Journal of Information Systems Engineering and Management
2025, 10(51s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Enhanced facial emotion detection based on deep feature whale optimization algorithm with Hyper Capsule Generative Adversarial Network

S. Sahaya Sugirtha Cindrella[1], R. Jayashree[2*]

[1]Department of Computer Applications, Faculty of Science and Humanities, SRM Institute of Science and Technology, Kattankulathur , Chengalpattu, Tamilnadu-603202, India

[2*]Department of Computer Applications, Faculty of Science and Humanities, SRM Institute of Science and Technology, Kattankulathur , Chengalpattu, Tamilnadu - 603202, India

Email: sc1905@srmist.edu.in, jayashrr@srmist.edu.in

ARTICLE INFO

Received: 18 Dec 2024

Revised: 20 Feb 2025

Accepted: 28 Feb 2025

ABSTRACT

Introduction: The emotion-based facial expression analysis combines both rule-based and data-driven approaches to detect and interpret human emotions from facial features. This method enhances emotion recognition in real-time applications like human-computer interaction and surveillance. Traditional facial expression systems rely heavily on handcrafted features, which limits flexibility. They are often sensitive to environmental noise and fail in dynamic or spontaneous expressions.

Objectives: The proposed system presents a hybrid approach for emotion-based facial expression recognition, integrating advanced techniques at multiple stages of the processing pipeline.

Methods: Initially, Discrete Wavelet Transform (DWT) is employed for pre-processing, enhancing feature extraction by decomposing facial images into multiple frequency components. This is followed by Facial Landmarks-Based Segmentation, which isolates critical facial regions that are most indicative of emotional expression. For feature selection, the Whale Optimization Algorithm (WOA) is utilized to identify the most relevant features, thereby reducing dimensionality and enhancing the model's efficiency. Classification is performed using a novel Hyper Capsule Generative Adversarial Network (HCGAN-G), which combines the representational strength of capsule networks with the generative capabilities of GANS to improve recognition performance, especially in complex and subtle emotional states.

Results: The effectiveness of the system is rigorously evaluated using a comprehensive set of performance metrics, including accuracy 95.4%, sensitivity 94.2%, specificity 96.3%, precision 93.5%, root mean square error 0.25, area under the curve 0.97, and F1-score 0.89, demonstrating its robustness and reliability in emotion recognition tasks.

Conclusions: The proposed hybrid emotion-based facial expression recognition system significantly improves on traditional methods by combining rule-based and data-driven approaches. This architecture integrates the hierarchical spatial feature learning capabilities of capsule networks with the generative abilities of GANs. As a result, the system can accurately classify even subtle and complex emotional expressions.

Keywords: Discrete Wavelet Transform, Hyper Capsule Generative Adversarial Network (HCGAN-G), Facial Landmarks-Based Segmentation, Whale Optimization Algorithm.

INTRODUCTION

Recent advancements in deep learning have fostered the development of various innovative models for emotion-based facial expression recognition, particularly in medical and affective computing domains. For instance, novel [1] While the model demonstrated potential for clinical application, its effectiveness was constrained by the limited volume and quality of existing emotional expression datasets. To address challenges in facial recognition, [2] introduced the Frequency Neural Network, which utilized frequency-based rather than spatial analysis. This method improved accuracy in diverse facial conditions but suffered from high sensitivity to image disturbances and required intensive image preprocessing.

In an effort to enhance recognition accuracy in non-frontal views, [3] developed a Pose-Guided Face Alignment model with Discriminative Features. Although this approach improved pose-invariant expression recognition, its performance was hindered by reliance on precise pose estimation, which remains difficult in dynamic environments. To overcome the limitations of handcrafted features, [4] Despite its success in feature learning, the model's performance diminished in low-light settings and with occluded faces due to inadequate data augmentation strategies. Lastly, [5] employed Deep Metric Learning with Synthetic Images (DML-SI) to reduce identity bias in facial emotion recognition. However, the synthetic data introduced a disconnect from real-world domain characteristics, limiting its practical effectiveness in reducing misclassification rates.

OBJECTIVES

To develop a robust and accurate hybrid system for emotion-based facial expression recognition by integrating both rule-based and data-driven techniques.

To enhance feature extraction and classification accuracy through advanced preprocessing, segmentation, feature selection, and deep learning methods.

To evaluate the proposed system using comprehensive performance metrics to ensure its effectiveness in real-world applications such as human-computer interaction and surveillance.

CONTRIBUTION OF WORK

Hybrid Approach Design Introduced a novel hybrid architecture that combines Discrete Wavelet Transform (DWT), Facial Landmarks-Based Segmentation, Whale Optimization Algorithm (WOA), and a custom deep learning model for effective emotion recognition.

Efficient Preprocessing and Segmentation Employed DWT to improve feature extraction by capturing key frequency components, and utilized facial landmark detection to isolate regions most relevant to emotional expression.

Optimized Feature Selection Implemented WOA to select the most significant features, thereby minimizing redundancy, reducing dimensionality, and enhancing model efficiency.

Advanced Classification Model Developed the Hyper Capsule Generative Adversarial Network (HCGAN-G), integrating the spatial awareness of capsule networks and the generative strength of GANs, enabling superior recognition of subtle and complex emotions.

The remaining portion of the document is divided into significant sections, which are described as follows: Section II examines the current research efforts in Hybrid Emotion Based Facial Expression used by different authors. The workflow of the suggested approach is explained in Section III and consists of feature selection, Segmentation, pre-processing, and classification models. Section IV presents the findings analysis and performance data. Section V presents the conclusion.

LITERATURE SURVEY

The authors in [6] developed Identity-Aware Contrastive Knowledge Distillation (IA-CKD) for training systems to excel at attribute recognition while enhancing generalization capabilities. The deployment of these models became complicated because their heightened complexity and computational requirements made them challenging to implement.

The authors in [7] demonstrated Joint DL for Facial Expression Synthesis and Recognition (JDL-FESR) to achieve combined expression recognition and synthesis through a unified network. The added dual functionality reduced learning time, but the process demanded performance compromises between generation and recognition abilities. The authors of [8] implemented Zoning-based DL (ZDL) for localized feature extraction through facial region division. The method showed better local pattern detection, but it failed to integrate global features, which made holistic expression recognition inefficient.

The author in [9] presented Subtle Facial Action Recognition (SFAR) for detecting driver yawning through the analysis of minor facial movements to maintain driver alertness. The research's specific detection capabilities were accurate for yawning, but they rendered the system inadequate for extensive face recognition work. In [10], an Equilibrium Optimizer and a Hybrid DL Model (EO-HDL) were designed to achieve superior emotion recognition in challenging environments. The excessive computational requirements of the optimization solution demanded by the system made real-time usage impossible.

Table 1. Survey of various DL methods based on Facial Detection

Author/Year	Used Methodology	Dataset	Performance Metrics	Drawbacks
C. Liu et al., (2021)	Spatial Attention Convolutional Neural Network and Long Short-term Memory networks with Attention mechanism	FER2013, CK+, JAFFE	Accuracy = 95.2%	The algorithm achieves reduced performance when operating in environments not managed by humans because of both obstruction and unforeseen changes in lighting conditions.
N. Yu et al., (2020)	Partial Image and Deep Metric Learning Method	CK+, Oulu-CASIA, MMI	Accuracy = 94.71%	Limited generalization due to overfitting on partial features.
U. Nawaz et al., (2025)	Transformer approach	FER2013, CK+, AffectNet, RAF-DB, and AFEW	Accuracy = 92.83%	The system operates at high processing costs while experiencing delays on real-time equipment.
S. Hossain et al., (2024)	Deep Quantum CNN (DQCNN)	Karolinska-directed emotional faces, FER-2013	Accuracy = 91.2%	Notable hardware limitations prevent quantum models from scaling up, as they require advanced simulations.
B. Fang et al., (2023)	Silhouette coefficient-based contrast clustering algorithm	RAF-DB, FERPlus and AffectNet	Accuracy = 89.65%	The proposed method produces noise that affects datasets with high degrees of imbalance.
V. S. Bhati et al., (2025)	Generalized Zero-Shot CNN	FER2013, AffectNet, RAF-DB, CK+,	Accuracy = 88.34%	The GZS-ConvNet encounters difficulties when processing unknown class variations

		KDEF, and JAFEE		along with unbalanced class distribution.
J. Wang et al., (2022)	Resnet	Casia Webb-Face	Accuracy = 93.56%	The system specializes for genetic syndromes but does not perform well for general FER applications.
R. Wadhawan et al., (2023)	Part-based ensemble transfer learning	CK+ and JAFFE	Accuracy = 96.18%	Scalability remains limited because the proposed method demands time-consuming landmark annotations.
C. Shi et al., (2021)	Multiple Branch Cross-Connected CNN (MBCC-CNN)	Fer2013, CK+, FER+ and RAF	Accuracy = 95.43%	Advanced design choices make the system hard to calibrate because adjusting parameters becomes challenging.
H. Zhang et al., (2020)	Deep Neural Network (DNN)	DEAP	Accuracy = 86.42%	The implementation of EEG data needs specialized hardware components in addition to user compliance for integration protocols.

The author [21] introduced a multimodal analysis that fused visible images with Infrared and Multispectral images through Early and Late Fusion of DL Models. Expressive data capture was enhanced under diverse situations using this technique yet scalability problems arose because it required more powerful computations alongside specialized hardware for multispectral information acquisition.

The authors in [22] developed Learning Transferable Sparse Representations to address domain adaptation problems in cross-corpus scenarios by training emotion features that are transferable across multiple datasets. Although it decreased dataset bias, the approach failed to be consistent because it reacted to both noise and changing data distributions in real-world applications. The model Phase Space Reconstruction Driven Spatio-Temporal Feature Learning by [23] derived spatial as well as temporal patterns from facial video sequences to achieve improved detection of evolving expressions. With its ability to process video sequences, the model relied on large training datasets and underwent extended training periods due to its sophisticated feature processing system.

In [24] established Cross-Dataset Adaptation, which enabled unbiased FER in the wild through improved generalization between different datasets. This assessment method enhanced real-world tasks, but it needed substantial pre-training quality and large-scale domain matching between source and target datasets, while remaining difficult to deploy. The author developed LQGDNet, which integrates features to recognize depression-related facial expressions through a combination of local and global features [25]. LQGDNet achieved excellent results in emotion-specific assessments for depression recognition. Yet, its performance remained restricted to depression-based tasks since it did not work across a broad spectrum of facial emotional displays.

PROPOSED METHODOLOGY

In order to efficiently analyze and understand human emotions in real-time applications including surveillance, smart environments, and human-computer interaction, the suggested emotion-based facial expression recognition system combines rule-based and data-driven methodologies. The system uses sophisticated signal processing, feature selection, and classification algorithms to get beyond the drawbacks of conventional approaches.

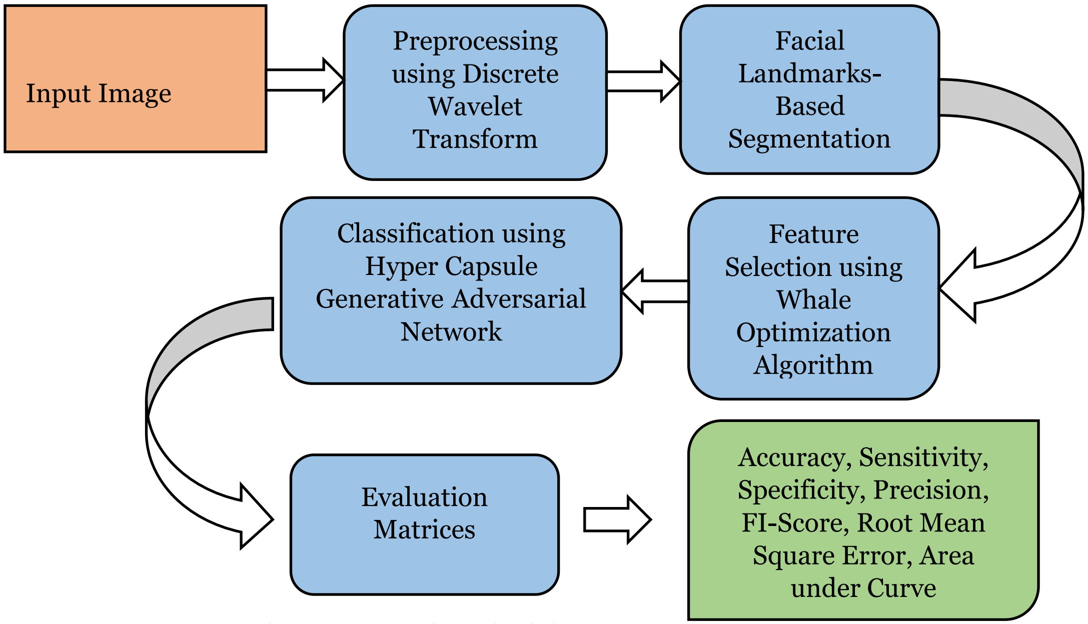
Input Image
Preprocessing using Discrete Wavelet Transform
Facial Landmarks-Based Segmentation
Feature Selection using Whale Optimization Algorithm
Classification using Hyper Capsule Generative Adversarial Network
Evaluation Matrices
Accuracy, Sensitivity, Specificity, Precision, FI-Score, Root Mean Square Error, Area under Curve

Fig.1. Proposed Methodology Diagram

The diagram in Figure 1 illustrates the proposed methodology for an emotion-based facial expression recognition system, which integrates both rule-based and data-driven techniques to enhance recognition performance in real-time environments. The process begins with the input facial image, which undergoes preprocessing using the Discrete Wavelet Transform. This step improves the quality of the image and enhances feature extraction by decomposing the image into various frequency sub bands. The segmented facial regions are then passed to the feature selection module, where the Whale Optimization Algorithm identifies and retains the most relevant features. This step effectively reduces dimensionality and improves computational efficiency without compromising recognition accuracy. The selected features are then fed into the classification module, which uses a Hyper Capsule Generative Adversarial Network.

PRE-PROCESSING: DISCRETE WAVELET TRANSFORM

The DWT is employed as a robust pre-processing technique for hybrid emotion-based facial expression recognition. DWT decomposes the input facial image into multiple frequency sub-bands, enabling the separation of fine and coarse features such as contours, textures, and subtle muscle movements that are crucial for detecting complex emotions. This method effectively captures both spatial and frequency-domain information, enhancing the representation of facial expressions while reducing noise and irrelevant details. By analyzing horizontal, vertical, and diagonal components at different resolution levels, DWT facilitates the extraction of prominent facial landmarks, such as the eyebrows, eyes, nose, and mouth areas, used for emotional feature analysis.

The equation for the DWT is given by:

$$F(x,y) = \sum\sum g(r,s) * e^\wedge(-j2\pi(\frac{xr}{M} + \frac{ys}{N}) \qquad (1)$$

Here, F(x, y) refers to the wavelet-transformed coefficients that represent localized frequency content at various scales and orientations, while g(r, s) denotes the original facial image in the spatial domain. The indices (x, y) correspond to spatial positions in the transformed sub-bands, and M and N are the dimensions of the image. Unlike the global nature of the DWT offers spatial and frequency localization, enabling the model to isolate and analyze subtle facial movements such as eyebrow twitches or mouth contour changes that signify complex or hybrid emotions.

This multi-resolution analysis supports effective facial by isolating emotionally relevant details such as furrowed brows, eye wrinkles, or lip curvature. Therefore, the application of DWT in this framework serves as a refined cross-domain approach, mapping spatial facial patterns into frequency sub-bands for more accurate and robust hybrid emotion classification, as outlined in Equation (2) and (3).

$$Q(e,f) = \sum_{a=1}^{M-1} \sum_{b=0}^{N-1} h(a,b)\, e\hat{}(-y2\pi(\frac{xa}{M} + \frac{yb}{M}) \tag{2}$$

In this framework, the spatial-domain image h (a, b) is transformed using wavelet basis functions rather than the global exponential terms used in the Fourier domain. While the exponential term in the Fourier Transform serves as a global basis for computing each frequency component Q (e, f), the DWT applies localized wavelet filters that capture both spatial and frequency information. Two fundamental properties of wavelets scale and shift enable the Discrete Wavelet Transform (DWT) to analyze facial features at multiple resolutions and positions. The wavelet coefficients $C_{j,k}$and the associated scaling function $\emptyset_{j,k}(t)$ can be computed to represent localized changes in facial structure. These coefficients are crucial for identifying emotional features such as furrowing of the brow, eye squinting, or asymmetric lip movements.

$$\emptyset_{j,k}(t) = 2^{\frac{j}{2}} \emptyset(2^j t - k) \tag{3}$$

In hybrid emotion-based facial expression recognition, the mother scaling function, denoted as $\emptyset_{j,k}$, plays a foundational role in the Discrete Wavelet Transform (DWT). This function governs the scaling behavior of the wavelet and serves as the basis for multi-resolution analysis of facial features. The dilatation parameter j controls the scale of the wave, allowing the system to zoom in on fine details or capture broader structural features across the face.

FACIAL LANDMARKS-BASED SEGMENTATION

After the pre-processing stage, the filter images segmented using the Facial Landmarks-Based Segmentation

Although typical pre-smoothing filters are effective in reducing noise and irregular features, they often lead to a loss of contour edge information. As a result, the region contour may shift positional in the reconstructed image, altering the original structure. This change occurs even though the smoothed image may appear visually appealing. Morphological reconstruction is a technique used to restore the shape and connectivity of regions in an image based on a marker and mask approach, helping to preserve meaningful structural details.

$$\emptyset_{j,k}(t) + 1 = (h_k + se) \cap f \tag{4}$$

Where: $h_k^{se} ere\ (g,f)$ denote the result of the morphological reconstruction process, where g, where f is the original image as a mask and se is the structural element that makes the marker image swell; h_k is the final iteration's outcome image, while h is the marker image g's first iteration. The final iteration of equation (4) occurs when $\emptyset_{j,k}(t) + 1$ Morphological closed reconstruction, like morphological open reconstruction, restores the target edge completely while excluding texture features smaller than structural elements.

The process continues until the reconstruction converges, meaning no further changes occur in the image across iterations. Additionally, it maintains the reconstructed image's outlines the definition of morphological restoration is:

$$M_{n+1} = (M_n \oplus StEl) \cap x, \tag{5}$$

Where is the result of the n-th iteration. $StEl$ Is the strutting element. X is the mask or constraint image. Additionally, initial and final restorative actions in morphological reconstruction are defined by Equation (6)

$$G_{StEl}^{uv} = I_{StEl}^{uv} \lfloor F_{StEl}^{uv}(x), x \rfloor \tag{6}$$

Journal of Information Systems Engineering and Management
2025, 10(51s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/
Research Article

Where I^{uv}_{StEl}(x) is the initial action restoration and $F^{uv}_{StEl}(x)$ is the final action restoration. By combining these morphological operations with facial landmarks-based segmentation, the method ensures precise localization of facial features while minimizing noise and preserving structural boundaries essential for accurate emotion recognition.

FEATURE SELECTION: WHALE OPTIMIZATION ALGORITHM (WOA)

This method allows for more accurate interpretation by analyzing facial expressions using elements that reflect basic and complex emotions. To recognize complicated emotional states, facial landmarks and emotional signals are recorded and processed in certain ways. After more investigation, researchers have shown that this behavior may be modelled to improve recognition systems using multi-modal feature extraction, emotion fusion, and adaptive learning across various facial datasets.

The currently most accurate or representative emotional expression is considered the reference or target expression. Each facial instance adjusts its representation by aligning toward this target expression, while other instances in the dataset adapt toward the optimal feature configuration. This alignment and feature enhancement process is achieved by updating their feature vectors using Equation (7) and (8)

$$G = |F.X*(S) - X(S)| \tag{7}$$

$$X(S+1) = X*(S) - A.G \tag{8}$$

Where, S is the current iteration number; X(s) is the current feature vector of the facial expression instance; X* represents the feature vector of the target or reference emotional expression. The coefficient vectors A and F are defined as follows:

$$A = 2b.r_1 - b \tag{9}$$

$$C = 2.r_2 \tag{10}$$

Where an is the convergence factor, which progressively drops from 2 to 0 as the number

Of iterations rises; r_1 and r_2 are random values in the interval [0,1]

$$b = 2 - \frac{2s}{s_{max}} \tag{11}$$

Where **S** is the current iteration number and **S** is the maximum number of iterations. In the context of hybrid emotion-based facial expression recognition, two distinct strategies are employed: the Feature Refinement Mechanism and the Spiral Expression Alignment method. Its mathematical formulation models this non-linear, spiral-like adjustment process as follows:

$$Y(s+1)D.e^{bl}.\cos(2\pi l) + X*(s) \tag{12}$$

Where D = |Y*(t) − X (t)| represents the distance between the current facial expression feature vector and the target (reference) hybrid emotion vector. Here, b is a constant used to define the logarithmic spiral shape, and l is a random number in the range [−1, 1]. The mathematical model for this random feature exploration can be represented as follows:

$$E = |C.Xrand(s) - X(s)| \tag{13}$$

$$X(s+1) = Xrand(s) - A.D \tag{14}$$

Where $Xrand(s)$ represents the feature vector of a randomly selected facial expression instance from the dataset. This mechanism introduces variability by comparing and adapting features based on randomly chosen expressions, thereby enhancing the diversity and generalization capability of the hybrid emotion recognition process.

CLASSIFICATION: A HYPER CAPSULE GENERATIVE ADVERSARIAL NETWORK GAN (HCGANG)

The Hyper Capsule Generative Adversarial Network is a hybrid deep learning framework designed for accurate classification of complex and blended facial emotions. These capsules are capable of capturing part-whole relationships, making them ideal for recognizing subtle emotional cues. In parallel, the generative adversarial component introduces a discriminative-learning phase where a generator produces emotion-rich facial representations while a discriminator refines the classification boundaries through adversarial training.

In the Hyper Capsule Generative Adversarial Network model for hybrid emotion-based facial expression classification, the generator is designed to minimize a combination of data-fidelity and adversarial loss functions. The data-fidelity term ensures that the generated facial expression features closely resemble the real emotional data in terms of semantic and spatial accuracy. This alignment helps in preserving crucial facial cues necessary for distinguishing between subtle and compound emotions. The data-fidelity objective is formally represented by Equation (15)

$$L_{data} = X(s+1)\sum_{i=1}^{N}||G(x_i) - y_i||1 \qquad (15)$$

Where L_{data}represents the data-fidelity loss function used in the HCGANG model. The summation symbol $\sum$ denotes the aggregation over all instances from 1 to N, where N is the total number of facial expression samples in the dataset. $G(x_i)$ Denotes the output of the generative component of the HCGANG, which attempts to reconstruct or generate the emotional representation corresponding to the input facial image (x_i) the variable y_i represents the ground truth emotional label or the true expression features associated with(x_i). The term $||G(x_i) - y_i||1$is the L1 norm, which calculates the absolute difference between the generated and true emotional data.

It ensures that the reconstructed hybrid emotion representations preserve the subtle characteristics of true facial expressions. Meanwhile, the adversarial loss component, which encourages the generator to produce emotionally realistic and indistinguishable outputs, is formulated by Equation (16) as:

$$L_{adv} = -\sum_{i=1}^{N}\log\big(G((x_i))\big) \qquad (16)$$

Where $L_{adv}$is the adversarial loss used to train the generator within the HCGANH model. This loss encourages the generator to produce hybrid emotion facial expressions that are indistinguishable from real ones. Here, N represents the number of facial expression samples in the dataset being summed over. The function G denotes the generator, which maps each input facial image $x_i$to a synthesized hybrid emotion representation.

RESULT & DISCUSSION

The result Hyper Capsule Generative Adversarial Network (HCGANG) classification framework presents a robust solution for addressing the complexities involved in hybrid emotion-based facial expression recognition. By leveraging Discrete Wavelet Transform (DWT) as a preprocessing step, the system effectively enhances the fine-grained details of facial images, allowing for more precise texture and edge preservation. This segmentation ensures that only the most relevant parts of the face contribute to the classification process. To further refine the model, the Whale Optimization Algorithm (WOA) is utilized for optimal feature selection, reducing dimensionality while retaining the most discriminative features. The integration of these components within the HCGANG architecture allows for superior learning and generation capabilities, particularly in capturing subtle emotional cues. The system's efficacy is validated through comprehensive performance metrics, including accuracy, sensitivity, and specificity, Precision, Root Mean Square (RMSE), Area under the Curve (AUC), and F1-score.

DATASET DESCRIPTION

Several emotion-based facial expression datasets are commonly used in research and machine learning. These include FER2013, AffectNet, and the Extended Cohn-Kanade Dataset (CK+), among others. These datasets provide a range of facial images labeled with specific emotions, enabling the development and evaluation of emotion recognition models. Figure 2 Presents of fear expressions for the two settings

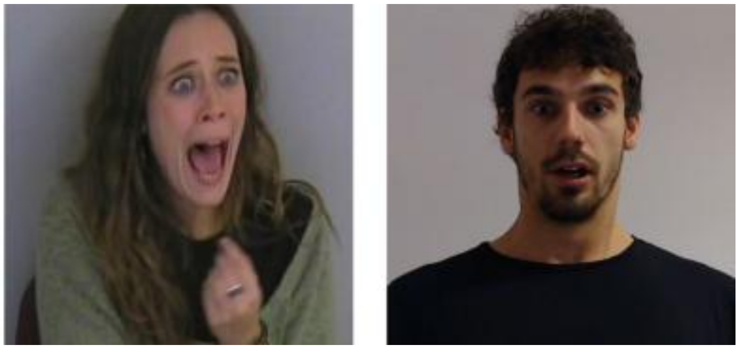

Fig. 2. Examples of fear expressions for the two settings

PERFORMANCE METRICS

Accuracy

Accuracy is achieved when the computational model correctly identifies the emotional states as intended, based on the percentage of true positive outcomes within the selected population. In this study, accuracy was realized when the algorithm's classification of facial expressions matched the emotional labels determined by human experts.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{17}$$

Sensitivity

It is particularly important in emotion recognition systems, where failing to detect subtle or complex emotions can significantly impact the accuracy and effectiveness of affective computing applications.

$$Sensitivity = \frac{TP}{TP+FN} \tag{18}$$

Specificity

High specificity is crucial in emotion recognition to prevent false positives, such as mistakenly classifying a neutral or different emotional expression as a targeted emotion, which could compromise the system's reliability in real-world applications.

$$Specificity = \frac{TN}{TN+FP} \tag{19}$$

Root Mean Squared Error

In the context of hybrid emotion-based facial expression recognition, RMSE can be employed to evaluate the discrepancy between the model's predicted confidence scores and the actual labeled emotions.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - y_i)^2} \tag{20}$$

F1-score

The F1-score is especially valuable when there is an imbalance in the distribution of emotional classes, such as underrepresented or subtle emotions. It provides a balanced measure by combining recall the ability to correctly identify all instances of a specific emotion and precision the accuracy of those predicted emotional instances.

$$FI - Score = 2 \times \frac{\text{Precision}\times\text{Recall}}{\text{Precision}+\text{ Recall}} \tag{21}$$

Area under the Curve (AUC)

The capacity of a classification model to differentiate between several classes in this example, emotional states is measured by the area under the Receiver Operating Characteristic (ROC) Curve. It shows the likelihood that a randomly selected positive instance would be ranked higher by the model than a randomly selected negative case.

Precision

In the context of hybrid emotion-based facial expression recognition, it indicates how accurate the model is when it claims that a particular emotion is present.

$$Precision = \frac{TP}{TP+FP} \tag{22}$$

COMPARISON RESULTS

The proposed Hyper Capsule Generative Adversarial Network (HCGAN) model was evaluated. The system's efficacy is validated through comprehensive performance metrics, including accuracy, sensitivity, and specificity, Precision, Root Mean Square (RMSE), Area under the Curve (AUC), and F1-score.

Table 2. Performance Metrics Comparison

Metrics	**CNN**	**KNN**	**Random Forest**	**SVM**	**HCGANG**
Accuracy	89.2%	88.7%	92.5%	90.1%	95.4%
Sensitivity	87.6%	85.4%	91.3%	88.7%	94.2%
Specificity	90.5%	89.9%	93.8%	91.4%	96.3%
Precision	88.0%	86.8%	91.0%	89.6%	93.5%
RMSE	0.38	0.40	0.32	0.35	0.25
AUC	0.89	0.88	0.93	0.91	0.97
F1-Score	0.85	0.87	0.92	0.87	0.89

Table 2 compares the performance metrics of the HCGANG framework against four traditional methods: CNN, KNN, Random Forest, and SVM. HCGANG consistently outperforms the traditional methods in most metrics, achieving the highest accuracy (95.4%), sensitivity (94.2%), and specificity (96.3%), demonstrating superior classification of emotional expressions. It also leads in precision (93.5%), AUC (0.97), and, highlighting its ability to balance correct positive predictions and minimize errors. Furthermore, HCGANG shows the lowest RMSE (0.25), indicating more accurate predictions compared to the other models.

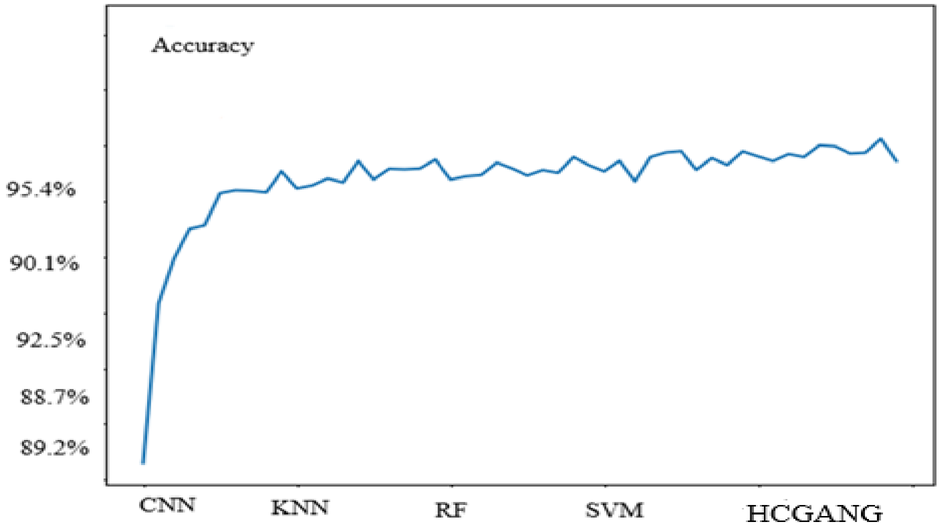

Fig. 3. Evaluation of accuracy for CNN, KNN, RF, SVM and Proposed HCGANG

The accuracy evaluation of many machine learning models, including CNN, KNN, RF, SVM, and the suggested Hybrid Conditional Generative Adversarial Network, is displayed in Figure 1. The accuracy statistic calculates each model's proportion of accurate predictions during testing. With an accuracy of 95.4%, the findings show that the suggested HCGANG model performs better than the other models. While Random Forest attains the greatest accuracy among the conventional models with 92.5%, KNN performs somewhat worse than CNN at 88.7%. SVM comes in second with a 90.1% accuracy rate.

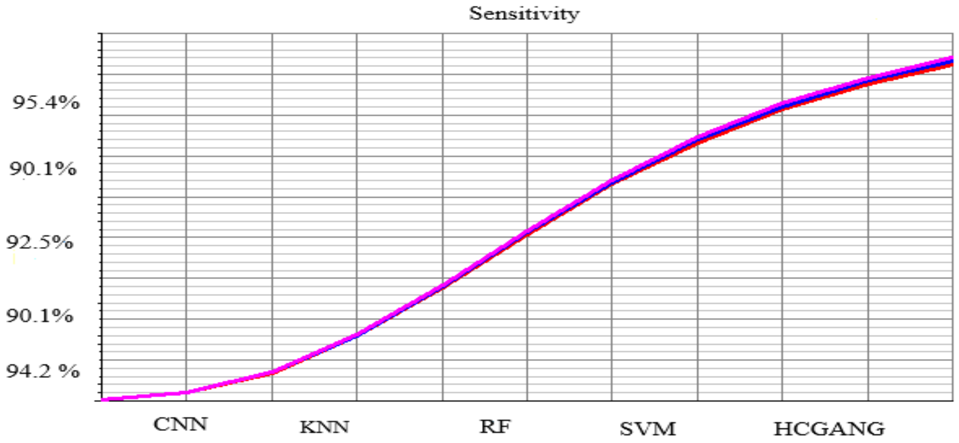

Fig. 4. Evaluation of Sensitivity

Figure 2 illustrates the evaluation of Sensitivity metrics for various machine learning models: CNN, KNN, RF, SVM, and the proposed Hybrid Conditional Generative Adversarial Network. Sensitivity, also known as recall or true positive rate, measures the model's ability to correctly identify positive instances. From the figure, it is evident that the HCGANH model achieves the highest sensitivity score at 94.2%, demonstrating its superior ability to detect true positive cases compared to the other models. CNN follows closely with a sensitivity of 87.6%, showing good performance in identifying positive instances. KNN performs slightly lower with a sensitivity of 85.4%, while Random Forest achieves 91.3%, highlighting its strong ability to correctly classify positives. SVM has a sensitivity of 88.7%, performing reasonably well but still not matching the HCGANG model.

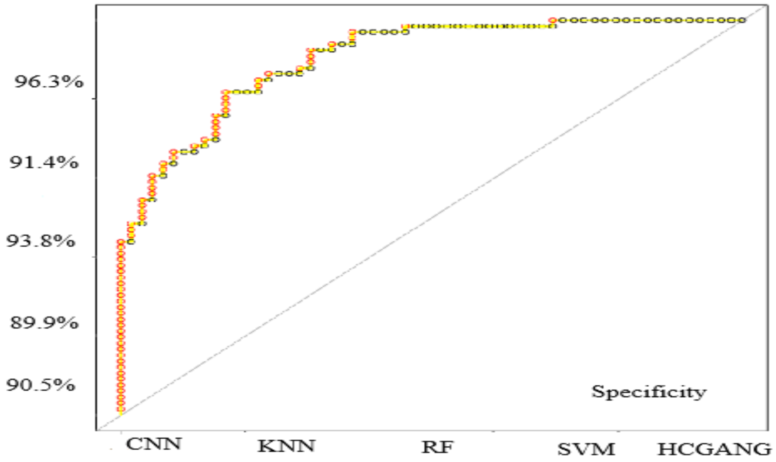

Fig. 5. Evaluation of Specificity

Figure 3 shows the evaluation of specificity measures for CNN, KNN, RF, SVM, and the proposed Hybrid Conditional Generative Adversarial Network, among other machine learning models. With a specificity of 96.3%, the suggested HCGANH model performs exceptionally well in accurately categorizing negative instances, according to the data. When false positives need to be reduced, this is very crucial. Additionally, Random Forest has great performance, with a specificity of 93.8%, followed by SVM at 91.4%. KNN has a little lower specificity of 89.9% than CNN, which reaches 90.5%.

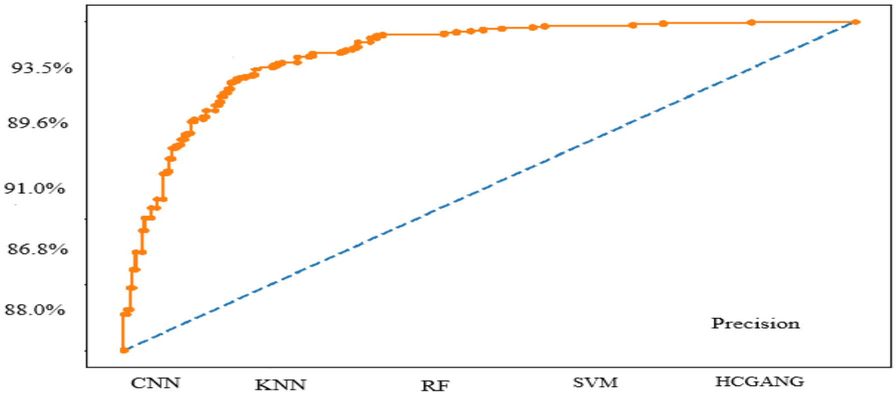

Fig. 6. Evaluation of Precision

Figure 4 presents the evaluation of Precision for various machine learning models: CNN, KNN, RF, SVM, and the proposed Hybrid Conditional Generative Adversarial Network. As shown in the results, the HCGANG model achieves the highest precision at 93.5%, indicating its superior ability to make accurate positive predictions with minimal false alarms. This is a significant advantage in applications where the cost of false positives is high. Random Forest also performs well, achieving a precision of 91.0%, followed by SVM at 89.6%. CNN records a precision of 88.0%, while KNN slightly trails behind with 86.8%.

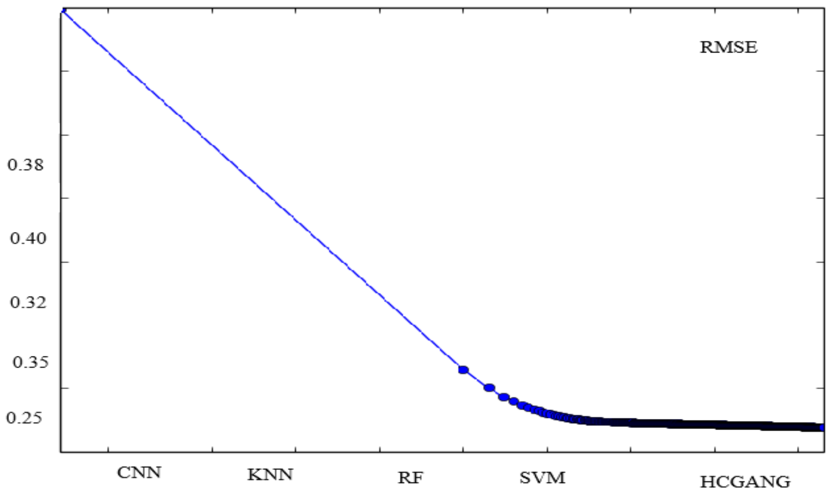

Fig. 7. Evaluation of RMSE

The assessment of the suggested Hybrid Conditional Generative Adversarial Network is displayed in Figure 5. A popular statistic for comparing expected and actual values is root mean square error (RMSE), where lower values suggest higher prediction accuracy. Out of all the models that were assessed, the HCGANH model produced the most accurate predictions, as seen by its lowest RMSE of 0.25. With an RMSE of 0.32, Random Forest performs well but is still less accurate than HCGANH. KNN has the greatest RMSE at 0.40, indicating somewhat less accurate predictions, whereas SVM and CNN record RMSE values of 0.35 and 0.38, respectively.

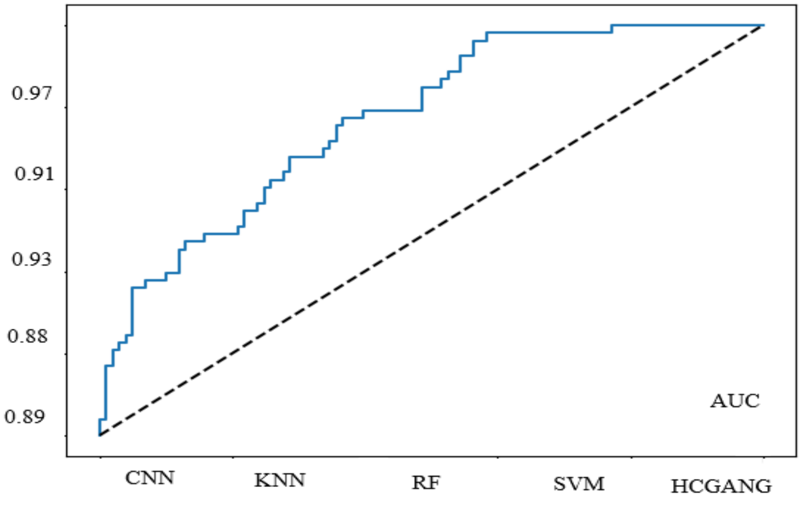

Fig. 8. Evaluation of AUC

AUC is evaluated for CNN, KNN, RF, SVM, and the proposed Hybrid Conditional Generative Adversarial Network, among other machine learning models, as shown in Figure 6. One important metric of overall classification performance is AUC, which quantifies the model's capacity to discriminate across classes. Effective discrimination is indicated by a higher AUC value. Its exceptional capacity to distinguish between positive and negative classes is demonstrated by the findings, which show that the suggested HCGANH model has the greatest AUC score of 0.97.

Random Forest performs well, coming in second with an AUC of 0.93. SVM outperforms KNN (0.88) and CNN (0.89) by a small margin with an AUC of 0.91.

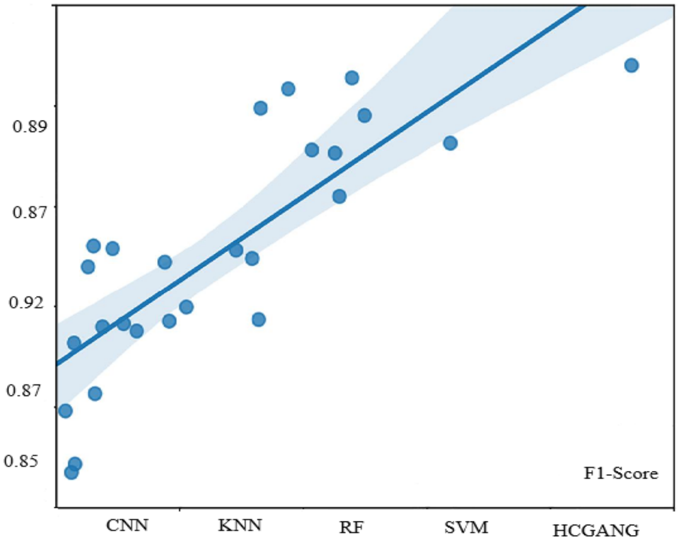

Fig. 9. Evaluation of FI-Score

The F1-Score metric evaluation for five models CNN, KNN, RF, SVM, and the suggested Hybrid Conditional Generative Adversarial Network is shown in Figure 7. The F1-Score, which is the harmonic mean of accuracy and recall (sensitivity), offers a fair evaluation of a model's performance, especially when the data is not balanced or when both erroneous positives and false negatives are costly. According to the results, Random Forest has the greatest F1-Score of 0.92, indicating a strong balance between accuracy and recall. The high F1-Score of 0.89 indicates that the proposed HCGANH model performs consistently and reliably across both criteria. CNN has the lowest F1-Score (0.85), followed by KNN and SVM (0.87).

CONCLUSION

In conclusion, the proposed hybrid emotion-based facial expression recognition system demonstrates a significant advancement over traditional methods by integrating both rule-based and data-driven approaches. By incorporating Discrete Wavelet Transform in the pre-processing phase, the system efficiently captures essential frequency components of facial expressions, thereby enhancing the quality of feature extraction. The use of Facial Landmarks-Based Segmentation ensures that only the most emotion-relevant facial regions are analyzed, which contributes to more accurate and focused recognition. Furthermore, the inclusion of the Whale Optimization Algorithm for feature selection effectively reduces computational complexity and improves the model's performance by prioritizing the most informative features. The classification stage, powered by the novel Hyper Capsule Generative Adversarial Network, is a key innovation. This architecture merges the hierarchical spatial feature learning capability of capsule networks with the generative power of GANs, enabling the system to accurately classify even subtle and complex emotional expressions. The proposed method exhibits outstanding performance, achieving high values across all key evaluation metrics such as 95.4% accuracy, 94.2% sensitivity, and 96.3% specificity, among others.

REFERENCES

[1] W. Huang, W. Xu, R. Wan, P. Zhang, Y. Zha and M. Pang, "Auto Diagnosis of Parkinson's Disease Via a Deep Learning Model Based on Mixed Emotional Facial Expressions," in IEEE Journal of Biomedical and Health Informatics, vol. 28, no. 5, pp. 2547-2557, May 2024, doi: 10.1109/JBHI.2023.3239780.

[2] Y. Tang, X. Zhang, X. Hu, S. Wang and H. Wang, "Facial Expression Recognition Using Frequency Neural Network," in IEEE Transactions on Image Processing, vol. 30, pp. 444-457, 2021, doi: 10.1109/TIP.2020.3037467.

[3] J. Liu, Y. Feng and H. Wang, "Facial Expression Recognition Using Pose-Guided Face Alignment and Discriminative Features Based on Deep Learning," in IEEE Access, vol. 9, pp. 69267-69277, 2021, doi: 10.1109/ACCESS.2021.3078258.

[4] J. Liu, H. Wang and Y. Feng, "An End-to-End Deep Model With Discriminative Facial Features for Facial Expression Recognition," in IEEE Access, vol. 9, pp. 12158-12166, 2021, doi: 10.1109/ACCESS.2021.3051403.

[5] W. Huang, S. Zhang, P. Zhang, Y. Zha, Y. Fang and Y. Zhang, "Identity-Aware Facial Expression Recognition Via Deep Metric Learning Based on Synthesized Images," in IEEE Transactions on Multimedia, vol. 24, pp. 3327-3339, 2022, doi: 10.1109/TMM.2021.3096068.

[6] S. Chen, X. Zhu, Y. Yan, S. Zhu, S. -Z. Li and D. -H. Wang, "Identity-Aware Contrastive Knowledge Distillation for Facial Attribute Recognition," in IEEE Transactions on Circuits and Systems for Video Technology, vol. 33, no. 10, pp. 5692-5706, Oct. 2023, doi: 10.1109/TCSVT.2023.3253799.

[7] Y. Yan, Y. Huang, S. Chen, C. Shen and H. Wang, "Joint Deep Learning of Facial Expression Synthesis and Recognition," in IEEE Transactions on Multimedia, vol. 22, no. 11, pp. 2792-2807, Nov. 2020, doi: 10.1109/TMM.2019.2962317.

[8] T. Shahzad, K. Iqbal, M. A. Khan, Imran and N. Iqbal, "Role of Zoning in Facial Expression Using Deep Learning," in IEEE Access, vol. 11, pp. 16493-16508, 2023, doi: 10.1109/ACCESS.2023.3243850.

[9] H. Yang, L. Liu, W. Min, X. Yang and X. Xiong, "Driver Yawning Detection Based on Subtle Facial Action Recognition," in IEEE Transactions on Multimedia, vol. 23, pp. 572-583, 2021, doi: 10.1109/TMM.2020.2985536.

[10] A. A. Alzahrani, "Bioinspired Image Processing Enabled Facial Emotion Recognition Using Equilibrium Optimizer With a Hybrid Deep Learning Model," in IEEE Access, vol. 12, pp. 22219-22229, 2024, doi: 10.1109/ACCESS.2024.3359436.

[11] C. Liu, K. Hirota, J. Ma, Z. Jia and Y. Dai, "Facial Expression Recognition Using Hybrid Features of Pixel and Geometry," in IEEE Access, vol. 9, pp. 18876-18889, 2021, doi: 10.1109/ACCESS.2021.3054332.

[12] N. Yu and D. Bai, "Facial Expression Recognition by Jointly Partial Image and Deep Metric Learning," in IEEE Access, vol. 8, pp. 4700-4707, 2020, doi: 10.1109/ACCESS.2019.2963201.

[13] U. Nawaz, Z. Saeed and K. Atif, "A Novel Transformer-Based Approach for Adult's Facial Emotion Recognition," in IEEE Access, vol. 13, pp. 56485-56508, 2025, doi: 10.1109/ACCESS.2025.3555510.

[14] S. Hossain, S. Umer, R. K. Rout and H. A. Marzouqi, "A Deep Quantum Convolutional Neural Network Based Facial Expression Recognition For Mental Health Analysis," in IEEE Transactions on Neural Systems and Rehabilitation Engineering, vol. 32, pp. 1556-1565, 2024, doi: 10.1109/TNSRE.2024.3385336.

[15] B. Fang, X. Li, G. Han and J. He, "Rethinking Pseudo-Labeling for Semi-Supervised Facial Expression Recognition With Contrastive Self-Supervised Learning," in IEEE Access, vol. 11, pp. 45547-45558, 2023, doi: 10.1109/ACCESS.2023.3274193.

[16] V. S. Bhati, N. Tiwari and M. Chawla, "A Generalized Zero-Shot Deep Learning Classifier for Emotion Recognition Using Facial Expression Images," in IEEE Access, vol. 13, pp. 18687-18700, 2025, doi: 10.1109/ACCESS.2025.3533580.

[17] J. Wang et al., "Multiple Genetic Syndromes Recognition Based on a Deep Learning Framework and Cross-Loss Training," in IEEE Access, vol. 10, pp. 117084-117092, 2022, doi: 10.1109/ACCESS.2022.3218160.

[18] R. Wadhawan and T. K. Gandhi, "Landmark-Aware and Part-Based Ensemble Transfer Learning Network for Static Facial Expression Recognition from Images," in IEEE Transactions on Artificial Intelligence, vol. 4, no. 2, pp. 349-361, April 2023, doi: 10.1109/TAI.2022.3172272.

[19] C. Shi, C. Tan and L. Wang, "A Facial Expression Recognition Method Based on a Multibranch Cross-Connection Convolutional Neural Network," in IEEE Access, vol. 9, pp. 39255-39274, 2021, doi: 10.1109/ACCESS.2021.3063493.

[20] H. Zhang, "Expression-EEG Based Collaborative Multimodal Emotion Recognition Using Deep AutoEncoder," in IEEE Access, vol. 8, pp. 164130-164143, 2020, doi: 10.1109/ACCESS.2020.3021994.

[21] M. T. Naseem, C. -S. Lee and N. -H. Kim, "Facial Expression Recognition Using Visible, IR, and MSX Images by Early and Late Fusion of Deep Learning Models," in IEEE Access, vol. 12, pp. 20692-20704, 2024, doi: 10.1109/ACCESS.2024.3362247.

[22] D. Chen, P. Song and W. Zheng, "Learning Transferable Sparse Representations for Cross-Corpus Facial Expression Recognition," in IEEE Transactions on Affective Computing, vol. 14, no. 2, pp. 1322-1333, 1 April-June 2023, doi: 10.1109/TAFFC.2021.3077489.

[23] S. Wang, H. Shuai and Q. Liu, "Phase Space Reconstruction Driven Spatio-Temporal Feature Learning for Dynamic Facial Expression Recognition," in IEEE Transactions on Affective Computing, vol. 13, no. 3, pp. 1466-1476, 1 July-Sept. 2022, doi: 10.1109/TAFFC.2020.3007531.

[24] B. Han, W. -H. Yun, J. -H. Yoo and W. H. Kim, "Toward Unbiased Facial Expression Recognition in the Wild via Cross-Dataset Adaptation," in IEEE Access, vol. 8, pp. 159172-159181, 2020, doi: 10.1109/ACCESS.2020.3018738.

[25] Y. Shang et al., "LQGDNet: A Local Quaternion and Global Deep Network for Facial Depression Recognition," in IEEE Transactions on Affective Computing, vol. 14, no. 3, pp. 2557-2563, 1 July-Sept. 2023, doi: 10.1109/TAFFC.2021.3139651.