Journal of Computer Science

Original Research Paper

An Innovative Method for Recognizing Face Expressions Based on Genetic Algorithm and Extreme Learning Based Hybrid Model

[1]Tushar Dhar Shukla, [2]Md Oqail Ahmad, [3]Puneet Sapra, [4]Ratna Kumari Tamma, [5]Sujatha Kamepalli, [6]C. Malarvizhi and [7]V. Bhoopathy

[1]*Department of Physical Education and Sports, Chandigarh University, Mohali, Punjab, India*
[2]*Department of Computer Science and Engineering, Vignan's Foundation for Science, Technology and Research, Vadlamudi, Guntur, Andhra Pradesh, India*
[3]*Department of Computer Science and Engineering, Rayat Bahra University, Mohali Punjab, India*
[4]*Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Green Fields, Vaddeswaram, Guntur, Andhra Pradesh, India*
[5]*Department of Information Technology, Vignan's Foundation for Science Technology and Research, Vadlamudi, Guntur, Andhra Pradesh, India*
[6]*Department of MCA, Erode Sengunthar Engineering College, Perundurai, Tamil Nadu, India*
[7]*Department of Computer Science and Engineering, Sree Rama Engineering College, Tirupathi, India*

Article history
Received: 31-05-2024
Revised: 18-07-2024
Accepted: 05-11-2024

Corresponding Author:
V. Bhoopathy
Department of Computer Science and Engineering, Sree Rama Engineering College, Tirupathi, India
Email: v.bhoopathy@gmail.com

Abstract: The detection of facial action coding system Action Units (AU) and other facial expressions has been a major focus of computer science for more than 20 years categorizing discrete emotional states based on facial expressions. While several widely used face expression databases exist, standardization and comparability remain key challenges. The lack of a universally recognized assessment procedure and insufficient information to replicate documented outcomes hinder progress and comparison in the field. To address these challenges, this research proposes a periodic challenge in facial expression recognition to facilitate fair comparisons and provide insights into the field's advancements, obstacles, and focal points. Additionally, a novel preprocessing method is introduced to remove illumination effects from facial images efficiently. Gabor filters are applied to preprocess images, enhancing subsequent digital image processing by improving image quality. The face extraction stage utilizes the Viola-Jones algorithm to identify faces in received images, followed by ROI segmentation to estimate the face's dimensions and automatically split ROIs into mouth and areas surrounding the eyes and forehead using projective integrals and picture moments. Feature extraction employs Shi Tomasi Corner Points to extract corner points, which are then utilized with ELM, ELM-CNN, or GA-ELM to train the model. In the proposed research study, the proposed GA-ELM model achieved recognition accuracy of 96.88%, outperforming other state-of-the-art methods such as ELM-CNN and ELM, which achieved accuracies of 90.53 and 86.24% respectively. This demonstrates the superior performance of the GA-ELM model in facial expression recognition tasks.

Keywords: Extreme Learning Machine (ELM), Facial Expression Recognition (FER), Shi Tomasi Corner Point, Genetic Algorithm (GA)

Introduction

The ability to automatically discern facial expressions is vital for its application in numerous domains due to its significance in communication. The development of socially competent robots, adaptive systems, and systems that are socially aware is dependent on HCI researchers' capacity to identify users' emotional states. Educators can improve the online learning environment by learning what students dislike. Diagnostic tools for Autism Spectrum Disorder (ASD) include arousal, valence, emotion, and action units; pain detection is used by therapists to monitor patients' treatment progress. Specialists in driver assistance need to monitor their emotional and attentive states, as well as their level of weariness, to ensure the safety and comfort of their customers while driving. Facial Expression Recognition (FER) has been a successful area for deep learning's application due to its exceptional ability to acquire visual semantic information. However, the excessive number of parameters and FLOPs in most deep-learning models makes them impractical. A number of researchers have considered developing a time-based FER model to improve upon the existing one. Having the capacity to converse in person is fundamental for carrying out day-to-day tasks.

Facial expression analysis relies on motion tracking and emotion detection as its two primary components. The three primary parts of an AFEA system are expression recognition, data extraction and representation, and face detection are illustrated in Fig. (1). Many practical uses rely on automatic Facial Expression Recognition (FER), such as in service robots, driver frailty detection and human-computer interface. Tian *et al*. (2011) thanks to large-scale datasets like AffectNet, RAF-DB, and EmotioNet, several deep learning-based FER algorithms have been developed and shown encouraging results. However, the ambiguity problem is still a threat to the efficiency of FER. In order to train the FER model, it is usual practice to label face photos with one of the three primary expressions. Having said that, "category" might signify something different to many individuals. In order to make things clearer, we poll users and randomly select two images from AffectNet to illustrate our points. By displaying the likelihood of each class, a label distribution can provide a more accurate depiction of the visual characteristic. Due to the significance of both verbal and nonverbal clues in interpersonal relationships, facial expressions have significant weight. As a kind of nonverbal communication, facial expressions allow us to send vital signals, such as making eye contact. Additionally, nonverbal communication can also be expressed through gestures and body language. Being able to read people's expressions and interpret their faces is a common ability. However, it is still very difficult to create an automated system that can understand at the same level. This area is fraught with difficulties, including but not limited to: Retrieving facial emotion data, recognizing facial characteristics, Classifying emotions, and Responding differently depending on head position, occlusions, or lighting. Online education, advertising, media, healthcare, security, law enforcement, and socially intelligent robots are just a few areas that could use artificial intelligence (FER). One example is software that recognizes expressions. Many areas of study that deal with human-computer interactions-including AI, psychology, multiplayer networking, and countless more-may benefit from automated facial expression recognition. Facial expressions are one of the most direct ways for a person to communicate their emotions. A lot of people are starting to pay attention to automatic Facial Expression Recognition (FER) lately because of all the cool things it could do in fields like digital entertainment, psychology, medicine, security, and driving monitoring. Subtle variations in facial appearance and huge subject-dependent differences make the FER problematic. There has been significant progress in the field of face expression detection in recent years, but it remains challenging to develop trustworthy algorithms for challenging scenarios. These problems include but are not limited to, insufficient training data, highly nonlinear changes in facial expression, large individual differences, and extremely variable poses. Facial expression recognition uses Ekman's six universal expression categories, contempt, fear, happiness, sorrow, and surprise recognize and categorize a specific face image into different emotional states. Many FER-specific algorithms have been created for this purpose; These algorithms, when trained on face photos taken from the front or almost frontal perspective, provide remarkable results. The findings of the research demonstrate that the proposed GA-ELM model significantly outperforms other methods in facial expression recognition tasks. The key quantitative results show that the GA-ELM model achieves a recognition accuracy of 96.88%, which is superior to the 90.53% accuracy of the ELM-CNN method and the 86.24% accuracy of another method tested. These results indicate that the GA-ELM model is highly effective in improving classification accuracy due to its optimization capabilities and robust training mechanisms.

Literature Survey

Facial expressions are among the most common, easy, and efficient ways for people to convey their feelings and intentions. There are 29,673 unique face pictures in the Real-world Affective Face Data Base (RAF-DB), which is a database that actually exists. Wu and Cui (2023) the samples are given seven basic emotion labels and eleven compound emotion labels by a combination of trustworthy estimation and manually crowd-sourced annotation. All of the images were taken from the basic emotion set. We uploaded 91,794 face photos to the Expression in the Wild database (ExpW) from a Google image search. Each photo of a face was given a number by a human annotator who expressed seven different kinds of views. Zhang *et al*. (2018a) We culled the images devoid of facial features during the annotation process. Cyclical picture formation was proposed by Sun *et al*. (2023) for GAN-based unsupervised cross-view face expression identification. As an example of our method's operation, we present face images with varying placements and occlusions.

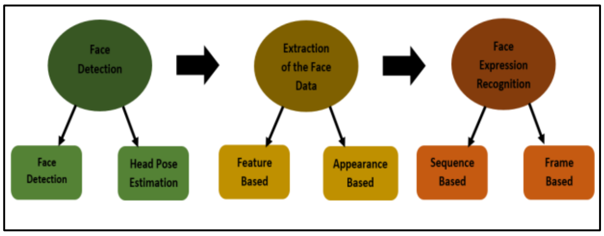

Fig. 1: Fundamental frameworks for systems that analyze facial expressions

Face deformation, seen as out-of-order sub-regions, occurs during the resizing process and in different positions, as opposed to frontal faces. In their proposal for dense prediction at the pixel level, Pyramid Vision Transformer (PVT) (Li *et al*., 2023) for feature extraction that doesn't require convolutions. Taking cues from the original transformer as well as other great derivatives, we first propose utilizing transformers for FER directly. Fan *et al*. (2020) haven't come across any endeavors that have sought to document the connections between fundamental traits with the aim of facial expression recognition. We simulate extensive interdependence between sequences of inputs employing transformers through the use of the global self-attention device. Our system is able to operate by utilizing label relations in the knowledge graph and "side" information from a small, clean dataset. Train a multi-task network to detect and remove unwanted annotations from images while it learns to identify them (Meng *et al*., 2019). For deep CNNs, select reliable images with the help of supplemental image regularization in cases where the labels are noisy. The remarkable outcomes of these two approaches (Li *et al*., 2018a), prove that FER benefits from deep features. In contrast, since region-based attention networks dynamically highlight critical face characteristics, they are perfect for FER in the field even in the presence of occlusions or clutters. Fan *et al*. (2022) Proposed utilizing Region Attention Networks (RAN) to prioritize certain facial features for occlusion and variant FER. According to (Li *et al*., 2020), a Convolutional Neural Network (CNN) with an attention mechanism for occlusion-aware facial recognition (FER) was trained to concentrate on the facial features that are most indicative of discrimination. Some additional constraints or distributions that can be assumed on the noisy samples include methods that model the noise with a softmax layer by connecting the latent correct labels to the noisy ones, methods that regularize the deep networks on corrupted labels using a MentorNet and a specific loss for randomly flipped labels (Zhang *et al*., 2018b). As a first step in the FER project, think about the problem of inconsistent annotation across different FER datasets and propose using these uncertainties to improve FER. Reference (Wu *et al*., 2020) explains how to construct a bottom-up and top-down architecture to get low-resolution attention features. Due to the materials' singular attention head, any facial flaws will most likely be the focal point. However, our method allowed us to simultaneously engage numerous non-overlapping attention zones, allowing us to collect data from different geographic regions. Discriminative loss functions have shown promise in a number of recent studies that aim to solve the FER problem. Reference (Zhu *et al*., 2019) creates a distribution-agnostic discriminative loss function by combining the strengths of center loss and softmax loss. A center loss and a softmax loss are two such examples; The former would combine features belonging to the same class and the latter would split features belonging to nearby classes. Rapid brow furrowing indicates a number of changes, such as facial muscle movements, changes in the shape and positioning of features, temporary wrinkles, and overall changes in the face's look. The duration of these facial flashes is quite brief. Although sluggish and static signals are infamously tough to deal with, each of these three signals has its own set of customizing choices. The face is a multi-message system in addition to a multi-signal system. According to (Farzaneh and Qi, 2020), a person's facial expressions reveal a great deal about their personality, status, age, intelligence, attractiveness, and quality of life. Different FER methods are the main emphasis of this study, which comprises three primary steps: Preprocessing, Feature extraction, and Classification. The group-wise spatial attention module (SGE) illustrated in reference (Xu *et al*., 2020) divides the spatial-wise data into multiple categories so that a potential spatial link can be learned. By making use of the complementary nature of channel-wise and spatial-wise information, as time goes on, the Convolutional Block Attention Module (CBAM) establishes a connection between spatial attention and channel attention to gain rich attention characteristics. The rectangular feature and the cascade AdaBoost algorithm were utilized in both methods, which are based on the Viola-Jones Algorithm. The fuzzy inference system is a well-liked method for automatic face expression recognition (Liu *et al*., 2019). Finally, new classification techniques have been proposed for application in this field by a couple of studies. We used many feature extraction techniques with several classification algorithms. The main objective of trying out different combinations was to find the best one for emotion recognition. Extra techniques like Principal Component Analysis (PCA), Haar Feature Selection Technique (HFST) (Liliana *et al*., 2019), feature distribution entropy, K-means clustering, stochastic neighbor embedding, distance vectors, gradient-based ternary texture patterns, Local Binary Patterns (LBP), spectral regression, Gaussian curvature, Hidden Markov Models, Dynamic Bayesian networks, Dynamic Bayesian networks and conditional random field models were used for face expression recognition. Engineering features can be divided into three subsets: Local features based on textures, Global features based on geometry, and Hybrid features. A few examples of texture-based features are SIFT, HOG, Histograms of LBP, Gabor wavelet coefficients, and many more. Abdulrazaq *et al*. (2021) Geometrically based global characteristics are based on landmark locations surrounding the mouth, eyes, and nose. Combining numerous artificial features is one approach to hybrid feature extraction, which can improve the representation. The learned properties of a shallow

DOI: 10.3844/jcssp.2025.388.398

CNN are robust against changes in face orientations and sizes. Deep Convolutional Neural Networks (CNNs) for feature extraction. Featuring an intuitive addition to the center loss, this research presents a Feature Clustering Network (FCN) that maximizes variances both within and between classes. In contrast to earlier methods (Li *et al*., 2019), our methodology requires just a few hyper-parameters and doesn't require any further computations beyond the variation of the cluster centers recent research has shown promising results by concentrating on attention mechanisms that pay close attention to local aspects (Li *et al*., 2018b). We believe that a person's expression is conveyed simultaneously by a multitude of facial characteristics, such as their eyes, nose, lips, chin, and eyebrows. Here the proposed approach uses four phases which are preprocessing, segmentation, feature extraction, and training the model using GA-ELM. The novelty of the research lies in the integration of the Genetic Algorithm (GA) with the Extreme Learning Machine (ELM) to form the GA-ELM model. This hybrid approach optimizes critical parameters such as activation functions, bias input weights, and regularization coefficients, which are crucial for enhancing model performance. By combining the evolutionary optimization strengths of GA with the fast learning capabilities of ELM, the GA-ELM model introduces a new paradigm in facial expression recognition that provides higher accuracy and better generalization compared to existing methods.

Proposed System

Many fields, including psychology, product marketing, and human-computer interaction, find facial expression detection to be an important topic of study. Image quality, illumination, and face orientation are the primary determinants of the accuracy with which an autonomous system can classify static photographs as input. Applying the Active Appearance Model, a thorough model-based approach can help alleviate some of these problems. An approach that reliably detects expressions will be described in detail into emotional categories such as joy, anger, sadness, surprise, fear, and contempt. It can also identify tiny, localized facial features by analyzing subtle muscle movements as outlined in the Facial Action Coding System (FACS). The proposed approach will demonstrate expression analysis and synthesis are two potential uses for the technology.

Here the proposed approach will go over the method, which involves preprocessing, segmentation, feature extraction, and classification; Fig. (2) shows the graphical representation of this framework.

Dataset

The FERV39k dataset is sourced from various scenes in videos. It contains 39,000 video clips capturing a wide range of facial expressions in diverse scenarios. The dataset is typically split into training and validation sets, though the specific percentage split can vary depending on the implementation. Commonly, an 80/20 split is used for training and validation, respectively (Yan *et al*., 2024).

Preprocessing-Gabor Filter

Noise reduction, contrast enhancement, edge detection, and outline highlighting are all possible with the use of filters. Linear and non-linear filters are the two main types. Many applications rely on filters due to their computational features and biological significance. These include pattern analysis, facial recognition, iris recognition, optical character recognition, and fingerprint recognition. One linear filter that can be used to improve edges is the Gabor filter. It accomplishes an optimal resolution in the frequency and spatial domains by acting as a bandpass filter for the local spatial frequency distribution. Gabor filters are quite similar to the way the human visual system works in terms of orientation and frequency. Saeed *et al*. (2022) in this case, the impulse response takes the form of an enhanced sinusoidal wave. With its multiplication convolutional property, this filter can handle a plethora of transformations, image attributes, operators, frequencies, and edge detection features. When collinear or extended segments are present, the filter adjusts its response based on the region value. The capacity to construct orientation-specific filters is an advantage. The input image's features are extracted pixel-by-pixel using 2D Gabor filters. Analysis of space-invariant intensity photographs and curved line traces are both accomplished with the use of Gabor filters. To examine features at grey levels and identify straight-line patterns, the extended Gabor filter was developed. One way to look at the filter is as a Gaussian-influenced sinusoidal plane, where the plane is determined by the frequency and the direction. To describe the Gabor filter, we need to know its orientation, standard deviation, and frequency. A predetermined combination of dilations and rotations forms the basis of the filter. "Gabor space" describes the result of applying this filter to a signal through convolution. An interesting aspect of the Fourier transform of the Gabor filter response is its multiplication convolution, which arises from the intersection of the Fourier transforms of a Gaussian function and a Harmonic function.

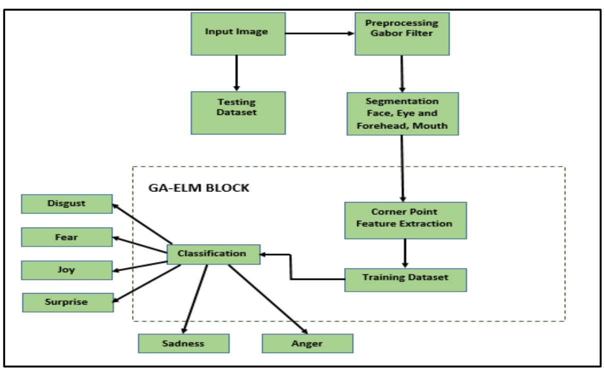

Fig. 2: Proposed architecture

The Gabor filter parameters are adjusted precisely by varying orientation and wavelength. The Gabor filter function is expressed as:

$$n(w, v, \alpha, \theta, \beta, \rho, \delta) = exp\left(\frac{-w'^2 + \delta^2 v'^2}{2\rho^2}\right) cos\left(2\pi \frac{w\prime}{\alpha} + \beta\right) \quad (1)$$

Here are the variables that are being discussed: α for the sinusoidal factor's wavelength, θ for the Gabor function's normal to the stripe's orientation, β for the phase offset, ρ for the Gaussian envelope's standard deviation and δ for the spatial ratio. Taken together, they characterize the Gabor function's elliptical support. With the filter's settings tweaked to match wavelength and orientation, edge enhancement is achieved at a certain orientation with distinct parameter values.

Segmentation

Face Segmentation

At first, the FER system uses the Viola-Jones method to detach the face from the backdrop during picture processing. Facial emotion detection could be hindered by irrelevant noise in certain areas of the detected face image, such as the hair, ears, or backdrop. Removing extraneous information by adjusting the dimensional parameters yields a more accurate estimate of the face dimensions, which in turn improves the facial expression recognition system's identification rate. The face picture is segmented into its component colors, green, and using the RGB color space. It takes out the red channel from the green channel to emphasize skin tones, denoted as 〖L〗 -*CN* (w, v) (Xu *et al*., 2020). The resulting image is binarized by applying a cutoff value found through experimentation, *Th*, according to Eq. (1):

$$L_s(w, v) = \begin{cases} 0 & L_{CN}(w, v) < Th \\ 225 & L_{CN}(w, v) \geq Th \end{cases} \quad (2)$$

When applying YUV color space for facial image segmentation, we substitute the plane 〖L〗 -CN (w, v) in Eq. (1) with either of the planes ( ), ( ), or ( ) using the threshold Th. The threshold value is determined by averaging the values when using planes Z or Y and by averaging the values when using the V plane.

Face Dimension Adjusting

Following the binarization of the image, the sec of the binarized picture that comes out of it is calculated in the subsequent steps:

$$H_{ed} = \sum_{w=1}^{G_S} \sum_{v=1}^{H_S} w^e v^d L_S(w, v) \quad (3)$$

$L_S(w, v)$ represents the intensity of a binary picture at position (w, v). G_S and H_S which correspond to the image's column and row counts, respectively. The image's moment order is denoted by e and d. To find the centroid, use Eq. (2) as follows:

$$w_r = {H_{1,0}}/{H_{0,0}} \quad (4)$$

$$v_r = {H_{0,1}}/{H_{0,0}} \quad (5)$$

Using Eqs. (3-5), the following variables are defined:

$$t = \frac{H_{2,0}}{H_{0,0}} - w_r^2 \quad (6)$$

$$s = 2\left(\frac{H_{1,1}}{H_{0,0}} - w_r v_r\right) \quad (7)$$

$$r = \frac{H_{0,2}}{H_{0,0}} - v_r^2 \quad (8)$$

Subsequently, by utilizing Eqs. (6-8), the breadth of the facial picture can be approximated in the following manner:

$$X = 2\sqrt{\frac{(t+s) - \sqrt{s^2 - (t-r)^2}}{2}} \quad (9)$$

By utilizing X, one may estimate the left (w_i) and right (w_c) borders of the face image:

$$w_i = [w_r] - \left[\frac{X}{2}\right] \quad (10)$$

$$w_i = [w_r] - \left[\frac{X}{2}\right] + [X] \quad (11)$$

Next, by utilizing X, the top boundary of the facial picture can be approximated in the following manner:

$$v_z = [v_r] - 0.84\left[\frac{X}{2}\right] \quad (12)$$

The facial image is segmented using Eqs. (10-12):

Eye and Forehead Segmentation

An important part of the current technology for detecting facial emotions is the ability to segment the face into areas around the eyes and the forehead. Equation (11) divides the divided face region horizontally into three sections (A-C) with respect to height, beginning at the point V-z. The area around the eyes and forehead is called Region A. Investment Return.

Mouth Segmentation

Using the detected facial region as a starting point, divide it into three equal horizontal pieces beginning at the edge edges. This will allow you to segment the mouth region. At the mouth lies Region C, the area of interest. In this case, segmenting just the mouth region is necessary, as opposed to the forehead/eye region. The image is subjected to a histogram equalization using the $LCN(w, v)$ function.

Extraction of Features from Corner Point

Using the Region-of-Interest (RoI) windows acquired for the lips and eyebrows, the Shi Tomasi corner point detector is employed to extract corner point features. The Tomoasi Shi approach identifies windows that exhibit significant intensity variations when shifted in both horizontal (W) and vertical (V) directions, hence calculating the gradients in the W and V directions. A score of C is calculated for each window discovered. Each window is associated with a window function denoted as $x(w, v)$:

$$H = x(w, v)wL$$

where:

$$L = \begin{bmatrix} \sum_{(w,v)} L_w^2 \sum_{(w,v)} L_w L_v \\ \sum_{(w,v)} L_v L_w \ \sum_{(w,v)} L_v^2 \end{bmatrix} \tag{13}$$

where, L_w and L_v represent picture derivatives in the w and v directions, respectively. The value C for each window:

$$C = min(\alpha_1, \alpha_2) \tag{14}$$

where, 1 and 2 represent the eigenvalues of matrix H. By applying a threshold to the red channel, significant corners are identified and highlighted. Adjusting to compute the derivative covariation matrix in order to extract corner points for fiducial landmarks, as well as factors such as the minimum picture corner quality, minimum Euclidean distance between corners, and window size.

The chosen feature corner points for the eyebrows consist of 12 points, with 6 on the left side and 6 on the right side. The eyes have 14 feature points, with 7 on the left side and 7 on the right side. 10 feature points for the nose. There are 20 feature points for the lips.

Classification of the Model

Extreme Learning Machine

The goal of ELM was to address the fundamental constraint of traditional feed-forward neural networks. Feed-forward neural networks are slow since they utilize gradient-based learning techniques for training. Gradient descent learning techniques cause the parameters, such as weights and biases in one layer, to be influenced by parameters in other layers and tend to converge towards local minima. Researchers have suggested strategies to enhance the efficiency of feed-forward neural networks over time. ELM has been successful in enhancing the performance of Single-hidden Layer Feedforward Neural Networks (SLFNs) as a solution. ELM initializes the parameters connecting the input layer to the hidden layer randomly, whereas the output weights are determined using the least-square technique. ELM's rapid learning skills stem from its capacity to learn without iteration, leading to quicker convergence compared to traditional feed-forward neural networks. Randomizing the input parameters in ELM prevents the occurrence of local minima commonly seen in traditional networks. ELM surpasses other learning algorithms in learning speed, ease of implementation, and generalization performance. Because of these characteristics, ELM has been utilized in several disciplines like as regression, classification, and clustering.

Algorithm 1: Process of training ELM as an estimator for proposed GA and for the proposed prediction model

Process of Training ELM as an Estimator for Proposed GA and Prediction Model
Input: A training set with N samples (From GA) consisting of event features and facial expression labels (w_l, v_l) | w_l, v_l ∈ ER, i = 1, 2, ..., G. Activation function $o(w)$ $(RELU)$
Output: Facial Expression Prediction

1. Randomly assign weights x_l and bias s_l, l = 1, 2, ..., G
2. Compute the output matrix based on the input event feature set, say M
3. Compute output weights based on input Facial Expression labels and output matrix Φ
4. Train the ELM network using the Least Square Solution φ' to the linear system $H\Phi = A$
5. Analytically tune output weights
6. Perform prediction for the test set and observe prediction performance

In Algorithm (2), we can see how to train the ELM network to estimate features for the feature selection model that is presented; this model doubles as the final prediction model. Based on the provided lab events or features set and face-specific facial expression labels, the output matrix is created after random parameter and weight initialization. Training involves iterative optimization of the weights between the hidden and output layers, culminating in the observation of performance in facial expression prediction (Krishnan and Kamath, 2019). Given N unique training sets (W_l, a_l), where $W_l = [w_{l1}, w_{l2}, \dots, w_{lg}] \in C^g$. Let A be a vector in C^g *and* a_l with components $[a_{l1}, a_{l2}, \dots, a_{lg}] \in C^g$. A is an element of the set of real numbers raised to the power of h. I represents the number of hidden nodes and $n(w)$ represents the activation function.

Figure (3) shows the results of training the proposed model using the ELM architecture. A support vector neural network (SLFNN) is the primary architecture component; the number of input nodes is determined by the quantity of training data. In order to forecast the probability of mortality, the hidden layer contains fifty nodes and the output layer contains one. We experimented with several setups for the hidden layer's node count, however beyond 50 nodes, training time skyrocketed and performance remained the same. Hence, we found that the optimal number of nodes for the hidden layer should be 50. ELM can be executed by randomly allocating the parameters of the hidden nodes (x, s), calculating the hidden layer output matrix (M), and determining the output weights (φ). Given G samples, the desired output A can be calculated using the following equation:

$$M\varphi = A \tag{15}$$

DOI: 10.3844/jcssp.2025.388.398

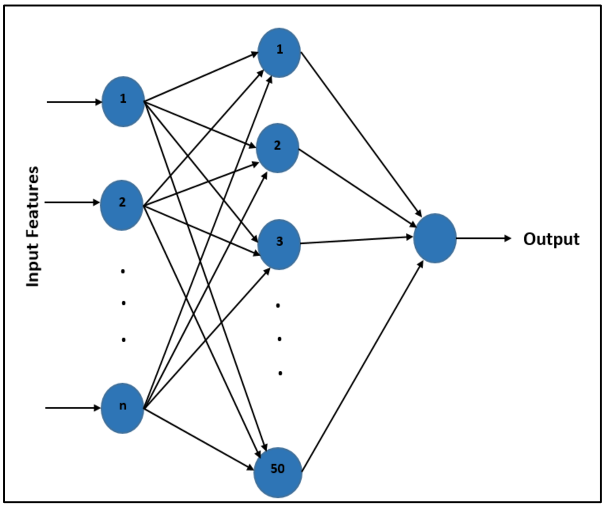

Fig. 3: Administration of the ELM prediction system

Were:

$$H(x_1,\dots,x_I,s_1,\dots,s_I,w_1,\dots,w_I)$$

$$\begin{bmatrix} n(x_1,w_1+s_1), & \dots, & n(x_I,w_1+s_I) \\ \vdots & \dots, & \vdots \\ n(x_1,w_G+s_1), & \dots, & n(x_I,w_G+s_G) \end{bmatrix}_{GI} \quad (16)$$

$$\varphi = \begin{bmatrix} \varphi_1^A \\ \varphi_2 \\ \vdots \\ \varphi_I^A \end{bmatrix}_{Ih} \quad (17)$$

$$A = \begin{bmatrix} a_1^A \\ a_2 \\ \vdots \\ a_I^A \end{bmatrix}_{Gh} \quad (18)$$

The weights between the hidden layer and the output layer, denoted by φ, are computed using the least-squares technique to minimize the error between the target and the output:

$$\hat{\varphi} = M \dagger A \quad (19)$$

$M \dagger$ represents the Moore-Penrose generalised inverse of matrix M and A is the target.

Genetic Algorithm (GA)

Genetic algorithms are a form of evolutionary algorithm that function based on the concepts of evolution and natural selection. Genetic algorithms randomly select a set of chromosomes to represent the problem being solved. These groups of chromosomes are known as populations during the stages of evolution. An assessment function is utilized to determine the optimal choice for every chromosome. Species evolution is influenced by crossover and mutation. When a genetic algorithm is utilized for problem-solving, three factors will influence the algorithm's effectiveness:

1) Choosing the fitness function
2) Determining how individuals are represented
3) Setting the settings for the genetic algorithm parameters

Proposed GA-ELM

Initiate the optimization process by specifying the number of hidden neurons and activation function. Following the setup of the ELM network, the initial input values are randomly generated. The input values comprise weights and biases. The model is trained to derive optimal output values. (Krishnan and Kamath, 2019) The output values generated by ELM will serve as the starting population for the Genetic Algorithm (GA). We will now implement the concepts of evolution. Genetic Algorithms are a set of bio-inspired optimization algorithms that utilize the principles of Darwinian evolution to determine the optimal weights for a given problem. The most optimal weights are determined by calculating a fitness score based on the principle of survival of the fittest in natural selection. The GA was implemented by designating each attribute of the intrusion, save for the type, as an equation input denoted by l and considering the complete set of intrusion observations as the population, g. A fitness score was calculated using the following formula:

$$fitness - score = \sum_{g=1}^{g} g \times l \quad (20)$$

Although the selection of a fitness score may seem random, it is recommended to opt for a fitness score that can be computed efficiently to avoid slowing down the genetic algorithm's computation speed. The purpose of calculating the fitness score for each individual in the set of equation inputs was to discover individuals with the highest fitness scores, as natural selection favors the fit individuals. Together, these individuals would constitute the mating pool responsible for producing future generations. The number of mating parents for this work was established as 4. Increasing the number of mating parents can enhance the quality of children in certain populations.

Algorithm 2: Proposed GA-ELM

The GA-ELM Algorithm

1. Normalize the sample data and map the training sample set $\{w_l, v_l\} \in w^h \times C^h$ to the interval [0,1]
2. Optimize the input dimension of the sample set.
3. Compute output weights based on input Facial Expression labels and output matrix Φ
4. Train the ELM network using the Least Square Solution to the linear system $H\Phi$=A

5. Construct the Face Expression prediction model based on ELM. In this paper, the sigmoid function and linear function are chosen as activation functions of the hidden layer node
6. Optimize the type of activation function $B_k \in \{0,1,2\}$, $k = 1, ...m$, regularization coefficient R, input weight $y_l \in [0,1]$, bias $s_l \in [0,1]$, and construct the Facial Expression prediction model based on GA-ELM
7. Apply the GA-ELM to a Facial Expression prediction in a special region. The model is constructed by this method, that is $\hat{y}_{a+m} = o(w_a)$, $\forall_a = \nabla ...$k
8. where $o(w_a)$, is constructed by ELM method, $\forall$ represents the embedding dimension of the prediction model, w_a is the multiple-dimension input vector constructed by the historical facial expression values $y_{t+1}, y_{t+2}, ...y_{t-\Delta}$

To start, GA is used to preprocess the data in order to efficiently extract the input dimension for the model from the feature space. This is used to build an ELM-based model for predicting facial expressions. Then, a GA-ELM hybrid prediction model for face expression is born by optimizing the ELM activation function's type, bias, input weight, and regularization coefficient. Algorithm (2) displays the GA-ELM algorithm. The parents mated by creating a function that simulated animal breeding. In this process, a fraction from each parent was selected to contribute to the creation of the offspring. Each parent contributed a fraction of genes that together formed the complete set for the offspring. Half of each parent's genetic material contributed to the offspring's genetic makeup in this study (Wang *et al*., 2017). Mutation is a process that causes alterations in the genetic makeup of an organism, leading to differentiation between organisms at different levels. Mutation plays a crucial role in deciding the survival or potential extinction of an organism via natural selection. An integer between 1 and 1000 was randomly selected in increments of 1 to simulate mutation. The mutagenic substance was introduced into the offspring's genetic composition by inserting it into a randomly chosen location within the individual's genetic framework. The new addition will be replicated to another random site to guarantee that the mutation occurs at random spots across the individual's whole structure. Finally, the procedure was allowed to continue for 100 iterations to replicate evolution across 100 epochs.

When the termination condition is met, the optimized weights are uniformly chosen to retrain the model. This selection was constrained by the minimum and maximum weights from the final generation of the GA evolution. The most effective approach was documented and then compared based on the precision of the weights derived by genetic algorithms and random uniform generation. The activation functions utilized in this work are the sigmoid, relu, and sin functions. The hidden layer's output was calculated by taking the dot product of the input layer's output and the beta weights of the hidden layer.

Materials and Methods

The experiment was conducted on a Windows 11 PC equipped with a 32-core, 128 GB of RAM, 3.20 GHz Intel Core i7-8700 processor, and dual NVIDIA GeForce GTX 3080 Ti graphics processing units (GPUs).

Results and Discussion

The experimental results show that the GA-ELM model outperforms other methods significantly. In comparison to other strategies, GA-ELM achieved the highest accuracy rate of 96.88%, whereas the ELM-CNN method came in second with an accuracy rate of 90.53% and another method ranked third with an accuracy rate of 86.24. The reasons behind the superior performance of the GA-ELM model are multifaceted.

Optimization of Parameters: GA-ELM optimizes the type of activation function, bias input weight, and regularization coefficient using Genetic Algorithm (GA), which enhances the model's ability to generalize and accurately classify facial expressions.

Feature Selection and Dimensionality Reduction: By preprocessing the data efficiently and extracting relevant input dimensions, GA-ELM reduces the complexity of the model while retaining critical information, which contributes to better performance.

Combination of GA and ELM: The hybrid model leverages the evolutionary capabilities of GA to optimize the initial population for ELM, thus combining the strengths of both approaches to improve learning and prediction accuracy.

Discriminative Information: The GA-ELM approach adds additional discriminative information to the features acquired and selected for facial expression recognition, leading to improved accuracy in identifying and classifying expressions.

Robust Training Mechanism: The training process involves multiple iterations and selection mechanisms that ensure the model evolves to a state of optimal performance through rigorous natural selection processes mimicking biological evolution.

In comparison to all of the strategies, ours performed the best in Fig. (4). This proved that our strategy added additional discriminative information to the characteristics acquired and selected for face expression recognition. The most effective performance is achieved by our GA-ELM approach. With an accuracy rate of 90.53 percent, the deep learning project ELM-CNN came in second. With an accuracy rate of 86.24%, it ranks third best. The most recent greatest performance was 96.88% accuracy, which was attained concurrently with the work proposed in this research.

One popular way to measure the precision of a predictive model, especially for jobs involving forecasting, is the Mean Absolute Percentage Error or MAPE. Averaging the absolute percentage discrepancy between the model's predicted values and the dataset's actual

DOI: 10.3844/jcssp.2025.388.398

observed values is what it measures. If you want to know what the average % difference is between your predictions and the actual numbers, MAPE can give you that number. More agreement between the actual and anticipated values is shown by lower MAPE values, whilst larger disparities are indicated by higher values. The percentage form of MAPE is common. Figure (5) illustrates the correlation between the number of iterations and the model's forecasting performance. Finding that MAPE values tend to remain stable when iteration counts exceed 50 suggests that this is a sensible upper bound to set.

When assessing the precision of a regression model, one frequent statistic is the Root Mean Square Error (RMSE). It is a measure of the average discordance between the model's projected values and the dataset's actual observations. It is possible to determine the Root-Mean-Squared Error (RMSE) by squaring the average of the squared discrepancies between the expected and observed values. The correlation between the model's forecasting performance and the amount of facial expression iterations is illustrated in Fig. (6). It is fair to set the maximum number of iterations at 70 since we know that RMSE values tend to be steady at higher iteration counts.

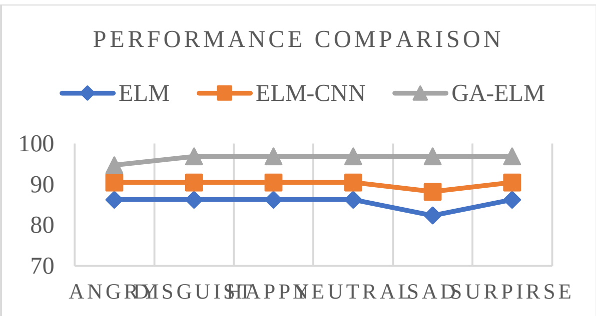

Fig. 4: Accuracy performance comparison

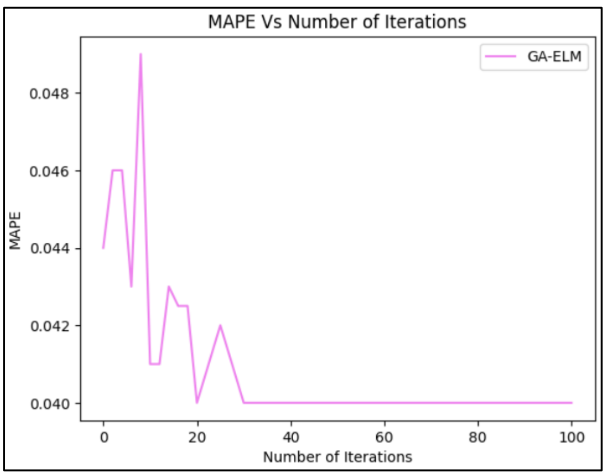

Fig. 5: MAPE vs number of iterations

In this study, the proposed GA-ELM is compared to twenty-one state-of-the-art models using four commonly used assessment metrics: Accuracy (A), Precision (P), Recall (R), and F1-score (F). However, the confusion matrix is where all the measurements get their values. Figure (7) displays the aforementioned datasets' training and validation accuracies. The dataset's training accuracy is highlighted in Peru and its validation accuracy is in chartreuse. To get confusion matrices, testing sets are put into trained GA-ELM sequentially.

The confusion matrix for GA-ELM is displayed in Fig. (8). The recognition rates for ELM-CNN and ELM were only 90.53 and 86.24%, respectively, due to the fact that there are only slight variations in the individuals' actual emotions. Additionally, the photographs are in grayscale, even though our model is designed to operate with RGB images. Neglecting to use RGB format can make it even more difficult for the network to differentiate between foreground and background objects. A 96.88% success rate was attained by our model.

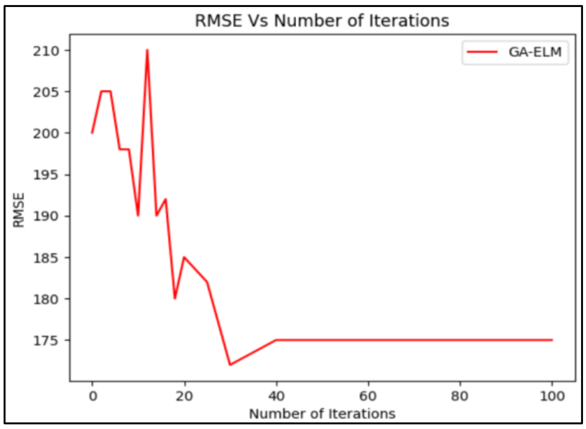

Fig. 6: RMSE vs number of iterations

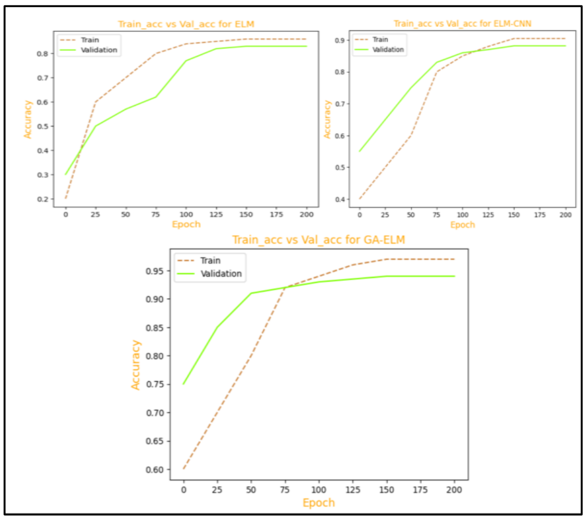

Fig. 7: Training and validation accuracies

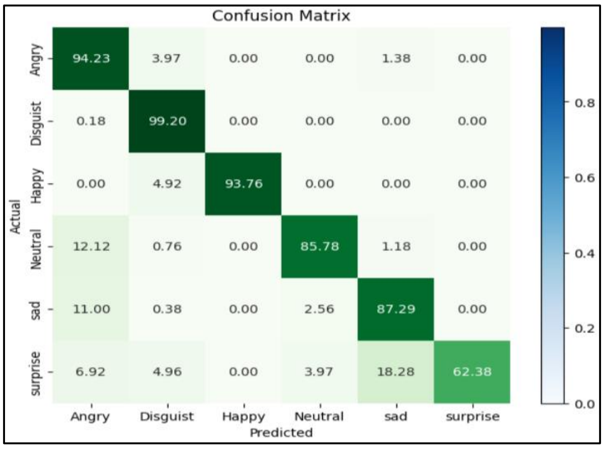

Fig. 8: Confusion matrix

Conclusion

The face conveys a variety of information regarding identification, age, sex, race, as well as emotional and mental state. Facial expressions are vital in social interactions and are frequently utilized to understand emotions through behavior. Automatic facial expression detection is a complex issue in computer vision with possible applications in Human-Computer Interaction (HCI), behavioral science, video games, etc. This research introduces a new method for automatically identifying face expressions by utilizing GA-ELM characteristics. The four steps that make up the suggested method are as follows: preprocessing, segmentation, feature extraction, and model training. The Gabor filter is employed during the preprocessing phase. Segmentation of the face, eyes, and mouth is done during the adjustment of face dimensions. When extracting characteristics, using the Shi Tomasi corner proximity detector. It uses CNN-ELM, ELM, and GA-ELM for model training. With an accuracy of approximately 96.88%, our suggested GA-ELM method outperforms the other two conventional approaches.

Future Research and Limitations

The study primarily uses specific datasets that might not cover all possible variations in facial expressions, thus limiting the model's generalization ability. The genetic algorithm's iterative optimization process can be computationally intensive and time-consuming, which might not be suitable for real-time applications without further optimization. The model is designed to operate with RGB images but the study uses grayscale images, which can affect the model's performance due to reduced information content compared to RGB images. While the study uses accuracy as a primary metric, other evaluation metrics such as precision, recall, and F1-score should also be thoroughly considered to provide a more comprehensive assessment of the model's performance across different scenarios. By addressing these limitations and focusing on the identified future research areas, the GA-ELM model's effectiveness and applicability can be further improved.

Acknowledgment

The authors would like to thank anonymous reviewers for their constructive comments and suggestions to update the manuscript.

Funding Information

This research received no external funding.

Author's Contributions

All authors equally contributed to this study.

Ethics

This manuscript is an original work. The authors declare that there are no ethical concerns associated with this submission.

References

Abdulrazaq, M. B., Mahmood, M. R., Zeebaree, S. R. M., Abdulwahab, M. H., Zebari, R. R., & Sallow, A. B. (2021). An Analytical Appraisal for Supervised Classifiers' Performance on Facial Expression Recognition Based on Relief-F Feature Selection. *Journal of Physics: Conference Series*, *1804*, 012055. https://doi.org/10.1088/1742-6596/1804/1/012055

Fan, X., Deng, Z., Wang, K., Peng, X., & Qiao, Y. (2020). Learning Discriminative Representation for Facial Expression Recognition from Uncertainties. *2020 IEEE International Conference on Image Processing (ICIP)*, 903–907. https://doi.org/10.1109/icip40778.2020.9190643

Fan, Y., Li, V. O. K., & Lam, J. C. K. (2022). Facial Expression Recognition with Deeply-Supervised Attention Network. *IEEE Transactions on Affective Computing*, *13*(2), 1057–1071. https://doi.org/10.1109/taffc.2020.2988264

Farzaneh, A. H., & Qi, X. (2020). Discriminant Distribution-Agnostic Loss for Facial Expression Recognition in the Wild. *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 1631–1639. https://doi.org/10.1109/cvprw50498.2020.00211

Li, J., Jin, K., Zhou, D., Kubota, N., & Ju, Z. (2020). Attention Mechanism-Based CNN for Facial Expression Recognition. *Neurocomputing*, *411*, 340–350. https://doi.org/10.1016/j.neucom.2020.06.014

Li, Y., Lu, G., Li, J., Zhang, Z., & Zhang, D. (2023). Facial Expression Recognition in the Wild Using Multi-Level Features and Attention Mechanisms. *IEEE Transactions on Affective Computing*, *14*(1), 451–462. https://doi.org/10.1109/taffc.2020.3031602

Li, Y., Zeng, J., Shan, S., & Chen, X. (2018a). Patch-Gated CNN for Occlusion-Aware Facial Expression Recognition. *2018 24th International Conference on Pattern Recognition (ICPR)*, 2209–2214. https://doi.org/10.1109/icpr.2018.8545853

Li, Y., Zeng, J., Shan, S., & Chen, X. (2019). Occlusion Aware Facial Expression Recognition Using CNN with Attention Mechanism. *IEEE Transactions on Image Processing*, *28*(5), 2439–2450. https://doi.org/10.1109/tip.2018.2886767

Li, Z., Wu, S., & Xiao, G. (2018b). Facial Expression Recognition by Multi-Scale CNN with Regularized Center Loss. *2018 24th International Conference on Pattern Recognition (ICPR)*, 3384–3389. https://doi.org/10.1109/icpr.2018.8545489

Liliana, D. Y., Basaruddin, T., Widyanto, M. R., & Oriza, I. I. D. (2019). Fuzzy Emotion: A Natural Approach to Automatic Facial Expression Recognition from Psychological Perspective Using Fuzzy System. *Cognitive Processing*, *20*(4), 391–403. https://doi.org/10.1007/s10339-019-00923-0

Liu, Y., Peng, J., Zeng, J., & Shan, S. (2019). Pose-adaptive Hierarchical Attention Network for Facial Expression Recognition. *ArXiv Preprint ArXiv:1905.10059*, *14*(8), 1–12. https://doi.org/10.48550/arXiv.1905.10059

Meng, D., Peng, X., Wang, K., & Qiao, Y. (2019). Frame Attention Networks for Facial Expression Recognition in Videos. *2019 IEEE International Conference on Image Processing (ICIP)*, 3866–3870. https://doi.org/10.1109/icip.2019.8803603

Krishnan, S. G., & Kamath S., S. (2019). A Novel GA-ELM Model for Patient-Specific Mortality Prediction Over Large-Scale Lab Event Data. *Applied Soft Computing*, *80*, 525–533. https://doi.org/10.1016/j.asoc.2019.04.019

Saeed, S., Shah, A. A., Ehsan, M. K., Amirzada, M. R., Mahmood, A., & Mezgebo, T. (2022). Automated Facial Expression Recognition Framework Using Deep Learning. *Journal of Healthcare Engineering*, *2022*(1), 1–11. https://doi.org/10.1155/2022/5707930

Sun, N., Lu, Q., Zheng, W., Liu, J., & Han, G. (2023). Unsupervised Cross-View Facial Expression Image Generation and Recognition. *IEEE Transactions on Affective Computing*, *14*(1), 718–731. https://doi.org/10.1109/taffc.2020.3029531

Tian, Y., Kanade, T., & Cohn, J. F. (2011). *Facial Expression Recognition*. Springer. https://doi.org/10.1007/978-0-85729-932-1_19

Wang, X., Wang, C., & Li, Q. (2017). Short-Term Wind Power Prediction Using GA-ELM. *The Open Electrical and Electronic Engineering Journal*, *11*(1), 48–56. https://doi.org/10.2174/1874129001711010048

Wu, R., Zhang, G., Lu, S., & Chen, T. (2020). Cascade EF-GAN: Progressive Facial Expression Editing with Local Focuses. *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA. https://doi.org/10.1109/cvpr42600.2020.00507

Wu, Z., & Cui, J. (2023). LA-Net: Landmark-Aware Learning for Reliable Facial Expression Recognition under Label Noise. *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*, 20641–20650. https://doi.org/10.1109/iccv51070.2023.01892

Xu, T., White, J., Kalkan, S., & Gunes, H. (2020). Investigating Bias and Fairness in Facial Expression Recognition. In A. Bartoli & A. Fusiello (Eds.), *Computer Vision – ECCV 2020 Workshops* (pp. 506–523). Springer International Publishing. https://doi.org/10.1007/978-3-030-65414-6_35

Yan, W., Yixuan, S., Yiwen, H., Zhongying, L., Shuyong, G., Wei, Z., Weifeng, G., & Wenqiang, Z. (2024). *FERV39k: A Large-Scale Multi-Scene Dataset for Facial Expression Recognition in Videos*. GitHub. https://github.com/wangyanckxx/FERV39k

Zhang, F., Zhang, T., Mao, Q., & Xu, C. (2018a). Joint Pose and Expression Modeling for Facial Expression Recognition. *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 3359–3368. https://doi.org/10.1109/cvpr.2018.00354

Zhang, Z., Luo, P., Loy, C. C., & Tang, X. (2018b). From Facial Expression Recognition to Interpersonal Relation Prediction. *International Journal of Computer Vision*, *126*(5), 550–569. https://doi.org/10.1007/s11263-017-1055-1

Zhu, K., Du, Z., Li, W., Huang, D., Wang, Y., & Chen, L. (2019). Discriminative Attention-Based Convolutional Neural Network for 3D Facial Expression Recognition. *2019 14th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2019)*, 1–8. https://doi.org/10.1109/fg.2019.8756524