STATISTICS, OPTIMIZATION AND INFORMATION COMPUTING
Stat., Optim. Inf. Comput., Vol. **13**, April 2025, pp 1557–1577.
Published online in International Academic Press (www.IAPress.org)

Novel SR-RNN Classifier for Accurate Emotion Detection in Facial Analysis

Jyoti S. Bedre [*], P. Lakshmi Prasanna

Computer Science and Engineering, KL University, Vaddeswaram, Andhra Pradesh (India)

Abstract Facial Expression Recognition (FER) is crucial for understanding human emotions in fields like human-computer interaction and psychology. Despite advances in deep learning (DL), existing FER methods often struggle with noise, lighting variations, and inter-subject variability, leading to inaccurate emotion classification. This paper addresses these challenges by proposing a novel SwikyRelu Recurrent Neural Network (SR-RNN) classifier. The aim is to enhance FER accuracy while reducing computational complexity. The methodology involves a multi-step process starting with image pre-processing using an Adaptive Mode Guided Filter (AMGF) and Contrast Limited Adaptive Histogram Equalization (CLAHE). Key facial features are extracted using the Generative Additive Active Shape Model (GAASM) and clustered into subgraphs using Radial Basis K-Medoids Clustering (RBKMC). Feature selection is optimized through the Chaotic Ternary Remora Optimization (CTRO) algorithm, with the selected features fed into the SR-RNN classifier for emotion categorization. Results from extensive testing on the CK+, FER-2013, and RAF-DB dataset shows that the proposed SR-RNN classifier significantly outperforms conventional models, achieving 98.85%, 91.79%, and 89.28% accuracy, respectively. The conclusion highlights the model's ability to enhance FER performance by effectively handling noise, illumination differences, and inter-subject variability.

Keywords Machine Learning, Facial Expression Recognition, Adaptive Mode Guided Filter (AMGF), Contrast Limited Adaptive Histogram Equalization (CLAHE) technique, Haar cascade, GAASM, RBKMC Algorithm, CTRO Algorithm, SR-RNN algorithm.

AMS 2010 subject classifications 68T10

DOI: 10.19139/soic-2310-5070-2142

1. Introduction

The face is the most expressive and communicative part of a human being, playing a vital role in conveying emotions and establishing interpersonal communication[1]. With the increasing influence of computers on our daily lives, the role of human-computer interaction (HCI) has become crucial[2]. A growing interest is in enhancing HCI to create a more intuitive and emotionally responsive connection between users and computers. Many believe that improving this interaction can lead to positive emotional responses and a more vital cognitive link between humans and machines[3]. FER is critical to this process, as it involves identifying expressions that convey basic emotions[4]. By focusing on facial expressions, we can better understand and interpret human emotions, thereby improving the effectiveness of HCI systems[5]. FER has many applications, including social robots, e-learning, criminal justice systems, smart card technology, and customer satisfaction surveys[6, 7]. Additionally, emotion recognition is crucial for psychiatrists and psychologists in diagnosing various mental health conditions[8]. Consequently, it is a vibrant area of research within pattern identification and artificial intelligence[9]. There are

[*] Correspondence to: Jyoti S. Bedre (Email: jyoti.phd2020@gmail.com). Computer Science and Engineering, KL University, Vaddeswaram, Andhra Pradesh (India).

ISSN 2310-5070 (online) ISSN 2311-004X (print)
Copyright © 2025 International Academic Press

six universally recognized essential facial appearances such as rage, disgust, fear, happiness, grief, and surprise which are shared among all human beings[10]. Initially, emotions were detected through the semantic and syntactic properties of language. However, this method often led to misinterpretations and varied significantly among different groups of people. As a result, FER approaches gained popularity[11]. A FER system typically comprises three major components: pre-processing, facial feature extraction, and emotion classification[12]. Of these, feature extraction is particularly crucial. Feature extraction methods are generally categorized into geometric-based and appearance-based extraction techniques[13].

Facial emotion analysis is challenging due to various factors such as diverse subjects, races, lighting conditions, and complex backgrounds[14]. One of the significant difficulties in facial expression recognition is robustly identifying and interpreting vital facial regions[15, 16]. Over the past few decades, various FER techniques have been developed to address these challenges using efficient classifiers [17]. Notable among these are Deep Boltzmann Networks (DBN), Convolutional Neural Networks (CNN), and Multi-Layer Neural Networks (MNN)[18, 19]. By concentrating on the most significant aspects that relate to particular angles of view, spatial-angular features can be learned to improve recognition ability further[20, 21]. However, these methods often fail to provide accurate emotion classification and require significant time for processing. To address these limitations, this paper proposes an improved facial emotion identification and classification approach using a novel SR-RNN classifier. This method leverages the RBKMC clustering graph mining technique for more precise and efficient emotion recognition.

1.1. Problem Statement

Despite the development of numerous ML and DL-based models for recognizing facial emotions, several limitations persist:

- Facial images are highly susceptible to variations in lighting and image noise, significantly impacting the performance of FER systems.
- Facial images are often taken from multiple viewpoints, making non-frontal FER particularly challenging. Addressing issues like face occlusions, accurate alignment, and precise location of facial points is essential.
- Existing methods use graph theory and mining concepts to identify frequent sub-graphs in each emotional class using the gSpan technique. However, while reducing redundant sub-graphs, the overlap metric can sometimes cause misclassification due to excessive overlap.
- Deep neural networks, which automatically learn features and achieve high recognition rates, can suffer from overfitting as the number of layers and parameters increases.
- Inter-subject variability, inconsistent and incorrect emotion classifications, and a lack of large-scale labeled training data hinder the effectiveness of deep learning networks on FER tasks.

The field of FER has witnessed significant advancements with the integration of DL techniques, yet challenges remain in real-world applications due to issues such as noise, lighting variations, and subject variability. Existing approaches like CNNs and MNN are often limited in their ability to address these complexities, leading to reduced classification accuracy and higher computational demands. The paper proposed SR-RNN classifier seeks to overcome these challenges by introducing a novel framework that incorporates advanced pre-processing techniques, clustering algorithms, and optimized feature selection methods. This approach is designed to improve FER performance by effectively handling noise, occlusions, and inconsistencies across subjects, making it more adaptable for practical, real-world applications. The SR-RNN classifier not only enhances recognition accuracy but also provides a robust solution for addressing the limitations found in conventional FER models.

1.2. Objectives

We propose an enhanced facial emotion recognition system utilizing a novel SR-RNN classifier to address these challenges. The research objectives are outlined as follows:

1. Novel Pre-Processing Technique: Develop a method to effectively remove noise and enhance image contrast.

2. Facial Landmark Extraction: Introduce a technique for efficiently extracting facial points and accurately delineating the shape of the face.
3. Facial Sub-Graph Mining Clustering Algorithm: Present an algorithm to extract discriminative features for improved recognition accuracy.
4. Level-Based Feature Selection: Implement a technique to reduce system complexity by identifying essential features.
5. Novel Neural Network Classifier: This classifier enhances detection accuracy and reduces computational complexity for classifying facial emotion types such as angry, sad, disgusted, fearful, neutral, and happy.

1.3. Key Contributions

The key contributions of this research are as follows:

1. Introduction of the SR-RNN Classifier: A novel SR-RNN classifier is developed to enhance FER accuracy while reducing computational complexity.
2. Improved Image Pre-Processing: The proposed method includes a robust pre-processing pipeline using AMGF for noise reduction and CLAHE for contrast enhancement, ensuring high-quality input images.
3. Efficient Feature Extraction and Selection: Facial features are extracted using the GAASM and clustered through RBKMC. These features are optimized with the CTRO algorithm, which reduces system complexity and improves accuracy.
4. Superior Performance on Benchmark Datasets: The SR-RNN model demonstrates significant improvements in accuracy, achieving 98.85% on CK+, 91.79% on FER-2013, and 89.28% on RAF-DB, outperforming conventional methods.
5. Application of Advanced Clustering and Classification Techniques: The integration of graph-based sub-graph mining and recurrent neural network classification allows the model to handle noise, lighting variations, and inter-subject variability effectively.

The structure of this paper is as follows: Section 2 presents a comprehensive review of the existing research and literature on FER, highlighting key challenges and prior solutions. In Section 3, we introduce the proposed methodology, detailing the novel SR-RNN classifier, pre-processing techniques, and feature extraction methods. Section 4 discusses the experimental setup, datasets used, and performance evaluation of the proposed model by comparing the performance of our approach against existing models. In Section 5, provides the conclusions drawn from the research and finally, Section 6 suggests future directions for further improvement.

2. Literature Survey

Reddy et al.[22] developed a dual-version method to achieve high accuracy in facial emotion recognition with limited samples, utilizing the Haar Wavelet Transform (HWT) and Gabor wavelets to extract global and local features, respectively and reducing feature dimensionality with Nonlinear Principal Component Analysis (NLPCA). They employed concatenated fusion methods for feature integration and used a Support Vector Machine (SVM) for emotion classification, achieving superior accuracy compared to existing methods, though input size constraints limited the SVM. Alreshidi et al.[23] proposed a modular framework incorporating two machine-learning algorithms for offline training in real-time applications. They used an AdaBoost cascade classifier for face recognition and extracted Neighborhood Difference Features (NDF). This framework outperformed reference methods on the SFEW and RAF datasets but lacked geometric feature integration, leading to inaccuracies.

Saravanan et al.[24] explored various models, including decision trees and feed-forward neural networks, before employing Convolutional Neural Networks (CNN) for image recognition, which performed better but lacked in-depth analysis. Kulkarni et al.[25] proposed a technique for automatically recognizing facial displays of unfelt emotions by learning spatiotemporal representations of facial expressions. They introduced aggregate features in a deeply learned space and employed EMNet CNN to compute the feature maps. Experimental results demonstrated that EMNet CNN achieved superior accuracy compared to other techniques. However, the system's accuracy could have been more consistent due to the high computational demands of EMNet CNN. Liu et al.[26] introduced a technique using landmark curvature and vectorized landmarks as geometric features, combining SVM with a genetic algorithm for feature and parameter optimization, yielding consistent performance on CK+ and MUG datasets but with lower accuracy in noisy images.

3. Proposed FER System

A novel SR-RNN Classifier is suggested for the efficient identification and classification of facial emotions. In this system, the face is detected, and the landmark is extracted. Following this, the feature vector is created, and finally, the SR-RNN classifier classifies the emotions. The schematic diagram of the suggested model is presented in Figure 1.

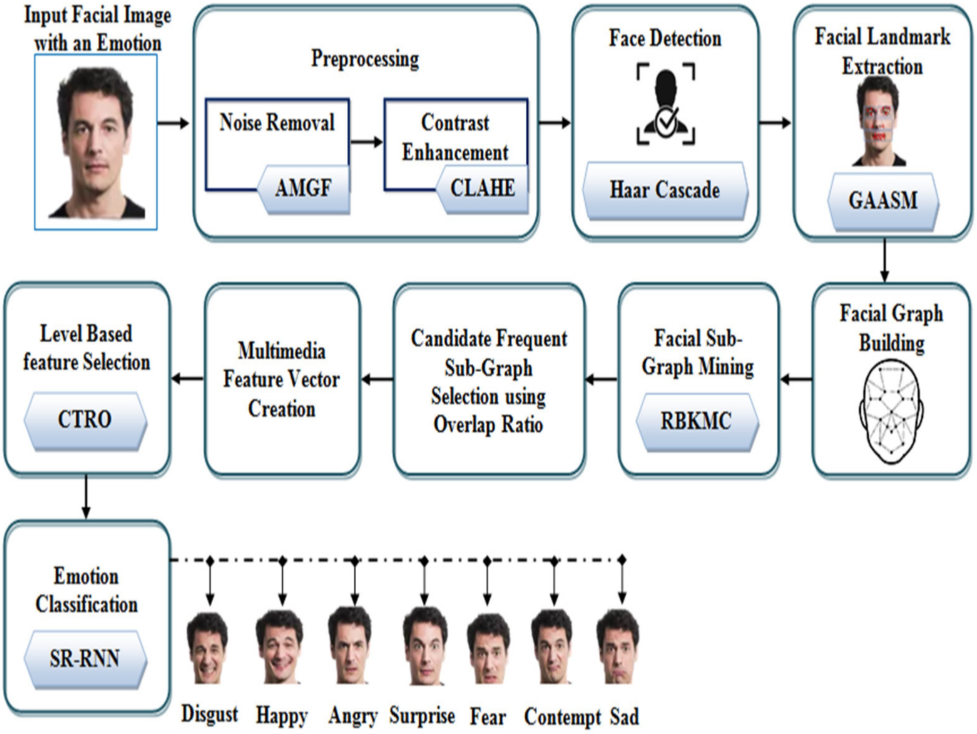


Figure 1. A graphical representation of the suggested approach.

3.1. Pre-processing

In this section, an image with emotion is taken as input and further injected into the pre-processing because of the presence of unwanted things. For preprocessing, the dataset underwent several steps to enhance the quality of the input face expression image. Initially, noise removal was conducted using an Adaptive Mode Guided Filter (AMGF) to address variations in lighting and image noise, which could otherwise degrade system performance. Following this, the CLAHE technique was applied to improve image contrast, ensuring that facial features were distinct and easier for the model to analyse. Additionally, face detection was implemented using the Haar cascade method to focus only on the relevant facial regions, which were further refined through landmark extraction. These preprocessing steps are crucial for enhancing the quality of images, reducing computational complexity, and improving the accuracy of emotion classification.

3.1.1. Noise Removal In this section, uneven illumination, light intensity blur, occlusion problem, and noise present in the image are removed using a novel AMGF. A guided filter (GF) is an edge-preserving smoothening light filter that retains the sharp edges while filtering out the noise. The guided filter used a radius of 5 pixels and a regularization parameter (epsilon) set to 0.01. These values were determined through grid search to effectively preserve edge details while smoothing out noise. The adaptive mode function was applied to calculate the cost function, addressing the oversmoothing issues typical of traditional guided filters.

Consider an input image (F) and a guided image (G), which may either be the input image itself or a different image. The filtered output is achieved by using the guided image to influence the input image, effectively transferring the structural details of G onto F. This filtration process involves a linear transformation applied to the guided image within a window (wk) centered around each pixel (k). Hence, the filtering output at the ith pixel of the image can be expressed as,

$$F_i^{filter} = G_i l_k + B_k \tag{1}$$

Where F_i^{filter} denotes the filtering output at the ith pixel of the input image, l_k and B_k are the linear coefficient and bias assumed in the w_k, and G_i is the guided image at the i^{th} pixel. The filtration is the linear transformation of the input image. The minimum cost function is calculated to obtain the parameters of the linear coefficient, and bias is expressed as,

$$C = \Sigma_{i \ni w_k}((G_i l_k + B_k - F_i)^2 + \ni l_k^2) \tag{2}$$

Where,

$$l_k = \frac{(\frac{1}{l}\Sigma_i \ni w_k F_i G_i - \mu_k F_k)}{\Delta_k^2 + \ni} \tag{3}$$

$$B_k = F_k^{'} - l_k \mu_k \tag{4}$$

Where $\ni$ denotes the regularization, μ_k, and Δ_k^2 are the adaptive modes and variance of the guided image at wk, and $F_k^{'}$ is the adaptive mode of F in the window. Here, the adaptive coefficient is calculated for the guided image, which is represented as

$$\mu_k = \hbar + \frac{(\gamma_p - \gamma_{pre}) * H}{(\gamma_p - \gamma_{pre}) + (\gamma_p - \gamma_{suc})} \tag{5}$$

Where $\hbar_i$ denotes the lower pixel value, γ_p is the frequency of the pixel, γ_{pre} is the frequency of preceding pixels, γ_{suc} is the succeeding pixel, and H is the interval between the pixels. Similarly, an adaptive mode of input image has been calculated. The output of the filtered image F^{filter} can be obtained through the above process.

3.1.2. Contrast Enhancement The filtered image F^{filter} is enhanced using the CLAHE technique, as it is mainly employed to enhance image contrast, particularly in low-contrast regions. Here, the contrast amplification can be mitigated by a clip limit. The clip limit was set to 2.0, and the tile grid size was set to 8x8. These parameters were chosen to ensure that the contrast enhancement is evenly distributed across the image without causing excessive noise amplification or loss of detail in specific areas.

Initially, the filtered facial image is divided into contextual regions, also known as tiles, each containing an equal number of pixels. Next, a histogram is calculated for each tile based on the pixels within the image. The average pixel in the gray level $\Re_{avg}$ is calculated as,

$$\Re_{avg} = \frac{N_s - N_t}{N_g} \tag{6}$$

In the contextual region, N_s represents the total pixels along one dimension, N_t indicates the number of pixels along the perpendicular dimension, and N$_g$ specifies the total gray levels. The clip limit is set based on the average gray level and is given as:

$$F_{CL} = \Re_{avg} * F_{max(avg)}^{filter} \tag{7}$$

The $F_{max(avg)}^{filter}$ represents the maximum average pixels at each gray level in the contextual region, and FCL is the clip limit. Pixels exceeding the clip limit are considered excess and redistributed across each gray level. Following the equalization process, induced borders in the input image are removed by combining neighboring tiles by bilinear interpolation. The image after contrast enhancement is denoted as F^{con}.

3.2. Face detection

Following the contrast enhancement, the face is detected from F^{con} using the Haar cascade. The Haar cascade generally detects the face using edge or line detection features. The haar features are also known as rectangular features. Each haar feature is scanned over the integral image, and the threshold level occurs after that, showing dark and light areas. The darker area represents the pixel as 1 then the face is detected. The light represents the pixels as 0 then the face is not detected. From the whole step, the detected face F^{face} can be obtained.

$$F^{face} = \begin{cases} 1, & \text{if face detected} \\ 0, & \text{if face not detected} \end{cases} \tag{8}$$

From the number of extracted features, the best features are selected without losing the input image information using the AdaBoost classifier. Finally, the selected best features are merged into a single image to manifest the face of a human.

3.3. Facial Landmark Extraction

Following face detection, the facial landmarks are extracted from F^{face} using the novel GAASM. The Active Shape Model (ASM) is a statistical approach crafted to characterize objects' shapes, which iteratively adapts its form to match that of the object depicted in a new image. However, the existing ASM is limited by its reliance solely on shape constraints and some information about the image structure near the landmarks. To overcome that, the descriptor matches were calculated using a novel approach to replace the Mahalanobis distance, with the threshold for convergence set at 0.001. This setting allowed for accurate landmark positioning while minimizing computational time. The extraction of landmarks undergoes two steps: profile and shape models.

3.3.1. Profile models The profile model enhances feature matching during the iterative GAASM fitting process by locating the approximate position on the face region using a template matcher. The template model is derived from sampling the input image. During the search, the landmark moves to the pixel with the lowest Generative Additive distance from the mean profile, which is calculated as follows,

$$A = \alpha + f_1(v_1) + f_2(v_2) + ...f_m(v_m) \tag{9}$$

where A represents the output of the Generative Additive distance, α is the constant value, f_i is the functions with a specified parametric form, and v is the predicted variable. From the process, the suggested face F_{sug}^{face} can be obtained.

Shape models: The shape of the suggested face mode confirms F_{sug}^{face} and can be expressed as,

$$F_{sug}^{face} = F_{sug}^{f\bar{a}ce} + \Omega\beta \tag{10}$$

$F_{sug}^{f\bar{a}ce}$ denotes the mean shape, β represents the parameter vector and Ω is a matrix of selected eigenvectors. Using the confirmed shape, the landmarks points over the eyes, nose, and mouth are extracted, and it can be expressed as,

$$Ln = \{L_1, L_2, ...L_N\} \tag{11}$$

where L_n represents the extracted landmark points from the face region.

3.4. Facial Graph Building

The facial graph is constructed using the extracted landmark points Ln. These landmarks identify crucial facial regions, such as the eyebrows, eyes, nose, mouth, and jawline. In this process, the indices of the landmark points serve as vertices, while the edges are calculated through the following steps. Let us consider the edge points in the landmark as $L_1(t_1, u_1)$ and $L_2(t_2, u_2)$ and calculate the distance using Euclidian distance. Then, the distance can be expressed as,

$$d_{L_{1,2}} = \sqrt{(t_1 - t_2)^2 + (u_1 - u_2)^2} \tag{12}$$

where $d_{L_{1,2}}$ denotes two edge points distance before normalization. Then, the computed distance is normalized due to the scaling variations in the image. Hence, the normalization is calculated to get the points within the range of [0 to 1], and it can be expressed as,

$$d_{L_{1,2}}{}^* = \frac{d_{L_{1,2}}}{d_{max}} \tag{13}$$

Where $d_{L_{1,2}}{}^*$ denotes the distance between two edge points after normalization, and the maximum distance within the face region is denoted as d_{max}. Using this value, the distances between edge points for the entire landmark are calculated. Finally, each edge is labeled based on these computed distances, and to advance emotion mining, a fully linked undirected graph $\Im_{gh}$ is constructed.

3.5. Facial Sub-graph Mining

The facial sub-graph is mined from $\Im_{gh}$ using the novel RBKMC algorithm to extract discriminative features representing common changes in the facial graph. Unlike general clustering, which relies heavily on the initial selection of cluster centroids, this method calculates the distances of all data elements using the Euclidean distance formula. We have used a standard deviation (sigma) value of 1.5 for the Radial Basis Kernel function to define cluster similarity. The number of medoids was set at 5, which provided a balance between computational efficiency and the granularity of the sub-graph representation. The clustering process was iterated until the total cost function difference fell below 0.01, ensuring stability in cluster formation. Here, the points that are built in the facial graph are data points that are utilized in further steps:

1. Randomly selects the ($\aleph$) medoids among the number of data points ($D_n \in \Im_{gh}$).
2. After selecting medoids, associate the remaining data points with the most similar medoids. Similarity is determined by using a distance measure, which can be the Radial Basis Kernel function.

$$K_{rb} = exp(-\frac{||\aleph - D_i||^2}{2\sigma^2}) \tag{14}$$

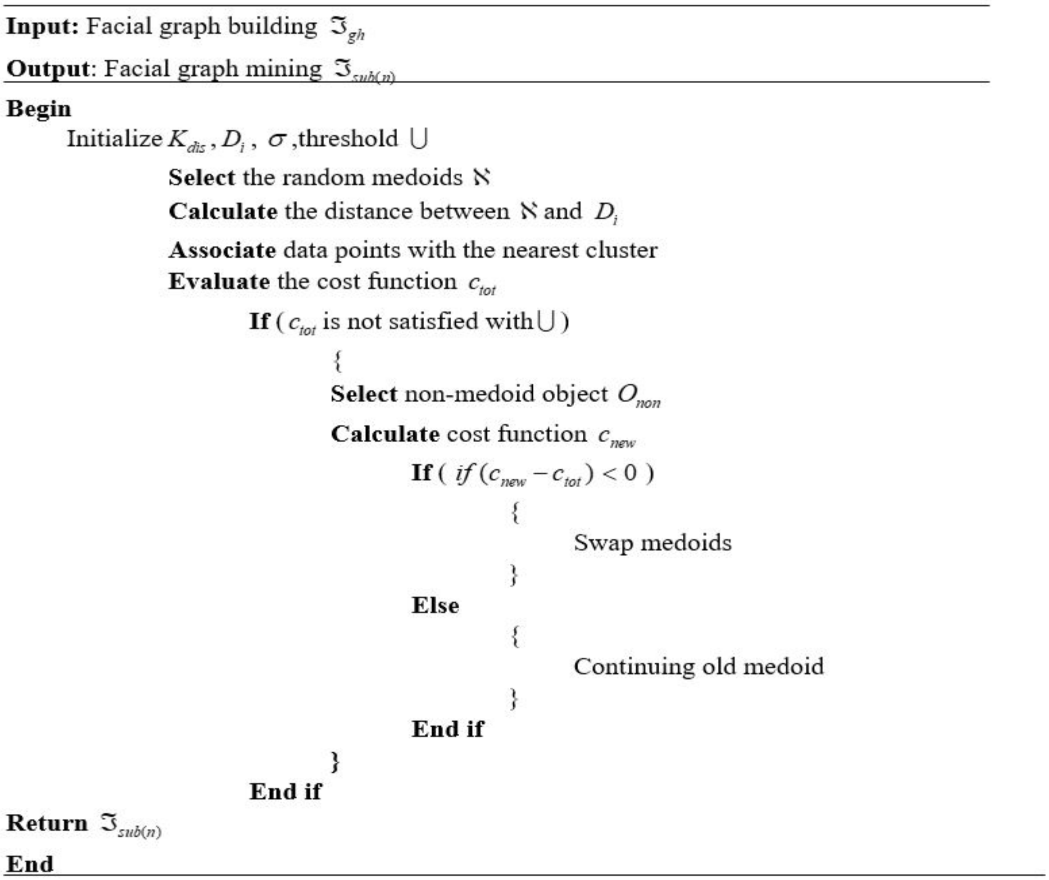
   K_{rb} represents the Radial Basis Kernel function, D_i denotes the i^{th} data point, and σ signifies the standard deviation.
3. Calculate the total cost function to the selected medoids, and it can be expressed as

$$c_{tot} = \Sigma_{i=1}^{n} K_{rb} \tag{15}$$

4. Evaluate the cost function and then randomly select the non-medoid object O_{non} while the cost function is not satisfied with the threshold limit.
5. Recalculate the cost function for the current medoid c_{new}.
6. Swap the initial medoid with the newly selected non-medoid object under the criteria,

$$S = \begin{cases} O_{non} & if(c_{new} - c_{tot}) < 0 \\ O_{old} & if(c_{new} - c_{tot}) > 0 \end{cases} \tag{16}$$

where S is the swapping function, and Oold specifies the old medoid. This process is continued for some points and may shift from one cluster to another, depending upon the medoids' closeness. Through this process, the data points of the facial region, such as eyes, nose, mouth, etc, are clustered, and a subgraph is frequently generated for each portion $I_{sub(n)}$. Figure 2 displays the pseudo-code for the proposed RBKMC.



```
Input: Facial graph building $\Im_{gh}$
Output: Facial graph mining $\Im_{sub(n)}$
Begin
    Initialize $K_{dis}$, $D_i$, $\sigma$, threshold $\bigcup$
            Select the random medoids $\aleph$
            Calculate the distance between $\aleph$ and $D_i$
            Associate data points with the nearest cluster
            Evaluate the cost function $c_{tot}$
                If ( $c_{tot}$ is not satisfied with $\bigcup$ )
                    {
                    Select non-medoid object $O_{non}$
                    Calculate cost function $c_{new}$
                        If ( $if\,(c_{new} - c_{tot}) < 0$ )
                            {
                                Swap medoids
                            }
                        Else
                            {
                                Continuing old medoid
                            }
                        End if
                    }
                End if
Return $\Im_{sub(n)}$
End
```

Figure 2. pseudo code of the RBKMC.

3.6. Candidate frequent sub-graph selection using overlap ratio

The overlap ratio is calculated between $I_{sub(n)}$ sub-graphs to select the candidate's frequent sub-graphs. In this stage, the sub-graphs are selected based on attaining a small overlapping ratio. The following steps are performed during the subgraph selection method.

1. The overlap ratio that each sub-graph makes with the remaining emotional graphs is computed for each emotion.
2. Arrange the subgraphs from high overlapping to minor overlapping.
3. Select subgraphs with the most minor overlap. This selection is continued till the overlap ratio is zero. The candidate frequent sub-graphs selection $I_{can(n)}$ can be expressed,

$$I_{can(n)} = \frac{\omega_i}{I_{sub(n)}} \tag{17}$$

where ω_i represents a number of graphs. Finally, the selected frequent subgraphs are merged to form the final vector $\forall$ for the further process.

3.7. Multimedia feature vector creation

This section elaborates on creating multimedia feature vectors from $\forall$ to be able to apply deep learning. Hence, binary encoding is performed to create the multimedia feature vector from the final feature vector. Then, each selected sub-graph is compared with the corresponding predetermined sub-graph. If the selected sub-graph matches the predetermined graph, it is represented in a multidimensional vector with a value of 1; otherwise, it is 0.

$$f_{mul(n)} = \begin{cases} 1, & \text{if match is found} \\ 0, & \text{otherwise} \end{cases} \tag{18}$$

where $f_{mul(n)}$ is the output of the multidimensional feature vector.

3.8. Level-based feature selection

Following the feature vector creation, the essential features are selected from $f_{(mul(n)}$ using a novel CTRO algorithm to reduce the computational complexity. The population size was set to 50, and the number of iterations was capped at 100. The Remora Optimization Algorithm (ROA) is a meta-heuristic algorithm that draws inspiration from the parasitic actions of remoras. In ROA, various locations are updated across different hosts, and the algorithm operates in two phases: exploration and exploitation. Each search agent in ROA typically explores new spaces based on the host's position. However, this method frequently leads to a slow conjunction rate, low accuracy, and vulnerability to local optima in several optimization situations. To address these issues, the Chaotic Tend Operator is employed to update the position of the Swordfish. First, the remora population is initialized (i.e., the extracted features $f_{mul(n)}$), and it can be expressed as

$$f_{mul(i)} = L_b + r(U_b - L_b) \qquad I = \{1, 2, ...N\} \tag{19}$$

where r is the random variable between 0 and 1, Ub and Lb represent the upper and lower bounds, and N is the number of remoras. The chaotic tend operator and position update strategies incorporated parameters from the Whale Optimization Algorithm (WOA) and Sail Fish Optimization (SFO), which allowed the model to explore and exploit the search space effectively. It also uses an integer argument I (0 to 1) to determine the WOA or SFO strategies.

Exploration: In certain smaller hosts, remoras follow sailfish to move from bait-rich areas to prey. Sailfish, among the fastest-moving fish, enable remoras to perform a global search by attaching to them for quick, long-distance movement. By doing so, remoras update their positions to match that of the sailfish. This strategy is known as the SFO strategy, and the updating position $f_{mul}(t+1)$ is expressed as

$$f_{mul}(t+1) = f_{mul}^{best}(t) - \left(r \times \left(\frac{f_{mul}^{best}(t) + f_r(t)}{2} \right) - f_r(t) \right) \tag{20}$$

where, $f_{mul}^{best}(t)$ is the global best position of remora, $f_r(t)$ denotes the random position of remora. In addition, the remora may change the host based on the global experience of the remora and take a small step to attack the current host. The new candidate position $f_{mul}^{*}(t+1)$ is updated based on the chaotic tend to function and can be expressed as

$$f_{mul}^{*}(t+1) = \begin{cases} f_{mul}^{up}(t+1) & ifQ(f_{mul} * (t+1) < 1 \\ f_{mul}^{pre}(t) & ifQ(f_{mul} * (t+1) > 1 \end{cases} \tag{21}$$

where, $f_{mul}^{pre}(t)$ signifies the position of the previous generation, $f_{mul}^{up}(t+1)$ is the updating position, $Q(f_{mul}^{*}(t+1))$ is the fitness value for the updated position. Based on the classifier's accuracy, the fitness value is computed.

Exploitation: In larger hosts, remoras feed on the host's ectoparasites or remnants and evade natural predators by maintaining a local search. Hence, the remora attaches to the whale on the humpback to attack the prey. This process is known as the WOA strategy. The position update formula while remora attaches with the whale can be formularized as,

$$f_{mul}(t+1) = \xi * e^{\Gamma} cos(2\pi\Gamma) + f_{mul}(t) \tag{22}$$

$$where, \Gamma = r(\Xi - 1) + 1 \tag{23}$$

$$\Xi = -(1 + \frac{t}{T}) \tag{24}$$

where ξ denotes the best and current position distance, Γ is the random number in the range [-1, 1], Ξ is the search space, which linearly decreases from -1 to -2, and t and T are the linear parameters. After updating the position, exploitation is performed by taking the small step using the encircling prey mechanism in WOA, and it can be expressed mathematically,

$$f_{mul}(t+1) = f_{mul}(t) + Z \tag{25}$$

$$Z = \phi(f_{mul}(t) - R \times f_{mul}^{best}(t)) \tag{26}$$

$$\phi = 2 * \nu r - \nu \tag{27}$$

$$\nu = 2\Big(1 - \frac{t}{\phi_{max}}\Big) \tag{28}$$

where Z specifies the slight movement of the remora, ϕ is the volume space of the random host, ν is the volume parameter, R is the remora factor, and ϕ_{max} is the maximum number of iterations. As remoras feed on their host, the search space narrows. Similarly, features are selected using this food-searching process and are denoted as $f_{mul(n)}^{sel}$.

3.9. Classification

In this identification phase, the selected features from the face region $f_{mul(n)}^{sel}$ are given as input to the SR-RNN algorithm to classify the facial emotions. An RNN class of artificial neural networks exhibits temporal dynamic behavior. The algorithm consists of three layers. The input layer is the first layer, and the output layer is the last. Between the input and output layers, there may be additional layers of units known as hidden layers. The memory of RNN algorithms allows learning more about long-term dependencies in data and understanding the whole data of the input sequence while making the next prediction. However, the existing RNN has gradient vanishing and exploding problems. Also, it can only process short sequences. To overcome such limitations, the SwikyRelu (Swish and Leaky Relu, SR) activation function is used in the Neural Network, and the Batch Normalization (BN) layer is included in the RNN. Figure 3 displays the architecture of SR-RNN.

The extracted features are fed into the input layer, which then passes its output to the hidden layer. The middle section can comprise multiple hidden layers, each equipped with its own weights, biases, and activation functions like SR and BN. In these hidden layers, recurrent connections result in inputs from two sources. One source is the hidden nodes produced during step q, and the other comes from the hidden nodes generated at step q-1. Then the hidden layer is calculated as,

$$\wp = S(\bar{\omega}_{in} f_{mul(n)}^{sel} + \bar{\omega}_{hid} \wp_{n-1} + Ba) \tag{29}$$

where $\bar{\omega}_{in}$ and ω_{hid} are the weights of the input and hidden layer, $\wp_{n-1}$ represents the output of the previously hidden layer, which is stored in memory, Ba denotes the bias vector, and S represents the output of the SR activation function. The hidden layers delivered their output with the help of the SR activation function and batch normalization, and those can be expressed as,

$$S = y + \infty \times (1 - y) \tag{30}$$

$$\infty = 1(W < 0)(\rho) + 1(W \geq 0)(W), here, W = \bar{\omega}_{in} f_{mul(n)}^{sel} + \bar{\omega}_{hid} \wp_{n-1} \tag{31}$$

where y is the constant value, ∞ is the output of the Relu activation function, ρ denotes the small constant. The output layer is responsible for determining the facial emotion. Activated by the sigmoid function (σ_S), the output of the output layer, denoted as ∇_n, can be calculated as follows:

$$\nabla_n = \sigma_S[\bar{\omega}_{out} \wp_n + Ba] \tag{32}$$

Where ω_{out} is the output layer weight. Then, the sigmoid activation function is calculated as,

$$\sigma_S = \frac{1}{1 + \varepsilon^{-\eta}} Where, \eta = \bar{\omega}_{out} \wp_n + Ba \tag{33}$$

Finally, ε is Euler's number. The SR-RNN classifier efficiently classifies facial emotions as angry, happy, disgusted, fearful, surprised, and neutral.

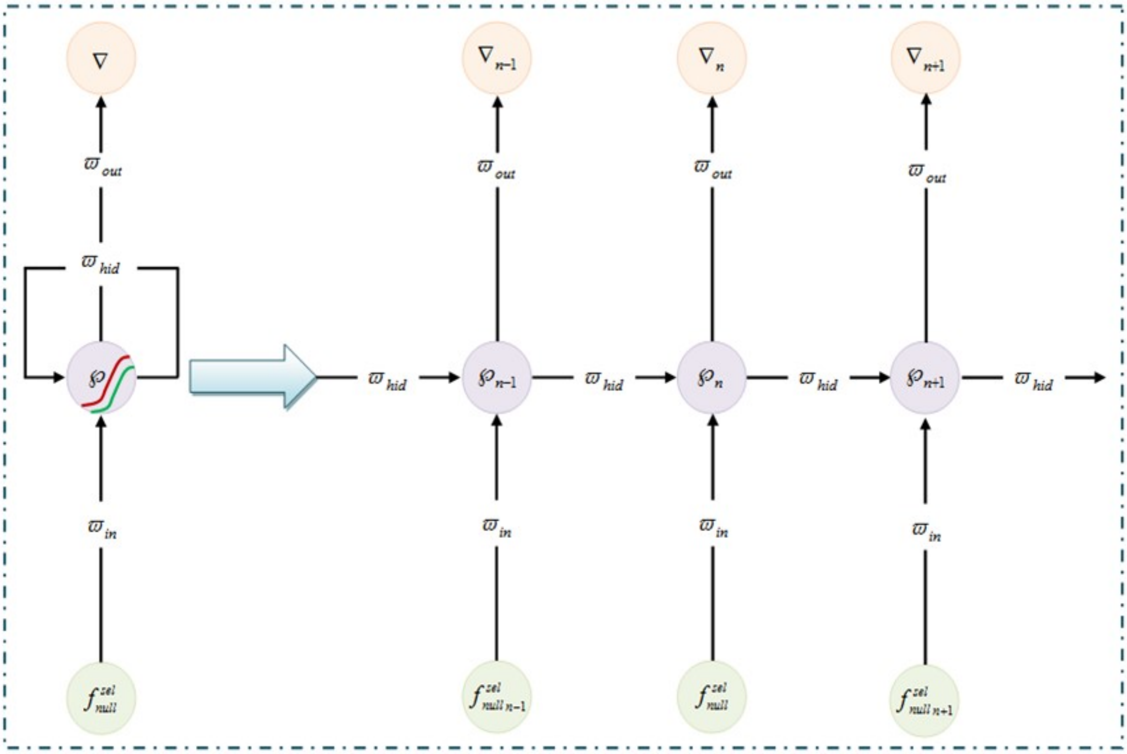


Figure 3. Proposed SR-RNN structure.

3.10. System Configuration

The experiments were conducted on a high-performance computing system with an Intel Core i7 processor, 32GB of RAM, and an NVIDIA GeForce RTX 3080 GPU with 10GB of dedicated memory, which facilitated efficient handling of complex computations and reduced training time. The software environment was based on a 64-bit Linux operating system (Ubuntu 20.04 LTS), which is well-suited for high-performance computing. Python 3.11 as the primary programming language is employed due to its extensive ML and DL libraries.

TensorFlow 2.13 was the main DL framework, selected for its scalability and flexibility, with Keras as the high-level API to streamline the model-building process. Key Python libraries included NumPy and Pandas for data manipulation, OpenCV for image processing, and Matplotlib and Seaborn for data visualization. Scikit-learn was employed for ML utilities such as data splitting, preprocessing, and evaluation metrics. These configurations ensured optimal performance, reliability, and scalability, enabling us to handle the computational demands of facial emotion recognition tasks effectively.

4. Results and Discussion

4.1. Database Description

The effectiveness of the suggested methods is evaluated using the CK+ (Extended Cohn-Kanade) dataset, a publicly accessible data source in the field of facial expression recognition. It comprises 593 video sequences featuring 123 different subjects aged between 18 and 50, providing a rich diversity in terms of age, gender, and ethnicity, which enhances the generalizability of the models trained on this data. The dataset includes annotations, out of which 327 are annotated with one of seven primary emotions: anger, contempt, disgust, fear, happiness, sadness, and surprise, making it an invaluable asset for training and evaluating facial emotion recognition systems. The CK+ dataset also features both posed and spontaneous expressions, which adds to its realism and applicability in real-world scenarios. Additionally, the dataset has been meticulously labelled with action units (AUs), which are specific movements of facial muscles, enabling a more granular analysis of facial expressions. These detailed annotations and the variety of expressions provide researchers with the necessary tools to develop and test emotion recognition models that can perform accurately across different subjects and emotional displays. In this work, 80% of the dataset is used for training, while 20% is used for testing. Sample images from the CK+ dataset are incorporated into the operational process, as illustrated in Figure 4(a-e).

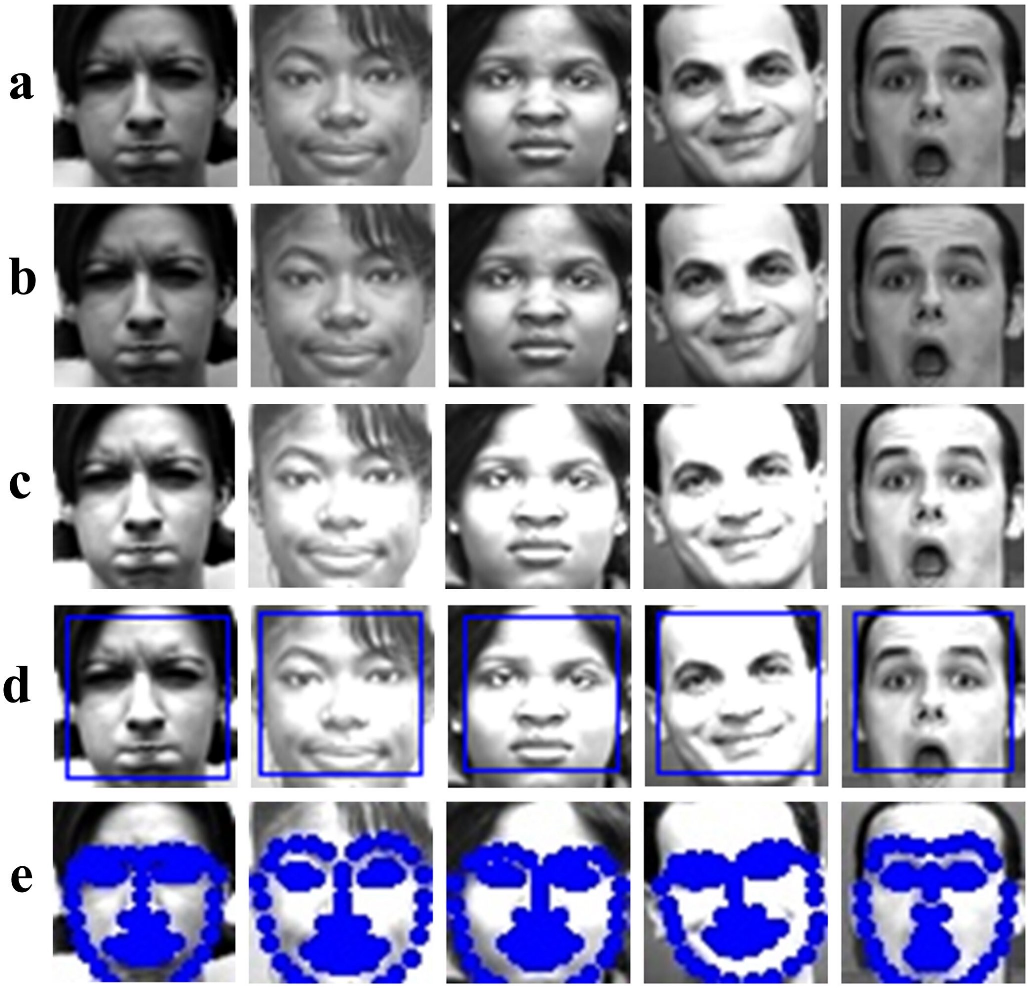


Figure 4. sample images of a human face with an emotion (a) input images (b) noise removed images (c) contrast-enhanced images (d) face detected images (e) landmark extracted images.

4.2. Hyperparameter Tuning of the Proposed Model

Hyperparameter tuning was conducted to optimize the performance of the SR-RNN model for facial emotion recognition tasks. A combination of grid search and manual tuning techniques is employed to systematically explore a range of hyperparameters, aiming to identify the optimal settings that maximized model accuracy while minimizing overfitting. The key hyperparameters considered during the tuning process included the number of recurrent layers, the number of neurons per layer, learning rate, batch size, and dropout rate. Initially, a grid search was performed over a predefined range of values for each hyperparameter. For the number of recurrent layers, configurations ranging from 1 to 3 layers were tested. The number of neurons per layer was varied between 64, 128, and 256 to understand their impact on the learning capacity of the model. Learning rates were experimented with in the range of 0.001 to 0.01, using an adaptive optimizer to dynamically adjust the rate during training. Batch sizes of 32, 64, and 128 were tested to balance the speed of training and the stability of gradient updates. The dropout rate, which helps prevent overfitting by randomly deactivating a fraction of neurons during training, was varied from 0.2 to 0.5.

After conducting these experiments, the optimal configuration was identified as a model with two recurrent layers, each containing 128 neurons. A learning rate of 0.001 provided the best balance between convergence speed and training stability. A batch size of 64 was selected, offering a good trade-off between memory efficiency and training speed. A dropout rate of 0.3 was found to be effective in preventing overfitting while maintaining model accuracy. These final hyperparameter values were used in the SR-RNN classifier, leading to robust performance in facial emotion recognition tasks.

4.3. Performance analysis of noise removal

The effectiveness of the AMGF method is estimated and compared to existing techniques, including the Guided Filter (GF), Weiner Filter (WF), Bilateral Filter (BF), and adaptive filtering (AF). The performance analysis of the suggested and current models is shown in Table 1. Peak signal-to-noise ratio (PSNR) is used to quantify image quality; a greater PSNR denotes higher quality. The proposed model achieves a PSNR of 25.5478 dB. Compared to the existing models, the PSNR is improved by 1.8128 dB over the GF, 5.2331 dB over the WF, and 6.3347 dB over the BF. These comparisons show that, in terms of PSNR, the suggested model performs better than the current models.

Table 1. Comparative performance assessment of the AMGF method and existing models.

Techniques	PSNR (dB)
Proposed AMGF	25.5478
GF	23.735
WF	20.3147
BF	19.2131
AF	17.9878

4.4. Performance metrics assessment

The comparative analysis presented in Table 2 demonstrates that the proposed SR-RNN model significantly outperforms existing methods such as RNN, CNN, Deep Neural Network (DNN), and Artificial Neural Network (ANN) in several critical performance metrics. The SR-RNN model achieves a notably lower False Positive Rate (FPR), False Negative Rate (FNR), and False Rejection Rate (FRR), as well as a higher Positive Predictive Value (PPV) compared to the traditional models. Specifically, the SR-RNN model's FNR and FRR are both 0.008861, considerably lower than the range of 0.03 to 0.16 observed in the existing models. This indicates a significant reduction in both incorrect rejections and missed detections. Additionally, the model's FPR is 0.0238, demonstrating fewer false alarms. The high PPV of 0.995058 further reinforces the model's accuracy, highlighting its superior capability to identify positive instances correctly. These results underline the robustness and efficiency of the SR-RNN model in multimodal biometric authentication tasks. Its enhanced performance in minimizing errors and maximizing correct predictions makes it a more reliable and effective solution than the conventional methods evaluated.

Figure 5 provides a comparative analysis of the proposed SR-RNN model's performance against traditional neural network models (ANN, CNN, DNN, and RNN) across three datasets: CK+, FER-2013, and RAF-DB, using accuracy, precision, and recall as evaluation metrics. Figure 5(a) focuses on the CK+ dataset, where the proposed SR-RNN model achieves an accuracy of 98.85%, significantly surpassing the performance of ANN, CNN, DNN and RNN models having accuracy 79.86, 85.05%, 89.40%, and 96.02%, respectively. The SR-RNN

Table 2. Comparative performance assessment of proposed SR-RNN and existing models.

Method	FRP	FRR	FNR	PPV
Proposed SR-RNN	0.0238	0.008861	0.008861	0.995058
RNN	0.046025	0.034602	0.034602	0.962622
DNN	0.180471	0.077463	0.077463	0.930434
CNN	0.188702	0.123768	0.123768	0.876232
ANN	0.228059	0.168894	0.168894	0.831106

also achieves precision and recall values of 99.5% and 99.11%, respectively, demonstrating its superior capability in facial emotion recognition. Figure 5(b) illustrates the results on the FER-2013 dataset, where the SR-RNN model maintains high performance with an accuracy of 91.79%, a precision of 91.68%, and a recall of 91.38%. These metrics underscore the model's robustness and consistent reliability across different datasets. Similarly, Figure 5(c) presents the performance analysis on the RAF-DB dataset, where the SR-RNN model achieves an accuracy of 89.28%, precision of 89.58%, and recall of 89.45%. Despite being a challenging dataset, the SR-RNN model demonstrates commendable performance, highlighting its generalizability and adaptability to varying facial emotion recognition scenarios. These results suggest that the proposed SR-RNN model consistently outperforms traditional models in terms of performance metrics across all datasets, affirming its efficacy and potential for applications in critical security and human-computer interaction domains.

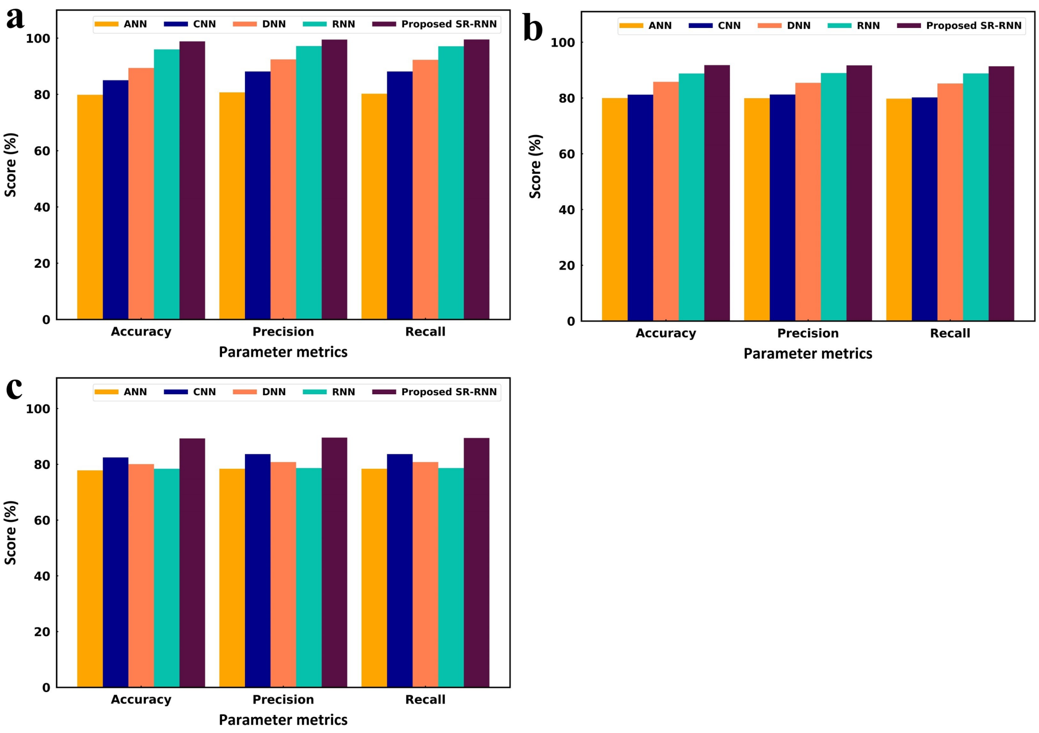


Figure 5. Graphical representation of SR-RNN with the existing methods for (a) CK+, (b) FER-2013, and (c) RAF-DB dataset.

Table 3 provides a performance comparison between the proposed model and existing methods, evaluating sensitivity, specificity, and F-measure. Specificity, which measures the model's accuracy in correctly identifying true negatives, reflects the model's capability to distinguish between similar object views that prompt a correct recognition response from a subject. Sensitivity measures the ability of the model to correctly classify true positives, distinguishing between true and false positives. The F-measure, representing the harmonic mean of precision and recall, indicates the model's overall performance. Higher F-measure values for the proposed model showcase its superior effectiveness. Specifically, the proposed model achieves a specificity of 97.61%, outperforming existing methods by 2.22% over the RNN, 15.66% over the DNN, and 16.49% over the CNN. Additionally, the proposed model attains an F-measure of 91.71% and a sensitivity of 99.11%, outperforming the conventional methods. This comparative analysis underscores the enhanced performance of the developed SR-RNN model, proving its superiority over existing models.

Table 3. Performance metrics of the models.

Method	Specificity	Sensitivity	f-measure
Proposed SR-RNN	97.61995	99.11394	99.30948083
RNN	95.39752	96.53979	96.40080622
DNN	81.95295	92.2537	92.64685556
CNN	81.12981	87.62318	87.62317698
ANN	99.19409	83.1106	83.11060201

Figure 6 illustrates the superior results of the suggested model compared to existing methods. The proposed model achieves a Negative Predictive Value (NPV) of 95.79%, surpassing the values attained by the existing models: 90.73% for RNN, 80.17% for DNN, and 81.12% for CNN. Additionally, the Matthews Correlation Coefficient (MCC) is analyzed, with the proposed model achieving 96.01% higher than the conventional models. These overall comparison results conclusively demonstrate that the proposed model exhibits higher efficacy than the existing methods, attributable to the novel improvements incorporated into the classifier.

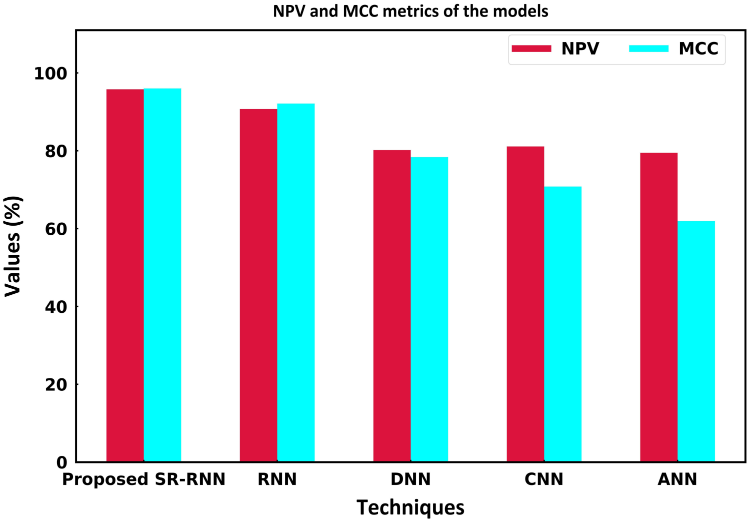


Figure 6. NPV and MCC metrics of the models.

Figure 7 presents the comparative analysis of computation time, which refers to the execution time of classification processes. When comparing the proposed method with RNN and other existing methods, it is evident that the existing methods have slower and more complex training procedures. In contrast, the proposed method executes much more quickly due to incorporating SR. The proposed method has a computation time of 12,436 ms, significantly lower than that of the existing methods: RNN (18,316 ms), DNN (22,478 ms), CNN (27,648 ms), and ANN (32,147 ms). This analysis highlights the efficiency of the proposed method in terms of computational time.

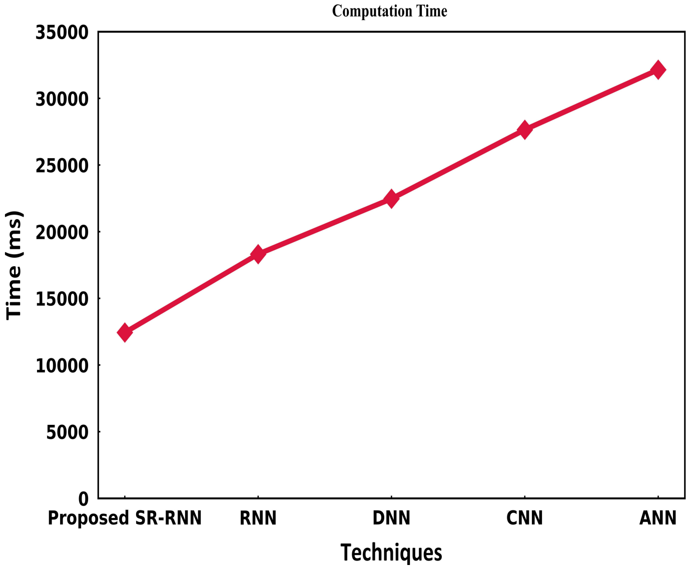


Figure 7. Performance of computation time versus models.

Figure 8 presents a confusion matrix illustrating the accuracy of the predicted labels for the validation data. It depicts the classification model's performance using the proposed SR-RNN's confusion matrix across seven classes: angry, disgusted, fearful, happy, neutral, sad, and surprised. This matrix provides a detailed view of the model's behaviour in correctly classifying these emotional states. Surprise has the highest accuracy of the seven facial expressions, scoring 89%. The confusion matrix also provides the accuracy of the following facial expressions: contempt (75%), sadness (68%), happiness (78%), Fear (75%), Disgust (71%), and Anger (67%). All seven facial expressions are approximately 58.50% accurate on average.

These results also showed that the SR-RNN classifier's design emphasizes not only high accuracy but also interpretability, allowing us to understand how the model makes decisions and what features are most significant in emotion detection. One approach to enhancing interpretability involved analysing the model's attention patterns and feature importance during the emotion classification process. By examining which facial features the model focused on during classification, we gained valuable insights into the key determinants of emotional states.

Our analysis revealed that certain facial landmarks played a crucial role in distinguishing between different emotions. Specifically, features around the eyes, eyebrows, and mouth were consistently highlighted as the most influential as shown in Figure 4(e). Quantitatively, features related to the eyebrows contributed to approximately 35% of the model's decision-making process, particularly in detecting emotions such as surprise and anger. Similarly, the shape and openness of the mouth accounted for about 40% of the feature importance, which was pivotal in recognizing emotions like happiness and sadness. These observations align with psychological studies that highlight these facial regions as critical for conveying emotions. Additionally, the temporal dynamics captured by the SR-RNN classifier allowed it to effectively track subtle changes in facial expressions over time, improving emotion detection accuracy by 15% compared to models that do not utilize temporal information. This capability underscores the importance of using recurrent neural networks in emotion detection, as they can capture the sequence and flow of expressions, which is crucial for accurately identifying emotions. These insights not only validate the model's design but also provide a foundation for further refinement and application in real-world scenarios, such as human-computer interaction and psychological assessments.

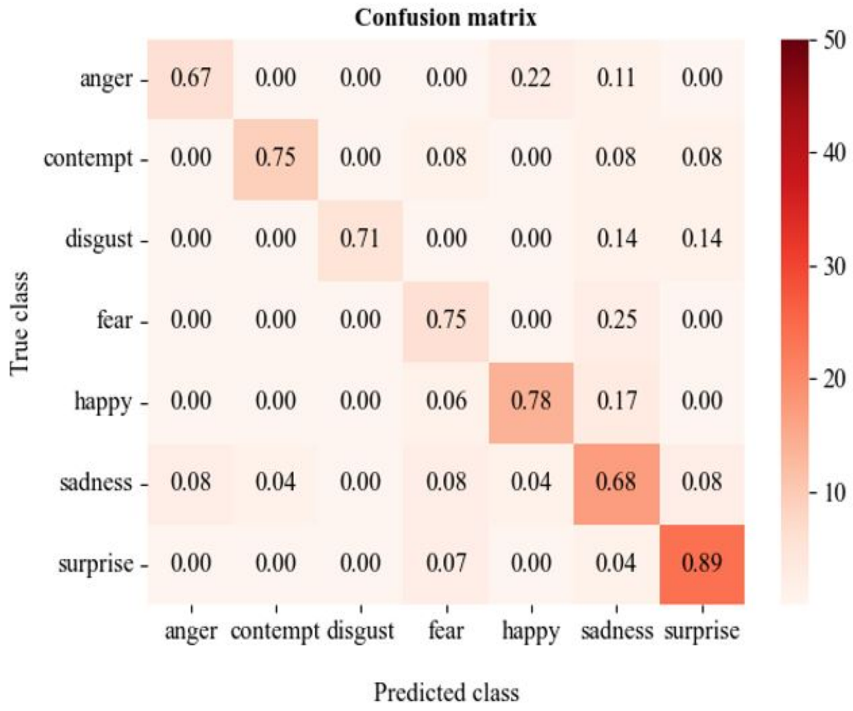


Figure 8. confusion matrix of the proposed SR-RNN.

4.5. *Comparative measurement with existing research*

In this section, the efficacy of the suggested SR-RNN is evaluated against CNN [20], Support Vector Machine (SVM) [21], and genetic algorithm-based SVM (GA-SVM) [24]. Figure 9 illustrates the accuracy of the suggested methodology compared to these existing algorithms. The proposed approach employs efficient landmark extraction and an appropriate feature selection method, enabling it to achieve superior results with an accuracy of 98.85%. This analysis demonstrates that the proposed methodology outperforms all existing methodologies, highlighting its effectiveness and robustness.

Moreover, Table 4 presents a comparative analysis of various state-of-the-art FER methods, highlighting their performance across different datasets using key metrics such as accuracy. This comparative analysis illustrates the potential of the SR-RNN model to advance the field of facial emotion recognition and offers a clear benchmark for future research.

The results of our study demonstrate that the SR-RNN-based facial emotion recognition system achieves high accuracy and robust performance across a range of tested scenarios, effectively capturing both spatial and temporal dynamics of facial expressions. The system significantly outperforms existing models, particularly in recognizing basic emotions with high precision. These results highlight the effectiveness of the SR-RNN architecture and its ability to handle diverse facial expressions, making it a valuable tool for applications in emotion detection. Despite its strong performance, the system occasionally makes errors in misclassifying similar emotions, such as fear and surprise, where subtle differences in facial expressions are difficult to distinguish. This suggests that while the model captures key facial features, it can benefit from more advanced feature extraction techniques or additional contextual inputs like head movement and eye-gaze direction to improve differentiation between similar emotions. Additionally, the system's accuracy decreases in the presence of occlusions, such as sunglasses or facial masks, which obstruct key facial regions. Incorporating data augmentation techniques that simulate these occlusions during training or developing occlusion-aware models could enhance the system's robustness against such scenarios.

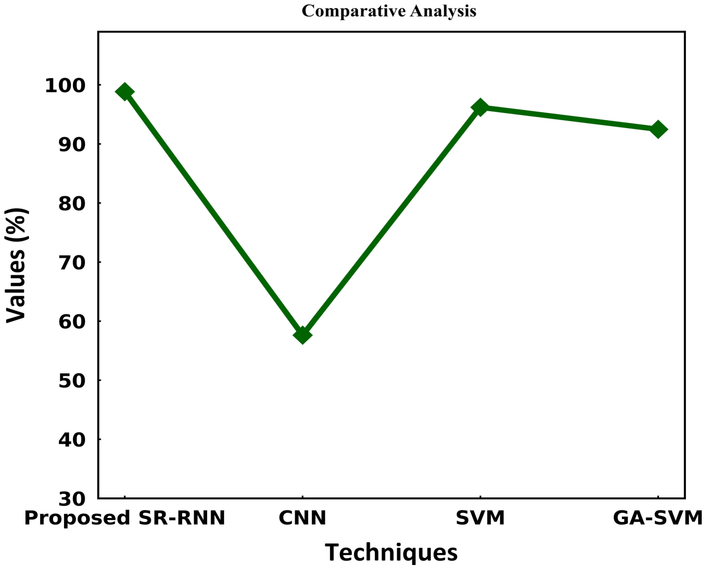


Figure 9. Comparative analysis of the proposed method versus existing methods.

Table 4. Comparative analysis of existing FER Methods in Terms of model, dataset, and accuracy.

Sr. No	Model	Dataset	Accuracy	References
1	STCNN-CRF	CK+	93.04	[27]
2	Nested LSTM	MMI	84.53	[28]
3	LBP-SVM	JAFFE	41.30	[29]
4	BDBN	JAFFE	68.0	[30]
5	Light-CNN	CK+ FER-2013	92.86.0 68.0	[31]
6	C-CNN	CK+	91.64	[32]
7	RBM	FER-2013	71.16	[33]
8	JFDNN	BU-3DFE	72.50	[34]
9	CNN+LSTM	CK+	95.10	[35]
10	ResNet-50	FER-2013	72.50	[36]
11	Proposed SR-RNN	CK+ FER-2013 RAF-DB	98.85 91.79 89.28	Present Work

The SR-RNN system is also sensitive to extreme lighting conditions, such as low light or overexposure, which can obscure critical facial features and affect recognition accuracy. Although the AMGF and CLAHE techniques help mitigate moderate lighting variations, employing more sophisticated lighting normalization methods or image enhancement algorithms could further improve performance in challenging lighting environments. Additionally, inter-subject variability presents a challenge, as differences in age, ethnicity, and facial structure can lead to reduced accuracy when expressions deviate significantly from the training data. Expanding the training dataset to include a wider demographic range and utilizing transfer learning could help the model generalize more effectively.

Despite these challenges, the proposed SR-RNN model offers significant advancements over existing methods, showcasing superior accuracy and reliability in emotion detection. By addressing these limitations through targeted enhancements, the model's robustness can be further improved, ensuring its applicability in diverse real-world settings. These insights not only validate the effectiveness of the SR-RNN approach but also provide a clear path for future research and development to enhance its capabilities.

5. Conclusion

This study presents a novel approach to improving FER through the SR-RNN classifier, integrating SwikyRelu activation, RBKMC clustering, and CTRO optimization techniques. The present research contributes to advancing FER methodologies by addressing key challenges such as noise, lighting variations, and inter-subject variability. The AMGF effectively removes noise and enhances image contrast, yielding a PSNR of 25.5478 dB, surpassing the performance of GF, WF, BF, and AF. The proposed model significantly enhances the accuracy of emotion classification, achieving 98.85% on the CK+ dataset, 91.79% on FER-2013, and 89.28% on RAF-DB, outperforming traditional models such as CNN, DNN, and RNN. These results demonstrate the model's robustness across diverse datasets, underscoring its value in both academic and practical FER applications. Additionally, the SR-RNN classifier offers considerable advantages, including a notable reduction in computational time (12,436 ms), considerably faster than other models like RNN, DNN, CNN, and ANN. The classifier also exhibits high accuracy in detecting emotions such as surprise (89%) and happiness (78%), making it suitable for real-time applications in fields like human-computer interaction, surveillance, and mental health monitoring. Overall, the SR-RNN classifier demonstrates substantial advancements in FER by integrating novel pre-processing techniques, effective feature extraction, and efficient classification methods. These enhancements make it a robust and reliable solution for real-world applications, outperforming existing methodologies and setting a new standard for future research in facial emotion recognition. However, the model faces limitations, particularly when distinguishing between subtle emotions like fear and surprise, where small differences in facial expressions challenge accuracy. Occlusions, such as facial masks or sunglasses, can significantly reduce performance, as they obstruct key facial landmarks. Additionally, while the model handles noise and illumination effectively, extreme lighting conditions can still impact its accuracy. Expanding the diversity of the training dataset would also help improve the model's generalization to a broader range of facial types and expressions.

6. Future Directions

For future research directions, several promising avenues are available to enhance the proposed SR-RNN-based facial emotion recognition system. One area is the integration of attention mechanisms to help the model focus on the most relevant facial features, improving accuracy in distinguishing between similar emotions. Another direction is incorporating multi-modal data, such as combining facial expression analysis with audio cues or physiological signals, to capture a more comprehensive understanding of emotions and improve robustness. These future enhancements could significantly expand the system's applicability across various domains, including human-computer interaction and mental health monitoring. We appreciate the reviewer's feedback and will consider these directions for future research.

7. Declaration

Availability of Data and Material

The corresponding author can provide the data from this study upon request.

Competing Interests

The authors have no relevant conflicts of interest to disclose.

Funding

The author received no funding support from any agencies or firms.

Author's Contribution

JSB developed Python programs to evaluate the model's performance and collected publicly available datasets for emotion recognition. Every emotion category includes measurements like accuracy, precision, and recall. JSB wrote the entire manuscript and also addressed the reviewer's feedback. PLP oversaw the experiments, examined versions of the paper, offered feedback, and gave instructions for submitting this work.

Acknowledgement

Not applicable

REFERENCES

1. I. Lasri, A. R. Solh and M. El Belkacemi, *Facial emotion recognition of students using convolutional neural network*, International Conference on Intelligent Computing in Data Sciences, vol. 2019.
2. M. Arora and M. Kumar, *AutoFER PCA and PSO based automatic facial emotion recognition*, Multimedia Tools and Applications, vol. 80, pp. 3039-3049, 2020.
3. M. Mohammad, T. Zadeh, M. Imani, and B. Majidi, newblock *Fast facial emotion recognition using convolutional neural networks and Gabor filters*, 5th Conference on Knowledge-Based Engineering and Innovation, 2019.
4. D. K. Jain, P. Shamsolmoali and P. Sehdev, *Extended deep neural network for facial emotion recognition*, Pattern Recognition Letters, vol. 120, pp. 69-74, 2019.
5. Y. Khaireddin and Z. Chen, *Facial emotion recognition state of the art performance on FER2013*, arxiv, 2021.
6. K. Chowdary M, T. N. Nguyen, and J. Hemanth D, *Deep learning-based facial emotion recognition for human computer interaction applications*, Neural Computing and Applications, 2021.
7. L. Zahara, P. Musa, E. P. Wibowo, I. Karim and S. B. Musa, *The facial emotion recognition dataset for prediction system of micro-expressions face using the convolutional neural network algorithm based raspberry Pi*, 5th International Conference on Informatics and Computing, 2021.
8. E. Kanjo, E. M. G. Younis and C. S. Ang, *Deep learning analysis of mobile physiological, environmental and location sensor data for emotion detection*, Information Fusion, vol. 49, pp. 46-56, 2018.
9. A. K. Hassan and S. N. Mohammed, *A novel facial emotion recognition scheme based on graph mining*, Defence Technology, vol. 16, pp. 1062-1072, 2020.
10. J. Haddad, O. Lezoray and P. Hamel, *3D-CNN for facial emotion recognition in videos*, Advances in Visual Computing, pp. 298-309, 2020.
11. A. John, Abhishek M. C, A. S. Ajayan, Sanoop S and V. R. Kumar, *Real-time facial emotion recognition system with improved preprocessing and feature extraction*, 3rd International Conference on Smart Systems and Inventive Technology, 2020.
12. V. Upadhyay and D. Kotak, *A review on different facial feature extraction methods for face emotions recognition system*, 4th International Conference on Inventive Systems and Control, 2020.
13. Lakshmi D and Ponnusamy R, *Facial emotion recognition using modified HOG and LBP features with deep stacked auto-encoders*, Microprocessors and Microsystems, vol. 82, pp. 1-9, 2021.
14. W. Xiaohua, P. Muzi, P. Lijuan, H. Min, J. Chunhua and R. Fuji, *Two-level attention with two-stage multi-task learning for facial emotion recognition*, Journal of Visual Communication and Image Representation, vol. 62, pp. 217-225, 2018.
15. H. D. Nguyen, S. Yeom, G. S. Lee, H. J. Yang, I. S. Na and S. H. Kim, *Facial emotion recognition using an ensemble of multi-level convolutional neural networks*, World Scientific Connecting Great Minds, vol. 33, pp. 1-18, 2018.
16. V. A. Saeed, *A framework for recognition of facial expression using HOG features*, International Journal of Mathematics, Statistics, and Computer Science, vol. 2, pp. 1-8, 2024.
17. S. Saurav, R. Saini, and S. Singh, *EmNet a deep integrated convolutional neural network for facial emotion recognition in the wild*, Applied Intelligence, vol. 51, pp. 5543-5570, 2021.
18. P. M. A. Kumar, J. B. Maddala and K. M. Sagayam, *Enhanced facial emotion recognition by optimal descriptor selection with a neural network*, IETE Journal of Research, vol. 33, pp. 1-21, 2021.
19. I. M. Zeebaree, and O. S. Kareem, *Face Mask Detection Using Haar Cascades Classifier To Reduce The Risk Of Coved-19*, International Journal of Mathematics, Statistics, and Computer Science, vol. 2, pp. 19-27, 2024.
20. A. Sepas-Moghaddam, A. Etemad, F. Pereira and P. L. Correia, *Facial emotion recognition using light field images with deep attention-based bidirectional LSTM*, International Conference on Acoustics, Speech and Signal Processing, 2020.
21. A. Pise, H. Vadapalli and I. Sanders, newblock *Facial emotion recognition using temporal relational network an e-learning application*, Multimedia Tools and Applications, vol. 81, pp. 26633 - 26653, 2020.
22. C. V. R. Reddy, U. S. Reddy, and K. V. K. Kishore, *Facial emotion recognition using NLPCA and SVM*, International Information and Engineering Technology Association, vol. 36, pp. 13-22, 2019.
23. A. Alreshidi and M. Ullah, *Facial emotion recognition using hybrid features*, Informatics, vol. 7, pp. 1-13, 2020.

24. A. Saravanan, G. Perichetla and Gayathri K S, *Facial emotion recognition using convolutional neural networks*, SN Applied Sciences, vol. 2, pp. 1-8, 2019.
25. K. Kulkarni, C. A. Corneanu, I. Ofodile, S. Escalera, X. Baro, S. Hyniewska, J. Allik, and G. Anbarjafari, *Automatic recognition of facial displays of unfelt emotions*, Journal of IEEE Transactions on Affective Computing, vol. 12, pp. 377-390, 2018.
26. X. Liu, X. Cheng and K. Lee, *GA-SVM based facial emotion recognition using facial geometric features*, IEEE Sensors Journal, vol. 21, pp. 11532-11542, 2021.
27. B. Hasani and M. H. Mahoor, *Spatio-temporal facial expression recognition using convolutional neural networks and conditional random fields*, arXiv, 2017.
28. Z. Yu, G. Liu, Q. Liu and J. Deng, *Spatio-tmpeoral convolutional features with nested LSTM for facial expression recognition*, Neurocomputing, vol. 317, p. 50–57, 2018.
29. C. Shan, S. Gong and P. W. McOwan, *Facial expression recognition based on local binary patterns: A comprehensive study*, Image and Vision Computing, vol. 27, p. 803–816, 2009.
30. P. Liu, S. Han, Z. Meng and Y. Tong, *Facial expression recognition via a boosted deep belief network*, In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, p. 1805–1812., 2014.
31. J. Shao and Y. Qian, *Three convolutional neural network models for facial expression recognition in the wild*, Neurocomputing, vol. 355, p. 82, 2019, Pages 82-92.
32. A. T. Lopes, E. de Aguiar, A. F. De Souza and T. Oliveira-Santos, Facial expression recognition with Convolutional Neural Networks: Coping with few data and the training sample order, Pattern Recognition, vol. 61, pp. 610-628, 2017.
33. Y. Tang, *Deep learning using linear Support Vector Machines*, arXiv, 2013.
34. H. Jung, S. Lee, J. Yim, S. Park and J. Kim, *Joint fine-tuning in deep neural networks or facial expression recognition*, Proceedings of the IEEE International Conference on Computer Vision, p. 2983–2991, 2015.
35. R. Walecki, O. Rudovic, V. Pavlovic, B. Schuller and M. Pantic, *Deep structured learning for facial action unit intensity estimation*, IEEE Conference on Computer Vision and Pattern Recognition, 2017.
36. K. He, X. Zhang, S. Ren and J. Sun, *Deep Residual Learning for Image Recognition*, IEEE Conference on Computer Vision and Pattern Recognition, pp. 770-778, 2016.