process, the data that are collected are pre-processed, and the facial features associated with the Dynamic social ski-driver optimization (DSSO) model are incorporated with Deep CapsuleNet approach. The derived features are used to detect the emotions using the proposed technique. Simulations are effectuated in the MATLAB simulator and analyzed the robustness of the proposed work. The proposed approach is compared with the previous works using the statistical parameters and achieved an increased accuracy of around 98%.

Keywords: *Facial Features, Characterization, Emotion Detection, Dynamic Social Ski-Driver Optimization, And Deep Learning.*

1. INTRODUCTION:

In the past decade, human-robot collaboration has taken on an important function in the progress of automation; the key to revealing emotional states between individuals and machines is the recognition of facial expressions [1]. Expressions of emotion convey implicit thoughts and might be interpreted as introspective cues that influence one's assumptions and decision-making. Emotional states are their feelings, which are created by synapses in the cerebral cortex that fire particles down minute paths, as such, it is quite possible to experience emotions without expressing them with your face. Emotion Classification [2] refers to the process of mechanically identifying sentiments from appearances to enable a device to understand individual spoken and physical sentiments. Even while computers find it difficult to understand reactions in people, individuals are highly skilled at it, modifications in lighting and facial positions, blockage, and processing intricacy are the causes. Since facial expressions reflect sensations [3], they serve as essential indicators of individual attitudes; the task of autonomously deducing an individual's emotions from their facial movements is both fascinating and challenging.

In the majority of occasions approximately 55% of situations, a person's facial expressions are a communicative means of expressing their emotions, and they can be used as conclusive proof to determine the extent to which they are telling the reality, moral conduct is somewhat influenced by sensations in day-to-day interactions. Emotion [4] is a characteristic of people with intellect that has a significant effect on the effectiveness of interpersonal sharing of knowledge and human labor. Communicating with human platforms must incorporate emotion detection to enhance engagement and advance the realization of computing's overall data, given the swift progress of digital technology [5]. It takes a lot of expertise and personal thought to judge another person's human feelings, which is difficult for humanity, much less gadgets. A series of new faces that showed pleased or furious emotions were presented to those who responded, the characters were either facing forward or in the other direction. When features with furious attitudes were originally demonstrated, retention for those grins was better when the participants had an oblique look rather than one that was focused. The nervous system is capable of feeling emotions when recognizing faces [6] and the adrenal cortex plays a major role in the identification mechanism, the integration site for incentive, sensations, and mental conduct is the cingulated.

However, recall for people with pleasant faces remained unchanged regardless of the angle of focus, it has been hypothesized that a person's ability to remember another person's face is somewhat dependent on how well they understand that person's motives for acting. Either freely or unintentionally, individuals may assume an attitude or the brain processes that govern the gesture are different in each scenario. On the other hand, it is thought that spontaneous gestures originate in the nervous system and travel through the region. Face-to-face communication [7] is crucial for networking because glances can convey genuine thoughts and objectives in a group setting. A person's visage is the primary means of conveying feelings in addition to information and concepts. Conversational signals [8] can be difficult to read and are readily misinterpreted by us. An electronic message lacks, for instance, an accent of pronunciation, gestures, and emotional expressions. This work proposed a novel Dynamic social ski-driver optimization (DSSO) model are incorporated with Deep CapsuleNet approach for the categorization of various kinds of human emotion such as sad, surprise, happy and fear.

This study proposes a DSSO-Deep CapsuleNet framework to improve facial emotion recognition by jointly optimizing network parameters and preserving spatial facial feature relationships. Readers can expect a comprehensive evaluation of the proposed framework against existing optimization-based methods, demonstrating its advantages in terms of accuracy, sensitivity, specificity, and convergence performance. The findings of this study provide a foundation for developing robust emotion recognition systems applicable to healthcare, intelligent human-computer interaction, surveillance, and other real-time computer vision applications.

This section organizes the remainder of the paper; the literature review is discussed in Section 2 and the proposed methodology details are given in Section 3. The exploratory analysis is described in Section 4 and the conclusion and future scope are addressed in Section 5.

2. LITERATURE SURVEY:

Alenazy et al. [9] have presented a deep belief network gravitational search algorithm (DBN-GSA) approach to precisely categorize facial reactions. A feature selection method removes any undesired knowledge from the graphic. The procedure for obtaining features is carried out to identify irregular elements and get more complex interactions through input parameters. Simulations were conducted using two popular databases, and the outcomes show how effective the suggested method is. It has the greatest durability range for the original information and reduces processing duration and storage usage. Hence, the capacity to generalize is limited when using a cross-dataset.
Sikkandar et al. [10] have described an Improved Cat Swarm Optimization (ICSO) algorithm to choose the best attributes from an individual's picture that allow one to discern an individual's reaction to the visage. A framework for automatically recognizing facial features is created to find photographs in the collection that are comparable to the inquiry photograph and to identify the attitudes on each of those photos. It is used for component selection, but it eliminates qualities that aren't important and chooses the best characteristics, enhances precision, and decreases calculation duration. However, gathering an enormous quantity of instructional information under diverse circumstances is hard.

Gao et al. [11] have implemented a gradient-priority particle swarm optimization (GPSO) a platform for dynamically optimizing using encoded binary data. A single network's characteristics and architecture can be altered at once via randomization. The outcome shows that the structure may create a level of precision that is, typically, higher than the reliability of the four preceding efforts, indicating the success of the strategy. Additionally, reliability is enhanced dramatically after applying the optimized technique. Therefore, this technique's extensive and quick application is restricted.
Liu et al. [12] have developed a genetic algorithm support vector machine (GA-SVM) for a multi-attribute efficiency issue of highlight and decision-making. A genetic algorithm is used to

Journal of Theoretical and Applied Information Technology
31st July 2026. Vol.104. No.14

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

automatically adjust variables to have the best condition functionality. Besides, the technique demonstrated continuous and controlled performance over diverse equipment-acquiring simulation assessments, indicating that the suggested paradigm is secure, harmonious, and receptive to realistic favorable occurrences numerically. More than 90 percent of the time, the GA-SVM approach for a facial process that was provided was recognized. Moreover, it's challenging to understand the signs in the indicated prominent places.

Said et al. [13] have evaluated a face-sensitive convolutional neural network (FS-CNN) is employed to identify features in massive amounts of photos, and then facial pattern analysis is utilized to forecast behaviors for recognizing feelings. In the initial step, faces in pictures with a great resolution are identified, and those images are cropped for additional editing. Dimensional consistency was processed on system photos using a CNN in the subsequent platform, which predicts facial emotion according to location statistics. The sample was used to educate and assess the suggested FS-CNN. Impressive results were obtained, with a standardized consistency of almost 95%. Thus, the wavelength coherence is generated, which is a disadvantage.

Mehendale et al. [14] have modified facial emotion recognition using convolutional neural networks (FERC) offered in this publication seeks to categorize the photograph supplied into the aforementioned five key sentiment types and examine representation. With an expressional vector (EV) of range 24 standards, sentiment may be accurately highlighted with a precision of 96%. The final component of the sensor modifies the coefficients and parameter quantities with every repetition, while the CNN operates in parallel. It improves precision by deviating from commonly used CNN techniques. Nevertheless, it's difficult to accomplish an identical goal with a computation for a computer.

Saurav et al. [15] suggested Emotion Network (EmNet) model provides three forecasts for a specific visual reputation, which are merged employing the standard combining and scaled greatest integration algorithms to provide the ultimate assessment. It uses two integration strategies, the normal amalgamation technique and the weighted-maximum synthesis strategy, to mix the EmNet predictive estimate outputs and get the ultimate choice point. The suggested approach is as far as processing speed and detection precision. Thus, quantification intelligence analysis of reactions is challenging.

Ch et al. [16] highlighted Deep Learning Neural Network-regression activation (DR) system especially compared to the current methods, is greater. Preliminary analysis, face feature removal, differentiation, and grouping steps make up this suggested approach. These procedures identify the following expressive sensations: unfavorable anxiety, joy, sorrow, dissatisfaction unexpectedly, and aggression. The effectiveness of the suggested and current approaches is examined for each feeling in the two samples. It follows that the suggested setup performs at an excellent standard. However, there are still difficulties with facial photographs in real-life scenarios.

3. PROPOSED FRAMEWORK:

Figure 1 explains the overall architecture of the proposed model. It incorporates steps like pre-processing, feature extraction, and human emotion categorization with respect to facial features, and every step is comprehensively explained in the following sub-sections.

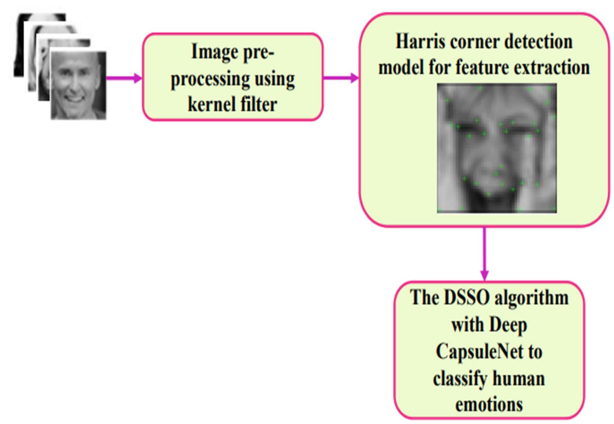


Fig 1: Proposed work architecture

3.1 Pre-processing using kernel filter:

Pre-processing of the input images is effectuated with the proposed Kernel filter [17] and it is a type of kernel-based smoothing technique. This mainly utilized input vector images of the facial expression detection to create the smoothed output of the images. This employs a particular type of filter and the process is effectuated with the convolution of the input sequence and the kernel. The convolution process is done with the Fast Discrete Fourier transform [18].

Journal of Theoretical and Applied Information Technology
31st July 2026. Vol.104. No.14

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

3.2 Feature Extraction:

The dynamic nature of the facial movements and varying expressions are made in the temporal domain and are used for feature extraction. Harris corner detection model like Haar-like feature extraction technique is used to obtain the temporal variations with 2D coordinate values with nonstationary and time-varying functions. When compared to the Gabor features the Haar-like feature has less computation and belongs to the wavelet family [19]. Moreover, using this, it acts as the filter and decomposes the raw signals to the coefficients such as detailed and approximation. Meanwhile, the high-frequency bands in the detailed coefficients are clustered using the technique. The Haar function is described as,

$$\Psi_{x,y}(t) = x^{-1/2}\Psi\left(\frac{t-y}{x}\right) \tag{1}$$

$$\Psi(t) = \begin{cases} 1; & 0 \le t < 1/2 \\ -1; & 1/2 \le t < 1 \\ 0; & otherwise \end{cases} \tag{2}$$

Using this the statistical features like, energy, mean, entropy, power, standard deviation, and variance of the features are extracted with the 4th-level approximation technique and are forwarded to the CapsuleNet network for further action.

3.3 Facial Emotion Detection based on Deep CapsuleNet with DSSO model:

The proposed Dynamic social ski-driver optimization (DSSO) with Deep CapsuleNet model to detect facial emotion recognition and classification is designed in the following section.

3.3.1 DSDO algorithm

The downhill path of ski drivers resembles the stochastic exploration based on the Social Ski Driver Optimization (SSO) algorithm. The non-linear functions are solved using this SSO algorithm. The search space based on the peak point is determined by using the SSO algorithm. There before meeting the termination criteria, four different steps inspire the SSO model.

(i) Initialization:

In the beginning, determines the required and fixed number of agents such as Y_j. Randomly select the search space position and the objective function on every position determines the fitness value [20].

(ii) Optimal position finding:

Calculate each agent fitness value with the usage of the objective function. According to the previous iteration, the value compares every agent with its fitness value. Store the best agent $B_{po}(t)$.

(iii) Global mean solution:

The global mean function ($m(t)$) is determined with the fitness value based on the best 3 mean agents. The global mean point tends to every agent.

$$m(t) = \frac{Y_\beta(t) + Y_\chi(t) + Y_\delta(t)}{3} \tag{3}$$

From the objective function, the fitness value determines the initial, secondary, and tertiary best positions as $Y_\beta(t)$, $Y_\chi(t)$ and $Y_\delta(t)$.

(iv) Update agent position and velocity:

The following equations update the velocity and agent position as $vl_j(t)$ and $Y_j(t)$. Where, [0, 1] is the random number stands for Rdn_1 and Rdn_2. According to the j^{th} agent, the best solution is $B_{po}(t)$. Utilize d to generate the balance between exploitation and exploration. Based on iterations, increasing the value d is reduced by using $0 < \beta < 1$ met the maximal iteration is $\max_t$.

$$d = d(t+1) = \beta \cdot d(t) \tag{4}$$

$$\begin{cases} \left(vl_j(t+1)\right) = d \cdot \sin(Rdn_1)\cdot(B_{po}(t) - Y_j(t)) + \\ \quad \sin(Rdn_1)\cdot(N - Y_j(t)) & if \quad Rdn_2 \le 0.5 \\ \left(vl_j(t+1)\right) = d \cdot \cos(Rdn_1)\cdot(B_{po}(t) - Y_j(t)) + \\ \quad \cos(Rdn_1)\cdot(N - Y_j(t)) & if \quad Rdn_2 > 0.5 \end{cases} \tag{5}$$

Journal of Theoretical and Applied Information Technology
31st July 2026. Vol.104. No.14

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

An opposition learning function effecuates the SSO enhancement for attaing best solution. For that reason, an opposite number and points rely a model of OBL. The ongoing solution approximation generates the oposite validation, which is named as dynamic social ski-driver (DSSO) model. Because of cosine and sine functions, the global maximum to the movement is not under a straight direction. Diversifies the search direction that generates good exploration capacity.

$$Y_j(t+1) = Y_j(t) + vl_j(t+1) + W^0 \tag{6}$$

$$W^0 = m + n - W \tag{7}$$

An opposite numbers m and n liked to the points $p(W_1, W_2, ..., W_m)$. Repeat the process at the end of the termination criteria after updating the position $Y_j(t+1)$.

3.3.2 DSSO with Deep CapsuleNet

CNN has had several successes in the deep learning field but it met some obstacles and its ability has failed in case of a few main tasks. Image detection and recognition features are easily learned due to the better feature extraction of CNN [21]. The representation of the upper scalar is transformed via the primary capsule layer. The loss affected by pooling is avoided and the network parameters are updated via a dynamic routing algorithm. Based on test samples towards a few classes, the probability is the length of the Eigenvector as the final output. The parametruic tuning of CapsuleNet achived with the help of DSSO algorithm thereby providing accurate classification outcomes.

Each neuron is scalar in the fully connected network and the weight is represented as wt. The numeric value and the scalar are every weight. The vector is every neuron capsule in DSSO with CapsNet. Based on the back-propagation, updates every capsule neuron weights.

$$S_k = \sum_j a_{jk} \hat{v}_{k/j}, .\hat{v}_{k/j} = wt_{jk} v_j \tag{8}$$

The learnable weight and the upper CapsuleNet outputs are wt_{jk} and v. Each output multiplies the matrix weight among j^{th} and k^{th} capsules. Calculates the coefficient (a) auxiliary with the linear sum.

$$a_{jk} = S_{\max}(c_{jk}) = \frac{Exp(c_{jk})}{\sum_i Ex(c_{jk})} \tag{9}$$

Forward propagation (S) to calculate the process and obtain the next layer belongs to DSSO. While squashing the activation function, the function of activation is sigmoid. Based on the loss function, update the weight with respect to entire network of other convolution parameters with repct to DSSO.

$$l_a = \sum_{i \in a_{num}} t_i Max\left(0, n^+ - \|U_i\|^2\right) + \kappa(1-t_i) Max\left(0, \|U_i\| - n^-\right)^2 \tag{10}$$

The facial class prediction values are n^+ and n^-. The accurate label is represented using $t_i = 1$ and the vector probability length is $\|U_i\|$ with category number is expressed in terms of i. This research suggested a novel DSSO algorithm with Deep CapsuleNet to detect and categorizes various kinds of human emotions like sad, surprise, happy and fear from human face.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

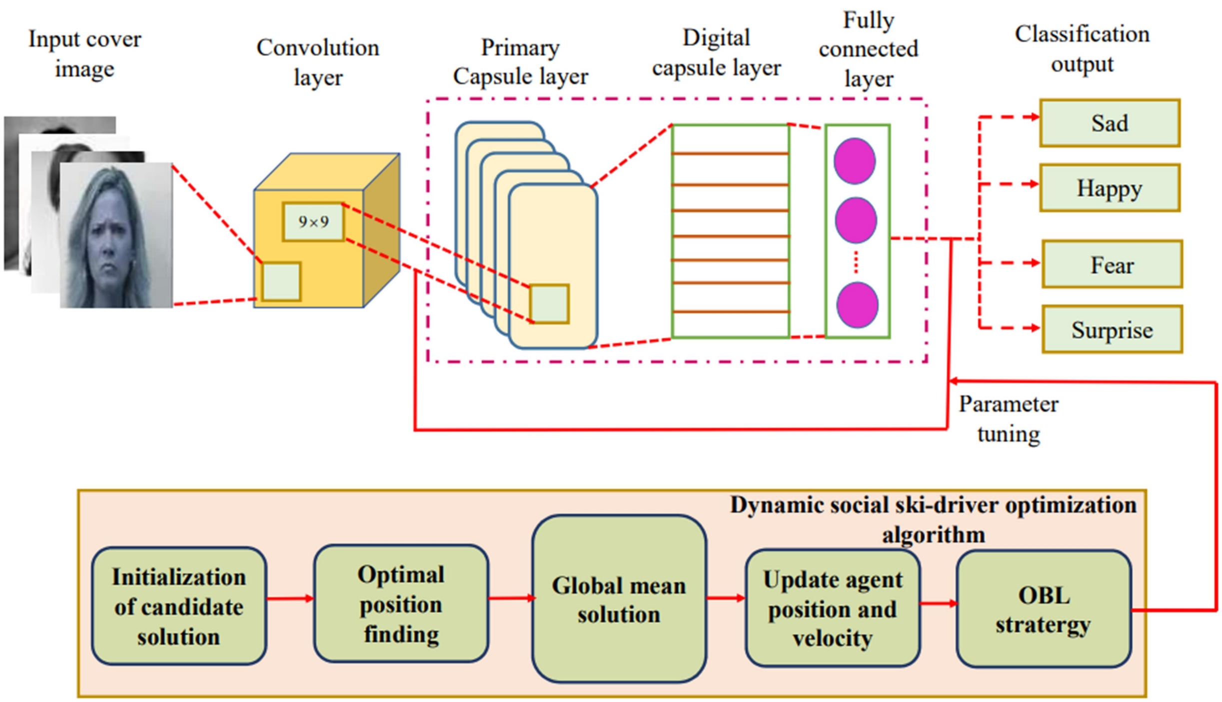


Fig 2: Proposed DSSO based CapsuelNet to classify human facial emotions

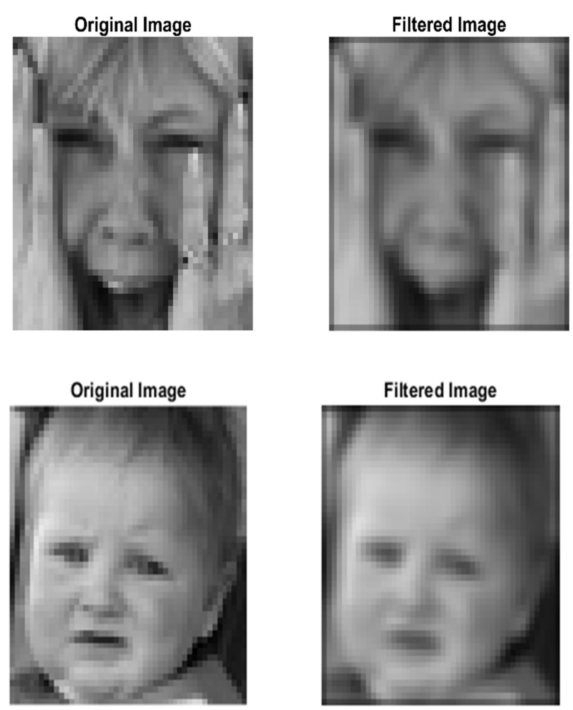
4. EXPERIMENTAL INVESTIGATION AND DISCUSSION:

This section extends the investigation of the performance of the proposed work while it is implemented along with the dataset description and experimental setup. This also presents the comparative study of proposed and prevision techniques. Simulation is conducted on MATLAB 2019a with the system specification of RAM 12 GB of Intel Xenon Gold operating system with 6145 CPU of 64 bit ELF structure with the memory of 96GB.

4.1 Dataset Description

For the analysis, we have taken the Kaggle dataset [22] which contains 35,686 grayscale images with 48X48 pixels. The images are split into training and testing datasets and include different expressions of anger, disgust, sadness, neutral, fear, surprise, and happiness.

For the pre-processing we have adopted the kernel filter and the output based on the pre-processing is shown in figure 3. Figure 3 illustrates both the original and filtered two sample images taken from the Kaggle dataset. The proposed filter robustly works in the images and removes the noises and other unnecessary factors from the images. The filtered images are smoothed from the original images as displayed in the figure.



Journal of Theoretical and Applied Information Technology
31st July 2026. Vol.104. No.14

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Fig 3: Preprocessing output

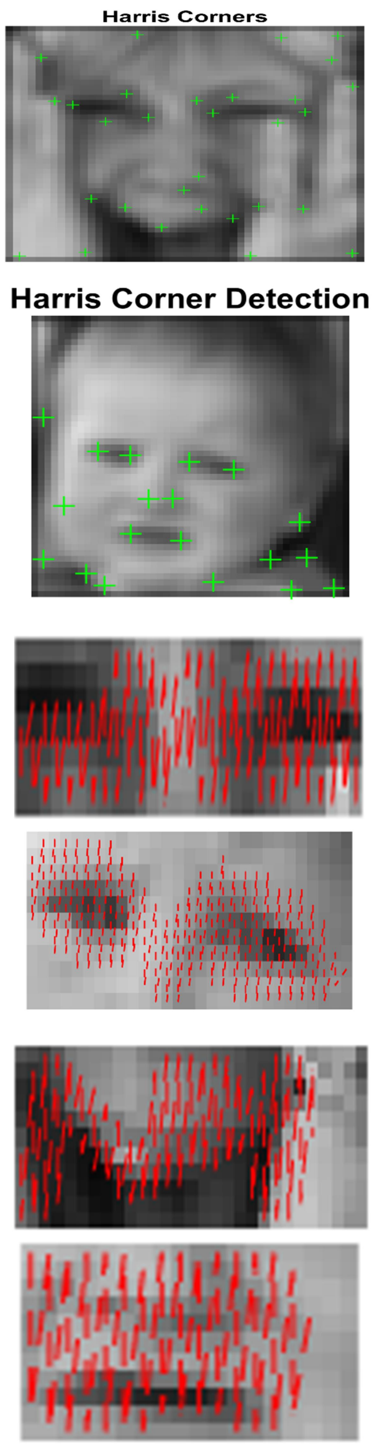


Fig 4: Feature extraction using Haar-like feature

The output of the proposed Haar-like feature extraction is displayed in Figure 4. In the figures, the Harris Corners are effectively spotted the particular parts to identify the different types of emotions. The pre-processed images are forwarded to the Haar-like feature extraction block and extract the required features like eyes and mouth as shown in the figure. The classification output of the proposed Deep CapsuleNet-based Dynamic social ski-driver optimization (DSSO) model is shown in Figure 5. It effectively classifies the emotions such as fear, happy, sad and surprises as shown. Since the selection of optimization is exact it provides better output.

Fig 5: Proposed DSSO- Deep CapsuleNet classification output

4.2 Performance Investigation

This section presents the in-depth analysis of the proposed approach based on confusion matrix, training and testing accuracy and loss, accuracy, sensitivity, specificity, and execution time. The training and testing accuracy and loss are plotted in Figure 6. From the figure it is observed that the learning rate of the proposed technique is 0.001 with the hardware requirement of a single CPU. The total iterations of 6040 are made with 20 epochs. One epoch consists of 302 iterations and the frequency of the proposed work is 10 iterations.

Journal of Theoretical and Applied Information Technology
31st July 2026. Vol.104. No.14

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

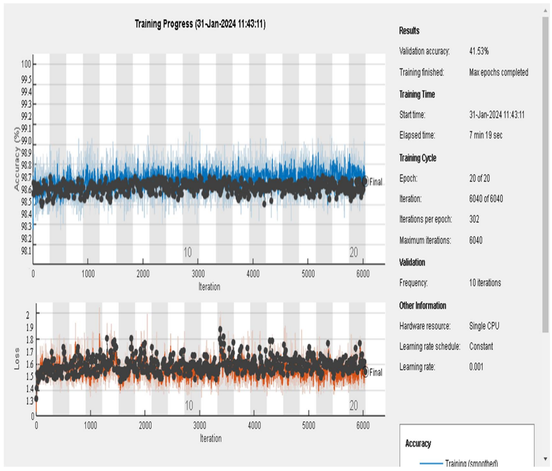


Fig 6: Training accuracy and loss for 20 epochs

The confusion matrix for all the samples is plotted in Figure 7. The confusion matrix is significantly sophisticated in the effectiveness of classifying emotions accurately as fear, happiness, sad, and surprises. From the samples about 800 images are classified as fear, 1215 are classified as Happy, 1024 as classified as sad, and 1798 as surprise as mentioned in the confusion matrix.

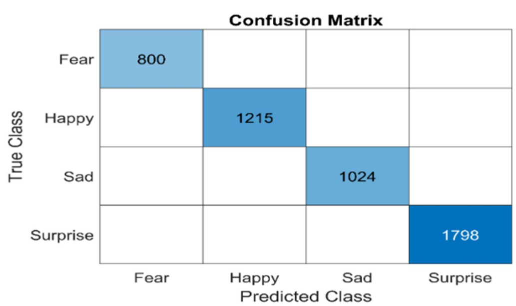


Fig 7: Confusion matrix for the proposed Deep CapsuleNet-based DSSO technique for facial features detection

The accuracy of the proposed approach when the epoch is about 20 is 95.2% and increases with the epochs. At the 40th epoch, the accuracy is about 96.23%, at the 60th epoch the accuracy is around 97. 5.7%, at the 80th epoch the accuracy is around 98.21%, and attains a saturation accuracy when the epoch is around 100. At the 100th epoch, the accuracy is about 98.62% as shown in figure 8. Figure 8 represents the graphical representation of the proposed work in terms of accuracy.

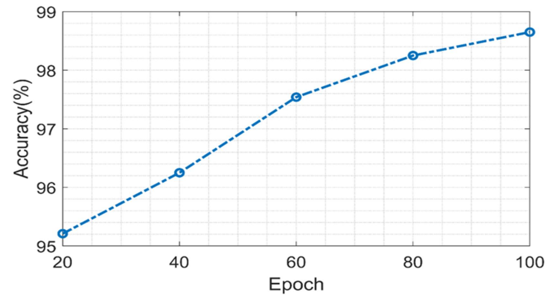


Fig 8: Performance analysis based on the accuracy for various epochs

The sensitivity of the proposed work is to measure the robustness of the proposed work in classifying the emotions as happy, sad, surprised, and fearful. The sensitivity for varying epochs is graphically plotted in Figure 9. When the epoch is 20, the sensitivity of the proposed work is 95.11%, at the 40th epoch the sensitivity is around 95.3%. At the 60th epoch, the sensitivity is around 95.65% and at the 80th epoch, the sensitivity is around 96.2.5%. At the 100th epoch, the saturation point begins and attains the saturation sensitivity of 96.87%. The sensitivity of the proposed approach is higher.

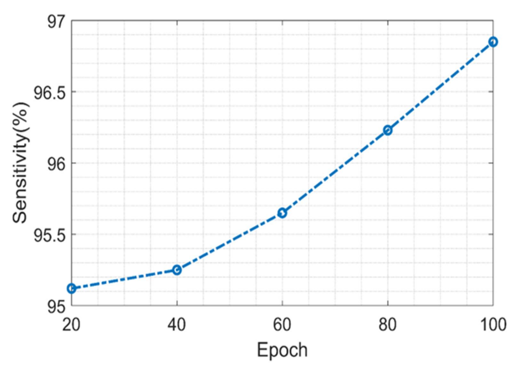


Fig 9: Performance analysis based on the Sensitivity for various epochs

The performance of the proposed technique in terms of specificity is graphically plotted in Figure 10. The specificity of the proposed technique is higher with 97.12% at the 100th epoch. Meanwhile, at the 20th epoch, the specificity is around 94.2%, at the 40th epoch the specificity is 94.56%, at the 60th epoch the specificity is 95.3%. At the 80th epoch, the specificity is around 96.15%. This is achieved with the inclusion of robust techniques from pre-processing to emotion classification.

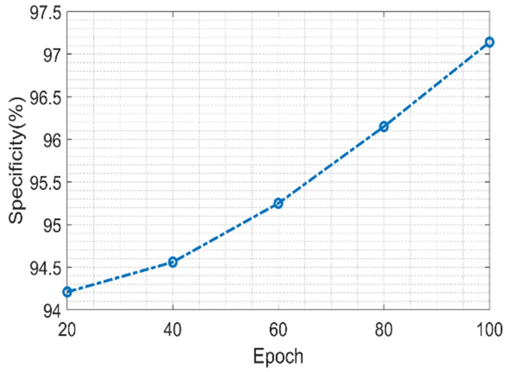


Fig 10: Performance analysis based on the Specificity for various epochs

The performance evaluation based on the execution time for the proposed technique is visually plotted in Figure 11. The figure is plotted for 100 epochs and at the 20th epoch the execution time is around 6.12 s, at the 40th epoch the proposed technique takes 6.14 s, at the 60th epoch the execution time of the proposed technique to complete the process is 6.18 s. when the epoch is around 80, the execution time for the proposed technique is around 6.24 s. Finally, for the 100th epoch, the execution time is around 6.32 s is depicted in figure 11.

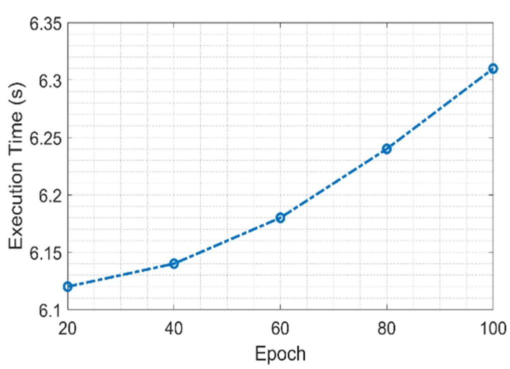


Fig 11: Performance analysis based on the Execution time for various epochs

For the comparative study, we have taken the parameters such as error vs. convergence time and overall performance is plotted in figure 12 and figure 13. The first parameter is compared with the various optimization techniques such as Ant Colony Optimization (ACO), Genetic Algorithm (GA), Black Widow optimization (BWO), and Zebra optimization (ZO) and compared with the proposed DSSO technique. The performance analysis with error vs. convergence time is graphically plotted in Figure 8. The error decreases with the elevation of convergence time as shown in the figure. When the convergence time is 1 the error rate of the proposed and ZO is almost equal to 0, showing the robustness of the stated optimization algorithm. Meanwhile, the other algorithm such as ACO attains an error of 0.2, GA attains 0.1, and BWO attains an error of 1 when the convergence time is equal to 1. This ensures the searchability and effectiveness of the various algorithms. From figure 13, we got that the performance of proposed work with respect to specificity, sensitivity and accuracy rates are superior than that of existing DBN-GSA, ICSO, GA-SVM and GPSO approaches.

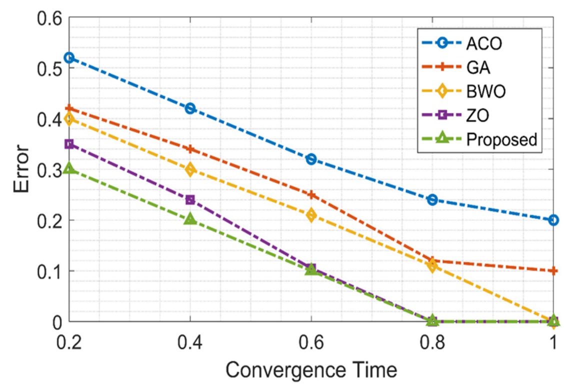


Fig 12: Performance analysis based on the Error vs. Convergence time

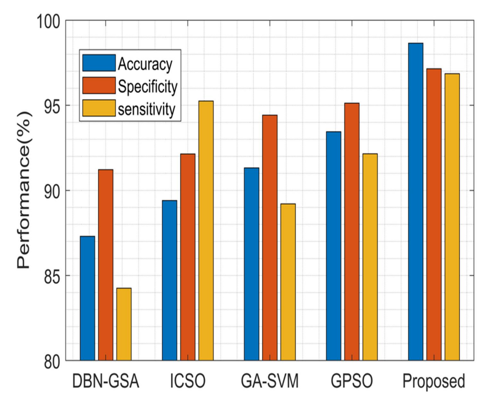


Fig 13: Performance analysis based on the performance percentage

4.3 Difference from Prior Work and Research Novelty

Previous facial emotion recognition methods have primarily focused on improving classification performance through optimized feature selection or deep learning architectures. For instance, DBN-GSA improves feature optimization but exhibits limited cross-dataset generalization. ICSO enhances feature selection while requiring large-scale training data. GPSO improves network optimization; however, its computational complexity restricts real-time deployment. GA-SVM achieves high recognition accuracy but struggles to effectively model complex facial variations. Similarly, CNN-based approaches such as FS-CNN, FERC, EmNet, and DR improve feature learning but remain sensitive to illumination changes, pose variations, occlusion, and local optimum problems during parameter optimization.

Unlike these existing approaches, the proposed framework introduces a Dynamic Social Ski-Driver Optimization (DSSO) algorithm integrated with Deep CapsuleNet for simultaneous parameter optimization and emotion classification. The DSSO algorithm enhances global exploration and exploitation using opposition-based learning, thereby avoiding premature convergence and improving optimal parameter selection for CapsuleNet. Furthermore, the CapsuleNet architecture effectively preserves spatial relationships among facial features, making the proposed model more robust against pose variation and facial deformation compared with conventional CNN-based methods.

Experimental evaluation on the Kaggle FER dataset demonstrates the effectiveness of the proposed framework. The proposed DSSO-Deep CapsuleNet achieves approximately 98.62% classification accuracy, 96.87% sensitivity, 97.12% specificity, and a lower execution time compared with existing optimization-based approaches including DBN-GSA, ICSO, GA-SVM, GPSO, ACO, BWO, and Zebra Optimization. These improvements confirm that the proposed framework not only enhances recognition accuracy but also provides faster convergence and better optimization capability.

Therefore, the major novelty of this work lies in the integration of DSSO-based parameter optimization with Deep CapsuleNet for facial emotion recognition, enabling improved feature representation, optimized network learning, higher classification performance, and superior convergence characteristics over existing state-of-the-art optimization-assisted emotion recognition methods.

Potential Applications

The proposed DSSO-Deep CapsuleNet framework has significant potential for real-world facial emotion recognition applications. It can be employed in intelligent human-computer interaction systems, healthcare monitoring for mental and emotional well-being assessment, driver fatigue and stress detection, online education platforms for student engagement analysis, surveillance systems for abnormal behavior identification, customer sentiment analysis in smart retail environments, and social robotics. Owing to its high recognition accuracy and efficient convergence, the proposed model is suitable for real-time emotion-aware systems where reliable facial expression analysis is essential.

4.4 Open Research Issues, Limitations and Computing Research Contribution

Although the proposed DSSO-Deep CapsuleNet framework achieves superior emotion recognition performance, several research challenges remain. Existing methods and the proposed model are primarily evaluated on benchmark datasets, limiting their generalization to real-world environments with varying illumination, occlusion, and head pose. Furthermore, the proposed work has been validated only on the Kaggle FER dataset and considers static facial images. Future research can focus on cross-dataset validation, real-time video emotion recognition, lightweight edge deployment, and explainable AI techniques. The main computing contribution of this work is the integration of DSSO with Deep CapsuleNet, which improves parameter optimization, preserves facial feature relationships, and enhances emotion classification accuracy compared with existing methods.

Results Interpretation

The effectiveness of the proposed DSSO-Deep CapsuleNet framework is interpreted using standard facial emotion recognition performance metrics, including classification accuracy, sensitivity, specificity, execution time, and convergence behavior. Higher values of accuracy, sensitivity, and specificity together with lower execution time indicate superior classification capability and computational efficiency. Compared with previous optimization-based approaches such as DBN-GSA, ICSO, GA-SVM, and GPSO, the proposed framework demonstrates improved performance across all evaluation metrics. These findings are consistent with previous studies that reported the benefits of optimization-assisted deep learning while further demonstrating that the integration of DSSO with Deep CapsuleNet provides enhanced parameter optimization, faster convergence, and improved recognition performance.

Table 1. Comparative Analysis of the Proposed DSSO-Deep CapsuleNet Framework with Existing Facial Emotion Recognition Methods

Method	**Main Limitation**	**Proposed DSSO-Deep CapsuleNet Improvement**
DBN-GSA	Limited cross-dataset generalization	Improved optimization and higher classification accuracy
ICSO	Requires large training dataset	Better parameter optimization with improved convergence

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

GA-SVM	Limited handling of complex facial variations	Better spatial feature learning using CapsuleNet
GPSO	Higher computational complexity	Faster convergence and improved optimization efficiency
CNN-based methods (FS-CNN, FERC, EmNet, DR)	Sensitive to pose variation, illumination and occlusion	DSSO optimizes CapsuleNet parameters while preserving facial spatial relationships
Proposed DSSO-Deep CapsuleNet	—	**98.62% Accuracy, 96.87% Sensitivity, 97.12% Specificity with improved convergence**

Table 1 presents a comparative overview of the limitations of existing facial emotion recognition methods and highlights the key improvements and research contributions achieved by the proposed DSSO-Deep CapsuleNet framework.

5. CONCLUSION:

The objective of this research was to address the limitations of existing facial emotion recognition methods, particularly their inability to jointly optimize feature extraction and emotion classification under challenging conditions such as pose variation, illumination changes, and facial occlusion. The proposed DSSO-Deep CapsuleNet framework successfully addresses these challenges by integrating Dynamic Social Ski-Driver Optimization with Deep CapsuleNet for efficient parameter optimization and robust feature learning. Experimental evaluation on the Kaggle FER dataset demonstrated superior classification accuracy (98.62%), sensitivity (96.87%), specificity (97.12%), and faster convergence compared with existing optimization-based methods. These findings confirm that the proposed framework effectively achieves the research objectives and provides an efficient solution for reliable facial emotion recognition.

REFERENCES:

[1] Yang, J., Qian, T., Zhang, F. and Khan, S.U., 2021. Real-time facial expression recognition based on edge computing. *IEEE Access*, *9*, pp.76178-76190.

[2] Selvan, Mercy Paul, and Renuka Selvaraj. "Mining user Message Pattern for Suspicious Behavior on Terrorism using NLP in Social Networks with Single Sign-On." Indian Journal of Science and Technology 10 (2017): 14.

[3] Hao, M., Chou, C.H., Zhang, J., Yang, F., Cao, C., Yin, P., Liang, W., Niu, C.M. and Lan, N., 2020. Restoring finger-specific sensory feedback for transradial amputees via non-invasive evoked tactile sensation. *IEEE Open Journal of Engineering in Medicine and Biology*, *1*, pp.98-107.

[4] Deng, J. and Ren, F., 2020. Multi-label emotion detection via emotion-specified feature extraction and emotion correlation learning. *IEEE Transactions on Affective Computing*.

[5] Selvan, Mercy Paul, Robert Bagio, Rithesh, and Viji Amutha Mary. "Computer Vision-Based Approach for Indian Sign Language Character Recognition Using CNN and ROI Segmentation." In Advances in Electronics, Communication and Computing: Select Proceedings of ETAEERE 2020, pp. 421-431. Springer Singapore, 2021..

[6] Jayaraman, U., Gupta, P., Gupta, S., Arora, G. and Tiwari, K., 2020. Recent development in face recognition. *Neurocomputing*, *408*, pp.231-245.

[7] Otsuka, K. and Tsumori, M., 2020. Analyzing multifunctionality of head movements in face-to-face conversations using deep convolutional neural networks. *IEEE Access*, *8*, pp.217169-217195.

[8] Guo, H., Zhang, S., Soong, F.K., He, L. and Xie, L., 2021, January. Conversational end-to-end tts for voice agents. In *2021 IEEE Spoken Language Technology Workshop (SLT)* (pp. 403-409). IEEE.

[9] Alenazy, W.M. and Alqahtani, A.S., 2021. Gravitational search algorithm based optimized deep learning model with diverse set of features for facial expression recognition. *Journal of ambient intelligence and humanized computing*, *12*, pp.1631-1646.

[10] Sikkandar, H. and Thiyagarajan, R., 2021. Deep learning based facial expression recognition using improved Cat Swarm Optimization. *Journal of Ambient Intelligence and Humanized Computing*, *12*, pp.3037-3053.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

[11] Gao, Z., Li, Y., Yang, Y., Wang, X., Dong, N. and Chiang, H.D., 2020. A GPSO-optimized convolutional neural networks for EEG-based emotion recognition. *Neurocomputing*, *380*, pp.225-235.

[12] Liu, X., Cheng, X. and Lee, K., 2020. GA-SVM-based facial emotion recognition using facial geometric features. *IEEE Sensors Journal*, *21*(10), pp.11532-11542.

[13] Said, Y. and Barr, M., 2021. Human emotion recognition based on facial expressions via deep learning on high-resolution images. *Multimedia Tools and Applications*, *80*(16), pp.25241-25253.

[14] Mehendale, N., 2020. Facial emotion recognition using convolutional neural networks (FERC). *SN Applied Sciences*, *2*(3), p.446.

[15] Saurav, S., Saini, R. and Singh, S., 2021. EmNet: a deep integrated convolutional neural network for facial emotion recognition in the wild. *Applied Intelligence*, *51*, pp.5543-5570.

[16] Ch, S., 2021. An efficient facial emotion recognition system using novel deep learning neural network-regression activation classifier. *Multimedia Tools and Applications*, *80*(12), pp.17543-17568.

[17] Tang, M. and Feng, J., 2015. Multi-kernel correlation filter for visual tracking. In *Proceedings of the IEEE international conference on computer vision* (pp. 3038-3046).

[18] Wahab, M.F., Gritti, F. and O'Haver, T.C., 2021. Discrete Fourier transform techniques for noise reduction and digital enhancement of analytical signals. *TrAC Trends in Analytical Chemistry*, *143*, p.116354.

[19] Murugappan, M. and Mutawa, A., 2021. Facial geometric feature extraction based emotional expression classification using machine learning algorithms. *Plos one*, *16*(2), p.e0247131.

[20] Chatterjee, B., Bhattacharyya, T., Ghosh, K.K., Singh, P.K., Geem, Z.W. and Sarkar, R., 2020. Late acceptance hill climbing based social ski driver algorithm for feature selection. IEEE Access, 8, pp.75393-75408.

[21] Li, X., Kiringa, I., Yeap, T., Zhu, X. and Li, Y., 2020. Exploring deep anomaly detection methods based on capsule net. In Advances in Artificial Intelligence: 33rd Canadian Conference on Artificial Intelligence, Canadian AI 2020, Ottawa, ON, Canada, May 13–15, 2020, Proceedings 33 (pp. 375-387). Springer International Publishing.

[22] https://www.kaggle.com/datasets/ananthu017/emotion-detection-fer?resource=download.