AN ACCURATE FACIAL EMOTION RECOGNITION USING A NOVEL SR-RNN CLASSIFIER

Jyoti Suresh Bedre,
Research scholar at Computer Science & Engineering department, KL Deemed to be University Vaddeswaram AP, India.

P. Laxmi Prasanna,
Associate professor, Computer Science & Engineering department, KL Deemed to be University Vaddeswaram AP, India

Abstract: Facial Expression Recognition (FER) is one of the most powerful, natural, and immediate means for human beings to communicate their emotions and intentions. Recently, automatic human emotion recognition has received much attention. Due to the fast advancement of artificial intelligence (AI) and machine learning, its application is actively being used in many domains. Most of the state-of-the-art works use those methodologies to attain better performance in facial emotion recognition. In the existing methods, the face images captured are highly susceptible to illumination differences and image noise. These outliers impact the performance of FER systems drastically. To solve that problem this research methodology proposed facial emotion recognition using a novel SR-RNN classification. Firstly, pre-processing is carried out and the face is detected using the Haar cascade**.** From this, the landmark features are extracted using the Generative Additive Active Shape Model (GAASM), and the facial graph and subgraph are built using Radial Basis K –Medoids clustering (RBKMC) algorithm. Following this, a candidate frequent sub-graph is selected based on the overlap ratio and a multimedia feature vector is created. From the created feature vector, important features are selected using the Chaotic tend Remora optimization (CTRO) Algorithm. Finally, the selected features are given to the SR-RNN classifier for classifying the emotions. The experimental results show that the method obtains better results compared to existing methods.
Keywords: *Adaptive Mode Guided Filter (AMGF), Contrast Limited Adaptive Histogram Equalization (CLAHE) technique,Haar cascade, Generative Additive Active Shape Model (GAASM), Radial Basis K –Medoids clustering (RBKMC) algorithm, Chaotic tend Remora optimization (CTRO) Algorithm, SwikyRelu Recurrent Neural Network (SR-RNN) algorithm.*

1. Introduction:

The face is the most expressive and communicative part of a human being [1]. Establishing communication between people is through emotions and facial expressions. Nowadays, with the influence of computers on human lives and the mechanization of the lives of individuals, the establishment of human and computer interaction (HCI) has played a crucial and very important role[2]. There is now a strong interest in improving the interaction between humans and computers. Many people believe in this theory and there is a positive and useful emotional response to establishing a good and useful cognitive link between computers with users [3]. So, facial expression is one of the most important features of human emotion recognition [4]. FER refers to identifying expressions that convey basic emotions [5]. Recognition of facial emotions is useful in so many tasks such as customer satisfaction identification, criminal justice systems, e-learning, security monitoring, social robots, smart card applications, etc [6,7]. In addition to that, emotion recognition plays a key role in determining various mental health conditions by psychiatrists and psychologists [8]. Hence the notion of emotion recognition is an active research topic in pattern recognition and artificial intelligence fields [9]. There are six basic facial expressions (anger, disgust, fear, happiness, sadness, and surprise) that have been identified as basic emotional expressions that are universal and common among human beings [10].

Initially, emotions were detected using the semantic and syntactic properties of a language. However, this faced many problems like messages were misinterpreted and the usage of a language differed among different groups of people. Hence Facial emotion recognition (FER) approaches were later used and became very popular [11]. In general, the FER system is divided into three major parts, such as pre-processing, facial feature extraction, and emotion classification [12]. Among these, feature extraction plays a vital role in FER. Feature extraction methods are categorized into two groups, namely, geometric-based and appearance-based feature extraction [13].

Due to the influence of many factors, such as different subjects, races, illumination, complex background, and so on, facial emotion analysis is an indubitable challenging task [14]. Huge challenges in facial expression recognition are finding and understanding facial interesting regions robustly [15]. Consequently, in the last few decades, several FER techniques developed in the literature to tackle these limitations with efficient classifiers [16] such as Deep Boltzmann Network (DBN), Convolution Neural Network (CNN), and Multi-Layer Neural Network (MNN) [17]. Moreover, to further boost the recognition performance, the spatio-angular features can also be learned by selectively focusing on the most important spatio-angular features corresponding to certain viewpoints [18, 19]. However, these approaches are failed to provide an accurate emotion classification and also take more time for classification. So, to obtain an accurate result, this paper proposed an enhanced face emotion identification and classification of the human using a novel SR-RNN classifier based on the RBKMC clustering graph mining technique.

1.1. Problem Statement:

Despite several ML and DL-based models being developed for recognizing facial emotions, there are still certain limitationsto be solved, which are mentioned below.

- The face images captured are highly susceptible to illumination differences and image noise. These outliers impact the performance of FER systems drastically.
- The facial images are always taken from multiple views and exhibit spontaneous, which makes non-frontal or in the wild FER more challenging and thus largely unexplored. In contrast to the frontal FER, expression recognition from non-frontal facial images is challenging because it needs to deal with the issues of face occlusions, accurate non-frontal face alignment, and accurate non-frontal facial points location.
- In Existing Work, Graph theory and mining concepts were employed for finding frequent sub-graphs in each emotional class by using the gSpan algorithm.The selection of the best candidate sub-graphs using the overlap ratio metric helps in reducing the number of redundant sub-graphs. But, sometimes excessive overlap ratio causes misclassification among classes.
- Compared to manual feature extraction methods, deep neural networks can learn features automatically and achieve a high recognition rate in facial expression recognition. In order to extract more facial expression features, the number of layers of neural networks is also increased gradually. However, these networks tend to overfit with the deepening of networks and the increase of parameters. The smaller the dataset, the more severe the overfitting phenomenon.
- Lack of large-scale labelled training data, inconsistent and unreliable emotion labels, and inter-subject variabilities all limit the performance of the deep learning network on the FER task.

1.2. Objectives:

To surpass these issues, an enhanced facial emotion recognition system utilizing a novel SR-NN classifier is proposed here. It comprises several research objectives, which are given further,

- To propose a novel pre-processing technique for removing the noise and enhancing the contrast of the image.

- To propose a novel Facial Landmark Extraction technique to efficiently produce the shape of the face by extracting the Facial points.
- To present a novel Facial Sub Graph mining Clustering Algorithm to extract discriminable featuresfor efficient recognition.
- To select the important features by using a novel Level Based Feature Selection technique for reducing the system complexity.
- To mount the detection accuracy and lower the computational complexity using the Novel Neural Network Classifier for classifying the facial emotion types, such as Angry, Sad, Disgust, Fear, Natural and Happy**.**

The paper is set as: Section 2 analyzes the associated work regarding the proposed facial emotion recognition. Section 3 displays a concise discussion about the detection and classification of facial expressions. Section 4 analyses the performance of the proposed system. Last, of all, section 5 wrapped up the paper.

2. Literature Survey:

ChirraVenkata Rami Reddy *et al.* [20] developed a method (involving two versions) for achieving high accuracy with limited samples. Global and local features of facial expression images were extracted using Haar Wavelet Transform (HWT) and Gabor wavelets respectively. Dimensionalities of extracted features were reduced using Nonlinear Principal Component Analysis (NLPCA). Concatenated and weighted fusion techniques had been employed for fusing the global and local features. To recognize and classify emotions from facial expressions, a Support Vector Machine (SVM) was used. From the experiment results, it was proved that better recognition accuracy of SVM was attained when compared with the existing methods. The SVM algorithm employed was not suitable for any size of input since the method had an input size limit.

AbdulrahmanAlreshidi and MohibUllah [21] suggested a modular framework for human facial emotion recognition. The framework comprised two machine learning algorithms (for detection and classification) that could be trained offline for real-time applications. AdaBoost cascade classifier was utilized for detecting face and Neighborhood Difference Features (NDF) was extracted. Then, the extracted features were classified using a random forest classifier with a latent emotional state that takes care of the mis/false detection. The experimental results demonstrated better results than the reference methods on the static facial expressions in the wild (SFEW) and real-world affective faces (RAF) datasets, respectively. However, the system failed to take the geometric feature, which led to inaccurate classification.

GuruduttPerichetla*et al*. [22] presented a facial emotion recognition using Convolutional Neural Networks (CNN). A number of different models were experimented with, including decision trees that predicted a value given a set of inputs by ”learning” rules based on a set of training data and feed-forward neural networks that attempted to identify underlying relationships in a set of data. Finally, CNN was employed for image recognition. The experimental results showed a better performance of the introduced model in comparison with other models. However, the system failed to attain good accuracy due to the absence of in-depth analysis.

Kaustubh Kulkarni *et al*. [23] introduced Automatic Recognition of Facial Displays of Unfelt Emotions. The expressions of the face were addressed by learning Spatio-temporal representations of the data. For this, aggregate features along fiducial trajectories in a deeply learned space were introduced and EMNet CNN was used to compute the feature maps. The experiment results showed a better accuracy of EMNet CNN in comparison with other techniques. The accuracy of the system was not stable due to the high computational time of EMNet CNN.

Xiao Liu *et al*. [24] suggested a facial emotion recognition technique using two newly defined geometric features, landmark curvature, and vectorized landmark. Those features were extracted from facial landmarks associated with individual components of facial muscle movements. The method combined support vector machine (SVM) based classification with a genetic algorithm (GA) for a multi-attribute optimization problem of feature and parameter selection. Experimental evaluations were conducted on the extended Cohn-Kanade dataset (CK+) and the Multimedia Understanding

Group (MUG) dataset. The experimental results showed consistent and balanced performance over varied machine learning model measurements. However, due to the presence of noise in the image, SVM showed lower detection accuracy.

3. Proposed Facial Emotion Recognition System:

In this paper, a novel SR-RNN Classifier is proposed for the efficient identification and classification of facial emotion. In this system, at first, 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 block diagram of the proposed model is shown in figure 1,

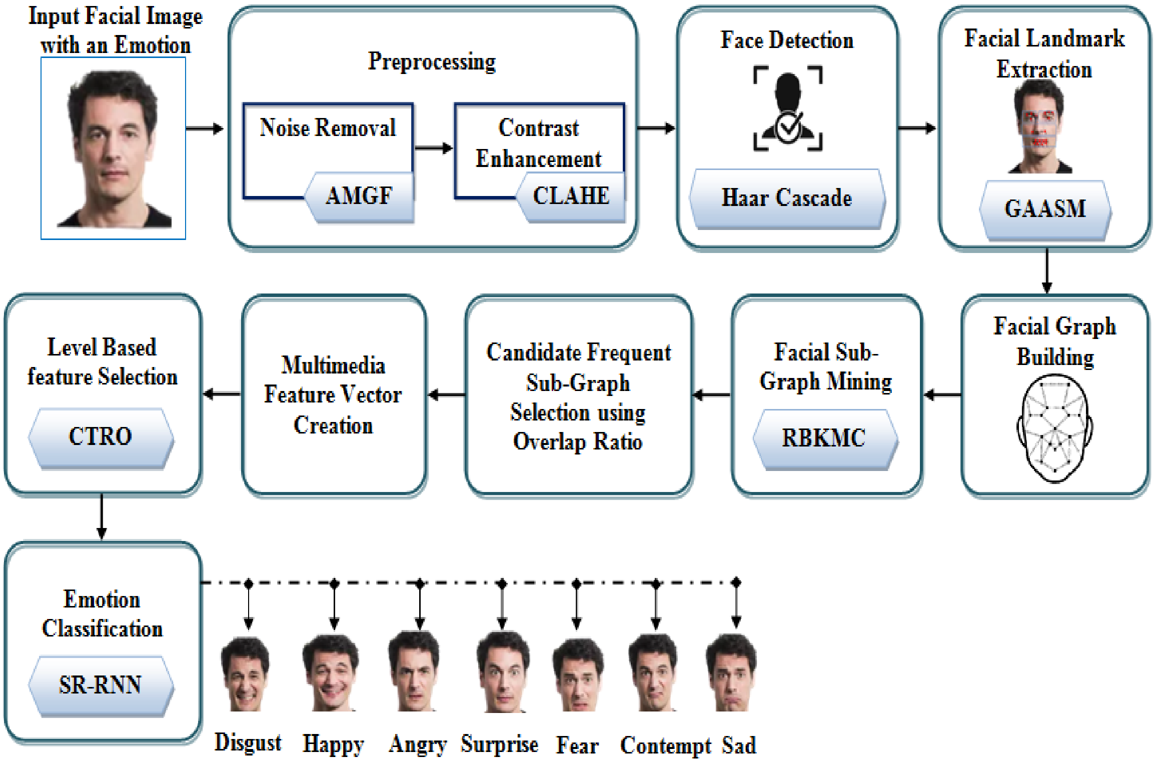


Figure 1: Block diagram of the proposed methodology

3.1 Pre-processing:

In this section, an Image with emotion is taken as an input and further, it is injected into the pre-processing because of the presence of unwanted things. The Input Face Expression Image undergoes the pre-processing operationin two stages namely, noise removal and contrast enhancement.

Noise removal: In this section, an uneven illumination, light intensity blur, Occlusion problem, and Noise present in the image are removed by using a novel Adaptive Mode Guided Filter (AMGF). A guided Filter is an edge-preserving smoothening light filter that retains the sharp edges while filtering out the noise. The existing Guided Filter has the limitations of ignorance of structural inconsistency between the reference and target image like color, infrared, and depth images captured under different conditions. Sometimes, it causes over smoothing and blurriness in an image; to overcome that the adaptive mode function is used to calculate the Cost Function instead of the mean.

Let the input image is (F) and the filtered output can be obtained by considering the guided image (G) which can be the input image itself or a different image. The guided image is filtering the input image by imprinting its structure over the original input image. The filtration is the linear transformation of a guided image in a window (w_k) centered at the pixel (k). Hence the filtering output at a $i\,th$ pixel of the image can be expressed as,

$$F_i^{filter} = G_i \ell_k + B_k \quad (1)$$

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, G_i is the guided image at the ith 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,

$$C = \sum_{i \ni w_k} ((G_i \ell_k + B_k - F_i)^2 + \ni \ell_k^2) \quad (2)$$

Where, $$\ell_k = \frac{\frac{1}{\ell} \sum_{i \ni w_k} F_i G_i - \mu_k F_k^{'}}{\Delta_k^2 + \ni} \quad (3)$$

$$B_k = F_k^{'} - \ell_k \mu_k \quad (4)$$

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

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

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

Contrast Enhancement: In this section, the filtered image F^{filter} is enhanced by using Contrast Limited Adaptive Histogram Equalization (CLAHE) technique. It is a variant of the adaptive histogram equalization technique, which is applied over all neighbourhood pixels to enhance the contrast of the image. Here the contrast amplification can be mitigated by a clip limit.

At first, the filtered facial image is divided into contextual regions with an equal number of pixels. It is also known as tiles. Then the histogram is calculated for each tile using the pixels presented in the image. Then the average pixel in the gray level $\Re_{avg}$ is calculated as,

$$\Re_{avg} = \frac{N_s - N_t}{N_g} \quad (6)$$

Where, N_s represents the number of pixels in x direction of the contextual region, N_t represents the number of pixels in y direction of the contextual region, N_g represents the number of gray levels. The clip limit is set based on the average gray level and it is expressed as,

$$F_{CL} = \Re_{avg} \times F_{\text{max(avg)}}^{\text{filter}} \quad (7)$$

Where, $F_{\text{max(avg)}}^{\text{filter}}$ is the maximum multiple of average pixels in each gray level of the contextual region, F_{CL} is the clip limit. Then, the pixels over the clip point limit are considered excess pixels and redistributed them for each gray level. After performing the equalization, it combines all the neighbouring tiles using bilinear interpolation to eliminate induced boundaries in the input image. Hence the image after the contrast enhancement can be denoted as F^{con}.

3.2 Face detection

Followed by the contrast enhancement, the face is detected from F^{con} by using the Haar cascade. In general, the Haar cascade detects the face using edge or line detection features. The haar features are

ISSN NO: 2230-5807

also known as rectangular features. In this paper, for the quick calculation of features from the input image, the concept of the integral image has been used. In the integral image, the value in pixels is the number of pixels above and on the left of the image. Each of the haar features is scanned over the integral image and the threshold level has occurred after that which shows 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.

$$F^{face} = \begin{cases} 1 & if \ \ face \det ected \\ 0 & if \ \ face \, is \, not \det ected \end{cases} \tag{8}$$

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

3.3 Facial Landmark Extraction:

Followed by face detection, the facial landmarks are extracted from F^{face} by using the novel Generative Additive Active Shape Model (GAASM). The active Shape Model (ASM) is a statistical model of the shape of objects, which iteratively deform to fit into the object in a new image. But the existing ASM has the limitation that it only uses shape constraints together with some information about the image structure near the landmarks. To overcome that the Generative Additive model is used to calculate descriptor matches replacing the Mahalanobis distance. The landmarks extraction undergoes two steps, namely the profile model and shape model.

Profile models: The profile model helps to improve the feature matching executions over the iterative GAASM fitting process by locating the approximate position over the face region. The approximate position is determined by the template matcher. Here the template model is obtained by a sampling of the input image. During searching, the landmark along the line moved to the pixel who has the lowest Generative Additive distance from the mean profile, and the distance is calculated as,

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

Where, A represents the output of the Generative Additive distance, α is the constant value, f_i are the functions with a specified parametric form, v is the predicted variable, Minimizing the Generative Additive distance is equivalent to maximizing the probability that originates from a multidimensional Gaussian distribution. From the process, the suggested face F_{sug}^{face} can be obtained.

Shape models: In this section, the shape of the suggested face mode F_{sug}^{face} is conformed to the shape model. Hence it generates the shape F_{shape}^{face} is expressed as,

$$F_{shape}^{face} = \bar{F}_{sug}^{face} + \Omega\beta \tag{10}$$

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

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

Where, L_n represent the extracted landmark points from the face region.

3.4 Facial Graph Building

In this section, the facial graph is built using the extracted landmark points L_n. Extracted facial landmarks are used to localize the location of key regions of the face, such as: eyebrows, eyes, nose, mouth, and jawline.In this paper, extracted landmark point indices are used as vertices while the edges are calculated in the following steps.

Let us consider the edge points in land mark is $L_1(t_1,u_1)$ and $L_2(t_2,u_2)$ and calculate the distance between these two points 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, $d_{L_{1,2}}$ is the distance between the two edge points 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, $d_{\max}$ represents the maximum distance in the face region. In this way, distances of edge points for the entire landmark are calculated. The range of distances is divided into sub-ranges with an assigned label value to cover all possible distance values. Then the label of each edge is computed. Finally, a fully connected undirected graph $\Im_{gh}$ is built for further steps of emotion mining.

3.5 Facial Sub-graph Mining:

In this section, the facial sub-graph is mined from $\Im_{gh}$ using the novel Radial Basis K –Medoids clustering (RBKMC) algorithm to extract discriminable features that represent the common change in the facial graph. In general clustering, clusters are fully dependent on the selection of the initial cluster centroids. Here the distances of all data elements are calculated by the Euclidean distance formula. Data elements having less distance to centroids are moved to the appropriate cluster. But the basic Euclidean distance is not suitable for a large amount of data so the Radial Basis Kernel function is calculated here, which provides a better result than the Euclidean distance calculation and also achieves better clustering accuracy. In this paper, the points that are built in the facial graph points are taken as a data points which are utilized in further steps are,

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

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

Where, K_{rb} represents the Radial Basis Kernel function, D_i denotes the ith data point, σ signifies the standard deviation

- Calculate the total cost function to the selected medoids and it can be expressed as,

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

- 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.
- Recalculate the cost function for the current medoid c_{new}.
- 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, O_{old} specifies the old medoid. This process is continued for some points and may shift from one cluster to another cluster depending upon the closeness of medoids. 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 $\Im_{sub(n)}$. The pseudo-code for the proposed RBKMC is shown in figure 2,

Input: Facial graph building $\Im_{gh}$

Output: Facial graph mining $\Im_{sub(n)}$

Begin

Initialize K_{dis}, D_i, σ, 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:

In this step, the overlap ratio is calculated between the sub-graphs $\Im_{sub(n)}$ for selecting the frequent sub-graphs of the candidate. 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.

i. For each emotion, the overlap ratio that each sub-graph makes with the remaining emotional graphs is computed.
ii. Arrange the subgraphs from high overlapping to small overlapping.
iii. Select subgraphs with the smallest overlap. This selection is continued till the overlap ratio is zero. The candidate frequent sub-graphs selection $\Im_{can(n)}$ can be expressed,

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

Where, ω_i denotes the number of the graph that contains. 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 the creation of multimedia feature vectors from $\forall$ for the purpose of being able to apply deep learning. Hence binary encoding is performed for creating 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 is matched with the predetermined graph then the graph is represented in a multidimensional vector with a value of 1 otherwise zero.

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

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

3.8 Level-based feature selection:

Followed by the feature vector creation, the important features are selected from $f_{mul(n)}$ by using a novel Chaotic tend Remora optimization (CTRO) Algorithm for reducing the computation time. Remora Optimization Algorithm (ROA) is a natural-inspired and meta-heuristic algorithm. The inspiration for ROA is mainly due to the parasitic behaviour of remora. In the ROA different locations are updated in different hosts. There are two phases namely Exploration and Exploitation. In general, in Remora Optimization, each search agent searches new space according to the position of the host, which makes the algorithm suffer from the drawbacks of slow convergence rate, poor solution accuracy, and local optima for some optimization problems. To tackle these problems, the Chaotic Tend Operator is used to update the Position of Swordfish.

First, the population of remora is initialized (i.e. the extracted features $f_{mul(n)}$) and it can be expressed as,

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

Where, r denotes a random variable between $[0,1]$, U_b and L_b represent the search space's upper and lower bounds, N is the number of remora. CTRO is updating position using Whale Optimization Algorithm (WOA) and SailFish Optimization (SFO) strategies and also uses an integer argument $I\ (0\ or\ 1)$ to determine WOA strategy or SFO strategy.

Exploration: In some small hosts, remora follows the sailfish to move from the bait-rich area to prey. The sailfish is one of the fastest-moving fish. Thus, remora attaches with sailfish to perform the global search for achieving quick long-distance movement and update their position to the host's position. This strategy is known as the SFO strategy. The updating position $f_{mul}(t+1)$ can be expressed as,

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

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 remora and take the small step to attack the current host. The new candidate position $f_{mul}{}^{*}(t+1)$ updated based on the chaotic tend function and can be expressed as,

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

Where, $f_{mul}^{pre}(t)$ signifies the position of the previous generation, $f_{mul}^{up}(t+1)$ is the updating position, $Q\left(f_{mul}{}^{*}(t+1)\right)$ is the fitness value for the updated position. The fitness value is calculated based on the accuracy of the classifier.

Exploitation: In some large hosts, remora feeds on the host's ectoparasites or wreckage and evades natural enemies by continuing the local search. Hence, the remora attaches to the whale on the humpback for attacking the prey. This process is known as the WOA strategy. The position update formula while remora attaches with the whale can be formularized as,

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

$$where, \Gamma = r(\Xi - 1) + 1 \qquad (23)$$

$$\Xi = -\left(1 + \frac{t}{T}\right) \qquad (24)$$

Where, ξ is the distance between best position and current position, Γ is the random number in the range of $[-1,1]$, Ξ is the search space which linearly decreases from -1 to -2, 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,

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

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

$$\phi = 2 * \vartheta r - \vartheta \qquad (27)$$

$$\vartheta = 2\left(1 - \frac{t}{\Phi_{\max}}\right) \qquad (28)$$

Where, Z specifies the small movement of the remora, ϕ is the volume space of the random host, ϑ is the volume parameter, R is the remora factor, $\Phi_{\max}$ is the maximum number of iterations. During the feeding of remora on the host, the search space can be seen as gradually decreasing. In this way, the features are selected as same as the process remora for searching food. The extracted features are denoted in the term of $f_{mul(n)}^{sel}$.

3.9 Classification:

In this identification phase, the selected features from the face region $f_{mul(n)}^{sel}$ are given as input to the SwikyRelu Recurrent Neural Network (SR-RNN) algorithm to classify the facial emotions. A recurrent Neural Network (RNN) is a class of artificial neural networksthat exhibits temporal dynamic behaviour. The algorithm consists of three layers. The first layer is the input layer and the last layer is the output layer. Between the input and output layer, there may be additional layers of units, called 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. But the existing RNN has gradient vanishing and exploding problems, also it cannot process very long sequences. To overcome such limitations, Swiky (Swish and Leaky Relu) activation Function is used in Neural Network, and also Batch Normalization layer is included in the Recurrent Neural Network. The architecture of SR-RNN is shown in figure 3,

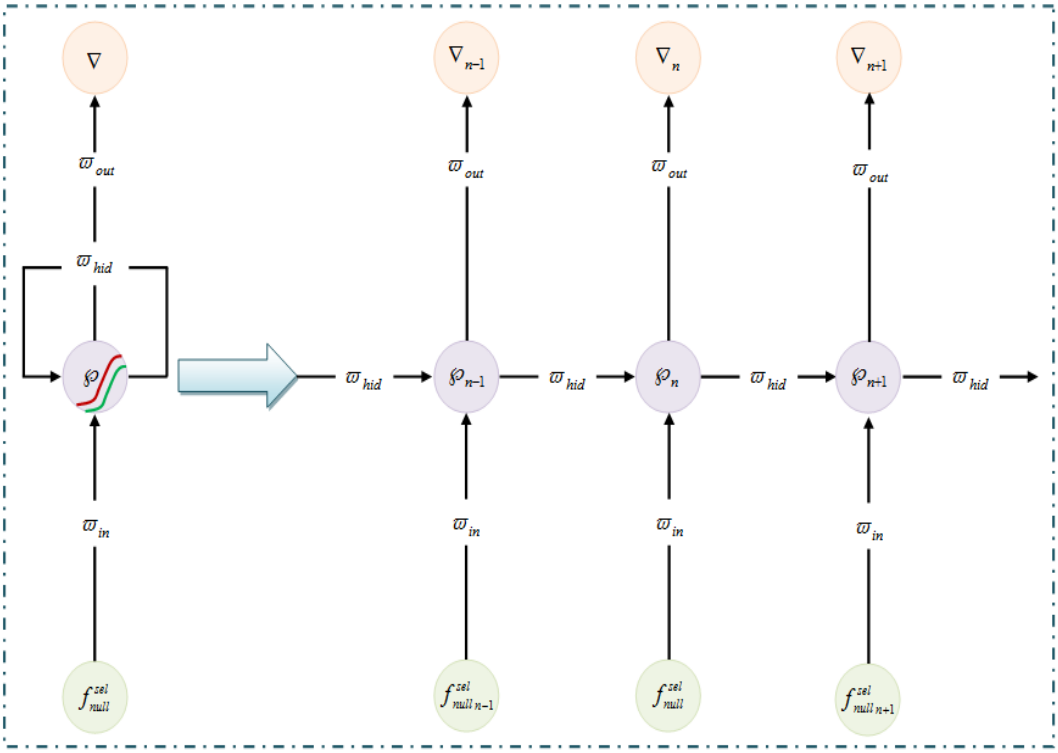


Figure 3: structure of proposed SR-RNN

The extracted features are given as an input to the input layer and then the output of the input layer is further injected into the hidden layer. The middle layer can consist of multiple hidden layers, each with its own weights and biases and activation functions such as SwikyRelu(SR) and Batch Normalization(BN). In the hidden layers, the edges that are connected recurrently yield input from two parts. One part comes from the hidden nodes that are generated at the time step q and the other comes from the hidden nodes that are generated at the time step $q-1$. Then the hidden layer is calculated as,

$$\wp_n = S\left(\varpi_{in} f^{sel}_{mul(n)} + \varpi_{hid}\wp_{n-1} + Ba\right) \tag{29}$$

Where, ϖ_{in} and ϖ_{hid} are the weights of input and hidden layer, $\wp_{n-1}$ represents the output of the previous hidden layer which stored in memory, Ba is the bias vector, S represents the output of SwikyRelu activation function. The hidden layers delivered their output with the help of the SwikyRelu activation function and batch normalization and those can be expressed as,

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

$$\infty = 1(W<0)(\rho) + 1(W>=0)(W) \quad Where, W = \varpi_{in} f^{sel}_{mul(n)} + \varpi_{hid}\wp_{n-1} \tag{31}$$

Where, y is the constant value, ∞ is the output of the Relu activation function, ρ denotes the small constant. Then, the output layer is responsible to determine the face emotion. The output layer is activated by the sigmoid activation function (σ_S), and the output layer ∇_n can be determined by,

$$\nabla_n = \sigma_S[\varpi_{out}\wp_n + Ba] \tag{32}$$

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

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

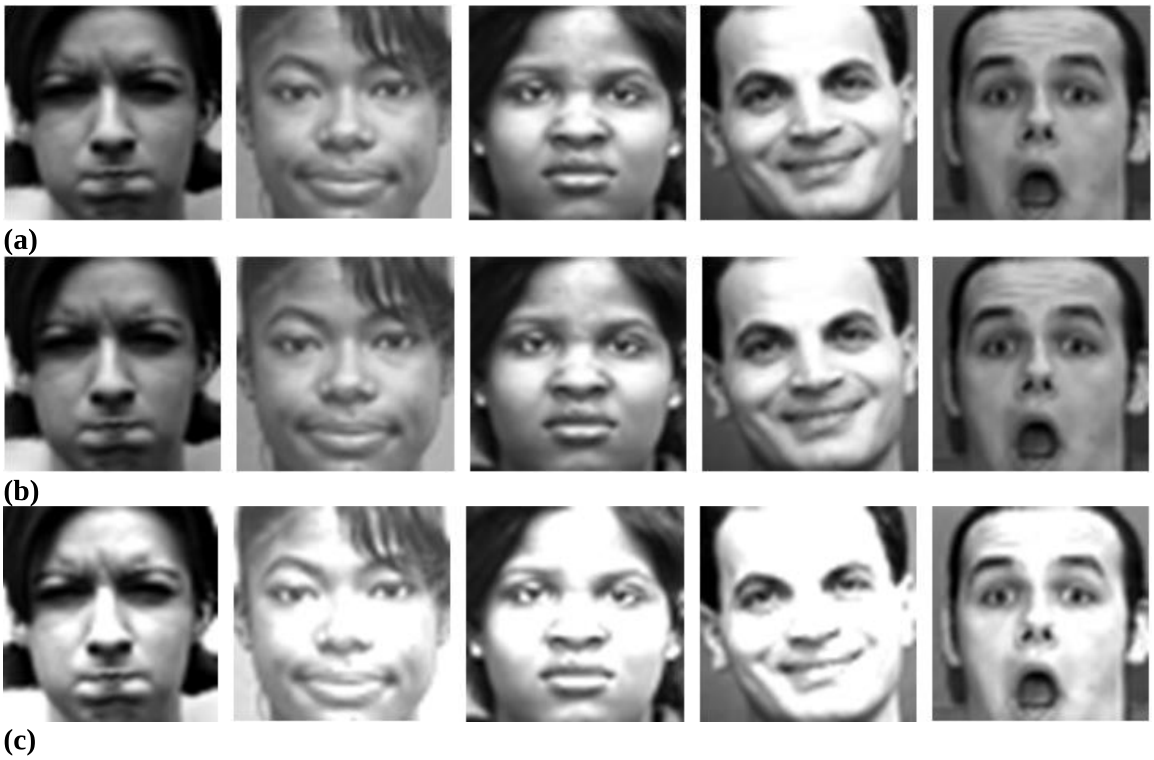
Finally, ε is the Eluer's number. The SR-RNN classifier efficiently classifies the facial emotion as Angry, Happy, Disgust, Fear, Surprise, and Neutral.

4. Result and Discussion:

In this section, the performance of the proposed facial emotion recognition system with SR-RNN is analyzed. The proposed system is implemented in the working platform of PYTHON.

4.1 Database Description:

In this research methodology, the performance of the methods is analyzed by using the **CK+** dataset, which is available in the publically available data source.**CK+** dataset contains 593 video sequences from a total of 123 different subjects, ranging from 18 to 50 years of age with a variety of genders and heritage.Out of these videos, 327 are labelled with one of seven expression classes: anger, contempt, disgust, fear, happiness, sadness, and surprise. From the database images, 80% of the data are used for training and 20% of the data are used for testing the data. Sample images are collected from the **CK+**dataset and it is injected into the operation as shown in figure 4,



(a)
(b)
(c)

(d)

(e)

Figure4: sample images of 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. Performance analysis of noise removal:

Here, the performance of the AMGF is evaluated and compared with the existing methods, such as the existing Guided filter (GF), Weiner filter (WF),Bilateral Filter (BF),and adaptive filtering (AF).

Table 1: Performance Analysis of Proposed AMGF and existing models

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

Discussion: Table 1 shows the performance analysis of the proposed model and existing models. The peak signal-to-noise ratio (PSNR) is used for measuring the quality of the image. A higher PSNR value gives better quality to the proposed model. The proposed model attains a PSNR of 25.5478db,whereas the PSNR of existing models is improved by 1.8128db, 5.2331db and 6.3347db than the existing GF, WF, and BF respectively. On comparing these values, the proposed model achieves a higher PSNR value than the existing models.

4.3. Performance analysis of the classification method:

To prove the effectiveness of the proposed SR-RNN method, the performance of the proposed method is analyzed and compared with the existing methods, such asRecurrent Neural Network (RNN), Deep Neural Network (DNN), Convolution Neural Network (CNN), and Artificial Neural Network (ANN).

Table 2: Performance Analysis of Proposed SR-RNN and existing models

Method	FPR	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

Discussion: Table 2 shows the performance analysis of the proposed SR-RNN and existing models in terms of False Positive Rate (FPR), False Negative Rate (FNR)False Rejection Rate (FRR), and Positive predictive value (PPV).Lower values of FPR, FNR, and FRR, and higher value of PPV showed a higher performance of the proposed method.The FNR and FRR of the proposed model are the same at 0.008861, which is lower compared with the existing models, whereas the existing models attain the FNR and FRR in the range of 0.03 to 0.16, which is higher compared to the proposed model. The proposed model attains an FPR of 0.0238, which is lower compared to existing models. Likewise, the PPV of the model is 0.995058, which is higher than the prevailing techniques. So, from the metrics, it is concluded that the proposed model shows better performance for multimodal biometric authentication

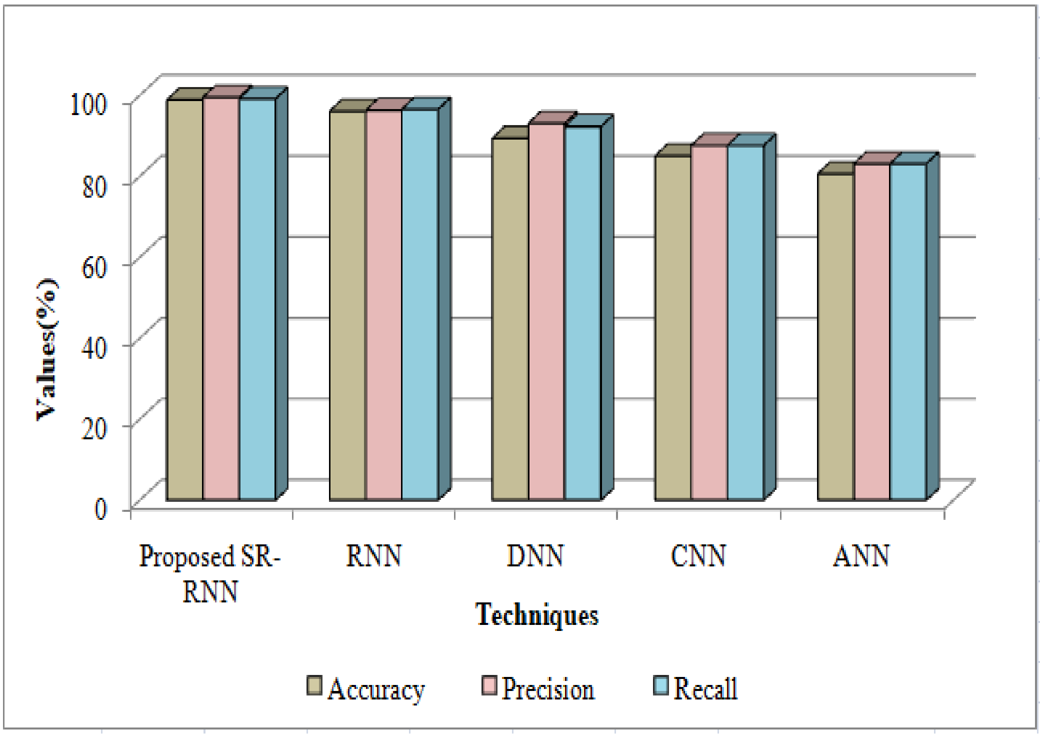


Figure 5: Graphical representation of SR-RNN with the existing methods

Discussion: Figure 5 displayed the performance analysis of the proposed SR-RNN and existing methods based on accuracy, precision, and recall. The higher value of the proposed model in above mentioned metrics indicates the higher performance of the recommended system. The proposed method recognizes facial emotion with an accuracy of 98.85%, which is higher than the existing methods. The accuracy of the existing CNN is 85.05%, DNN is 89.40%, and RNN is 96.02%.Similarly, the value of the precision and recall of the proposed SR-RNN are 99.5% and 99.11%, respectively, which arealso higher than the conventional models. So from the results, it can be concluded that the developed model is highly superior to prevailing methods.

Table 3: Performance Analysis of Proposed SR-RNN and existing 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	77.19409	83.1106	83.11060201

Discussion: The performance analysis of the proposed model and existing methods based on Sensitivity, specificity, and f-measure is represented in above table 3.The degree of similarity between views of an object that elicits a correct recognition response from a given subject is referred to as specificity. The ability to correctly classify a test between True and False positive performance is referred to as sensitivity. Finally, the combination of precision and recall is known as F-Measure. A higher value for the proposed model in the above-mentioned metrics indicates that the recommended system performs better. According to this state, the specificityof the proposed model is97.61%,which is 2.22% improved than existing RNN, 15.66% improved than the DNN, and 16.49% improved than the CNN.Likewise, the F-measure of the proposed model is 91.71% and the sensitivity is 99.11% compared with the conventional methods. The comparison proved the higher performance of the developed model.

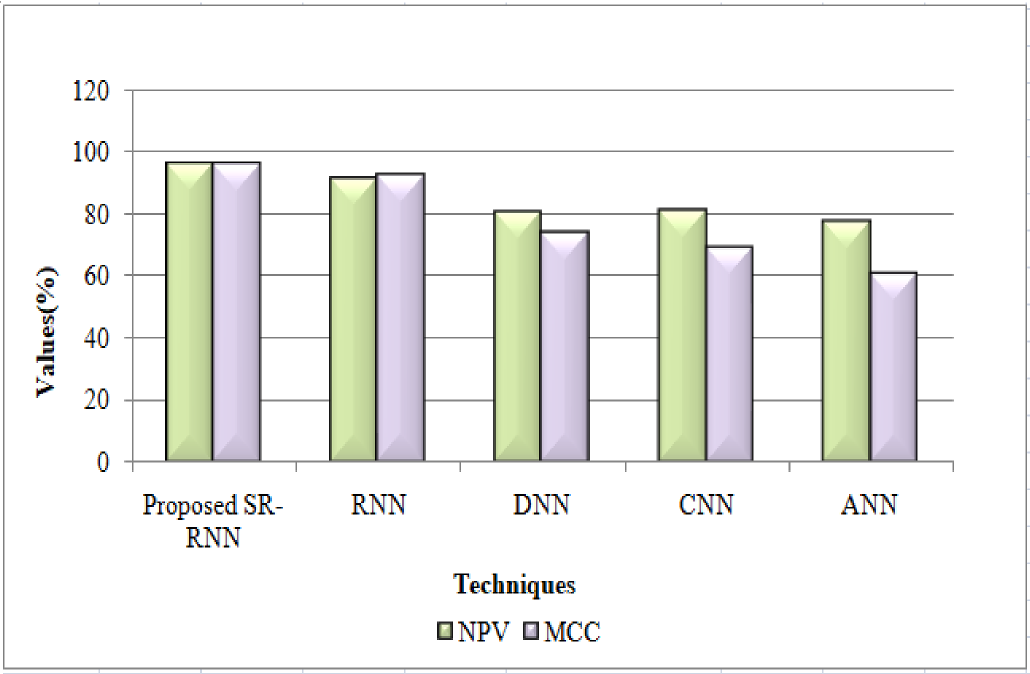


Figure 6: Performance Analysis of Proposed SR-RNN and existing models in terms of NPV and MCC

Figure 6 unveils the better results of the proposed model with comparison. To attain better performance for the proposed model, it should reach higher values. The NPV of the proposed model is 95.79%, which is higher than the existing models. Existing RNN attains 90.73%, DNN attains 80.17%, and CNN attains 81.12%. Likewise, the MCC of the proposed and existing methods are analyzed and compared. The MCC of the proposed model is 96.01%, which is also higher than the conventional models. So, from this overall comparison results, it can be proved that the proposed model has high efficacy than the prevailing methods by inducing the novelty in the classifier.

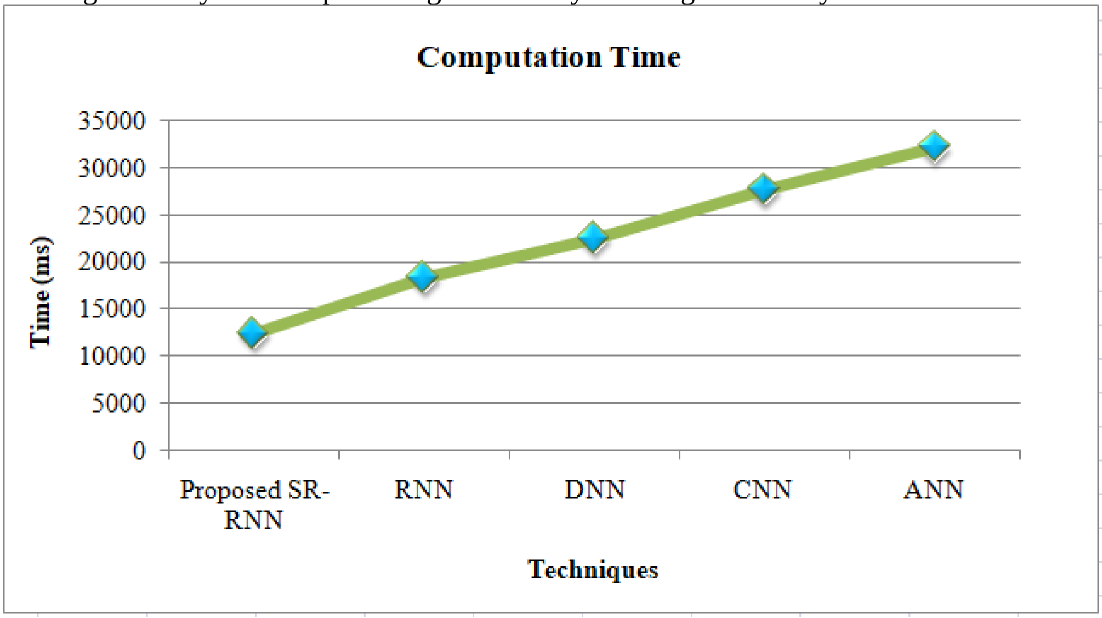


Figure 7: Demonstrate the performance of computation time

Discussion: Figure 7 shows the representation of the comparative time analysis. Computation time is the process of the execution time of classification. While making the comparison of the proposed method with RNN and other existing methods, existing methods have a slow and complex training procedure, and the proposed method executes its process very fastly due to the inducement of SwikyRelu.Hence, the computational time of the proposed method is 12436ms, which is too much low when compared to the other existing methods of RNN (18316ms), DNN (22478ms), CNN (27648ms), and ANN (32147ms).

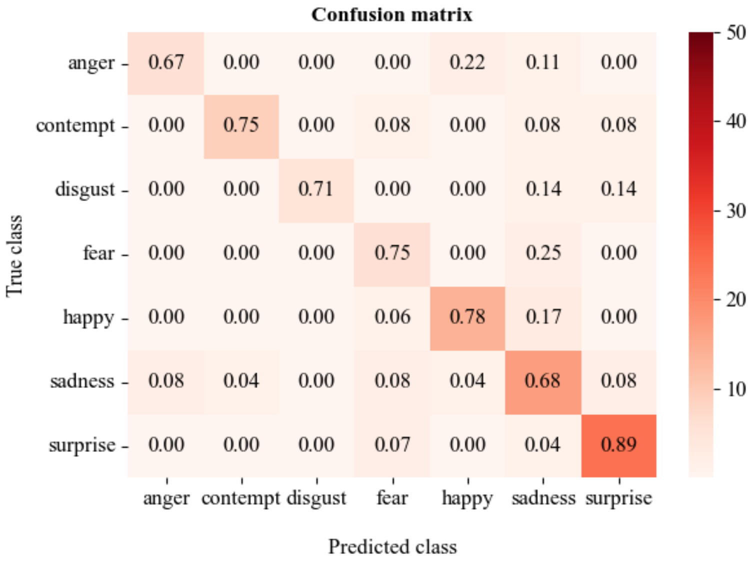


Figure 8: confusion matrix of the proposed SR-RNN

Figure 8 shows a confusion matrix that shows the accuracy of the predicted labels of the validation data. It depicts the classification model's behaviour using the proposed SR-confusion RNN's matrix in terms of seven classes: angry, disgusted, fearful, happy, neutral, sad, and surprised.The accuracy of a classifier is used to determine the proportion of correctly classified instances.Surprise has the highest accuracy of the seven facial expressions with a score of 89%. The confusion matrix also provides the accuracy of the following facial expressions: contempt (75%), Sad (68%), happy (78%), Fear (75%),Disgust (71%), and Anger (67%). All seven facial expressions are approximately 58.50% accurate on average.

4.4. Comparative measurement with literature papers:

In this section, the performance of the proposed SR-RNN is compared with the CNN [20], Support Vector Machine (SVM) [21], and genetic algorithm based SVM (GA-SVM) [24].

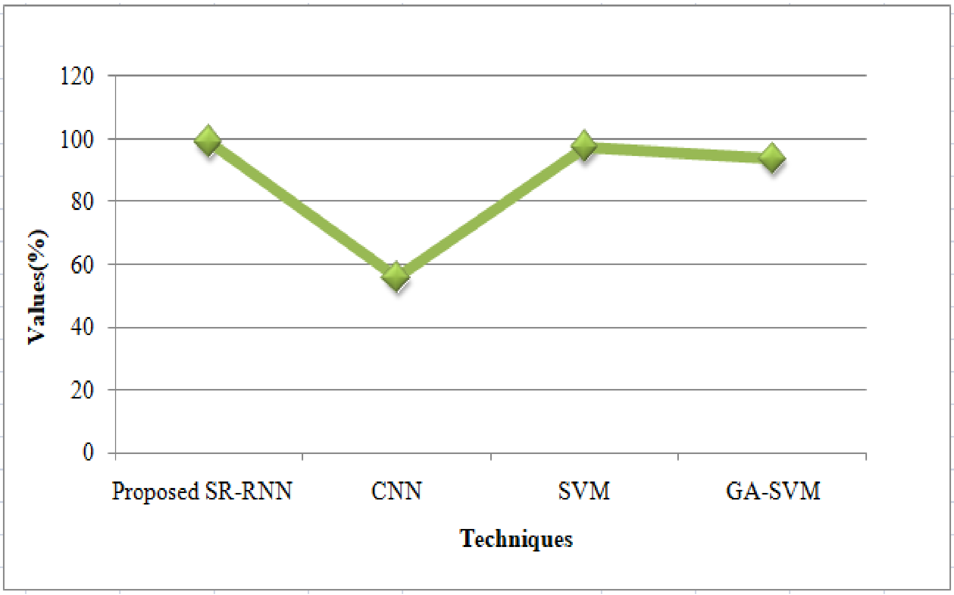


Figure 9: A comparative analysis of the proposed method with the existing method

Figure 9 displays the performance of the proposed methodology with the existing algorithms based on accuracy. Here, the proposed work is considering the efficient landmark extraction and appropriate feature selection method. By this accurate processing, the proposed methodology can attain better results when compared with all other existing methodologies that are, 98.4%. Finally, the overall discussion shows that the performance of the proposed methodology is better than the existing methodologies.

5. Conclusion:

The work has proposed facial emotion recognition using a novel SR-RNN classifier. The proposed method undergoes the following process such as attribute extraction, grouping, data clustering, crime data mapping, word to vector conversion and classification. After executing all the steps, the experimental analysis are performed in the performance of the proposed SR-RNN was analyzed and compared with the existing techniques in terms of various metrics. The final outcomes reveal that the proposed model achieves an accuracy of 90.5821% and the computation time is 45617ms. Similarly, the proposed model achieves the best result for all other metrics such as sensitivity, specificity, precision, recall, f-measure, NPV, MCC, FPR, FNR, and FRR. So, from the results of all metrics, it is concluded that the proposed model is highly efficient for the classification of facial emotion when compared with the existing methods. Facial emotion recognition has numerous applications in human-computer interactions, augmented reality, virtual reality, affective computing, and advanced driver assistance systems (ADSSS). FER can also be used in areas, such as website customization, gaming, healthcare, and education. It has greater potential in surveillance and military applications. In the future, the work will be extended for advanced recognition under different head poses.

References:

1. ImaneLasri, AnouarRiadSolh and Mourad El Belkacemi, “Facial emotion recognition of students using convolutional neural network”, International Conference on Intelligent Computing in Data Sciences, IEEE, 28-30 October 2019, Marrakech, 2019.
2. Malika Arora and Munish Kumar, “AutoFER PCA and PSO based automatic facial emotion recognition”, Multimedia Tools and Applications, vol. 80, no. 1-2, pp. 3039-3049, 2020.
3. Milad Mohammad TaghiZadeh, Maryam Imani and BabakMajidi, “Fast facial emotion recognition using convolutional neural networks and gabor filters”, 5th Conference on

Knowledge Based Engineering and Innovation, 28 February - 1 March 2019, Tehran, Iran, 2019.
4. Deepak Kumar Jain, PouryaShamsolmoaliandParamjitSehdev, “Extended deep neural network for facial emotion recognition”, Pattern Recognition Letters, vol. 120, pp. 69-74, 2019.
5. YousifKhaireddin and Zhuofa Chen, “Facial emotion recognition state of the art performance on FER2013”, 2021. https://arxiv.org/abs/2105.03588v1
6. KalpanaChowdary M, Tu N Nguyen and Jude Hemanth D, “Deep learning-based facial emotion recognition for human computer interaction applications”, Neural Computing and Applications, 2021.https://doi.org/10.1007/s00521-021-06012-8
7. LutfiahZahara, Purnawarman Musa, EriPrasetyoWibowo, Irwan Karim and SaifulBahri 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, 3-4 November 2021, Gorontalo, Indonesia, 2021.
8. EimanKanjo, Eman M. G Younis and Chee Siang Ang, “Deep learning analysis of mobile physiological, environmental and location sensor data for emotion detection”, Information Fusion, vol. 49, pp. 46-56, 2018.
9. Alia K Hassan and Suhaila N Mohammed, “A novel facial emotion recognition scheme based on graph mining”, Defence Technology, vol. 16, no. 5, pp. 1062-1072, 2020.
10. Jad Haddad, Olivier Lezoray and Philippe Hamel, “3D-CNN for facial emotion recognition in videos”, Springer, Cham, 1st Edition, ISBN: 978-3-030-64558-8, 2020.
11. Ansamma John, Abhishek M. C, Ananthu S Ajayan, Sanoop S and Vishnu R Kumar, “Real time facial emotion recognition system with improved preprocessing and feature extraction”, 3rd International Conference on Smart Systems and Inventive Technology, IEEE, 20-22 August 2020, Tirunelveli, 2020.
12. VihaUpadhyay and Devangi Kotak, “A review on different facial feature extraction methods for face emotions recognition system”, 4thInternational Conference on Inventive Systems and Control, IEEE, 8-10 January 2020, Coimbatore, India, 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, no. 10, pp. 1-9, 2021.
14. Wang Xiaohua, Peng Muzi, Pan Lijuan, Hu Min, JinChunhua and Ren Fuji, “Two level attention with twostage multi-task learning for facial emotion recognition”, Journal of Visual Communication and Image Representation , vol. 62, pp. 217-225, 2018.
15. Hai-Duong Nguyen, SoonjaYeom, Guee-Sang Lee, Hyung-Jeong Yang, In-Seop Na and Soo-Hyung Kim, “Facial emotion recognition using an ensemble of multi-level convolutional neural networks”, World Scientific Connecting Great Minds, vol. 33, no. 11, pp. 1-18, 2018.
16. SumeetSaurav, Ravi Saini and Sanjay Singh, “EmNet a deep integrated convolutional neural network for facial emotion recognition in the wild”, Applied Intelligence, vol. 51, no. 16, pp. 5543-5570, 2021.
17. Ashok Kumar P. M, JeevanBabuMaddala and Martin Sagayam K, “Enhanced facial emotion recognition by optimal descriptor selection with neural network”, IETE Journal of Research, vol. 33, no. 11, pp. 1-21, 2021.
18. AlirezaSepas-Moghaddam, Ali Etemad, Fernando Pereira and Paulo LobatoCorreia, “Facial emotion recognition using light field images with deep attention-based bidirectional LSTM”, International Conference on Acoustics, Speech and Signal Processing, IEEE, 4-8 May 2020, Barcelona, Spain, 2020.
19. Anil Pise, HimaVadapalli and Ian Sanders, “Facial emotion recognition using temporal relational network an application to e-learning”, Multimedia Tools and Applications, 2020.https://doi.org/10.1007/s11042-020-10133-y

20. AbdulrahmanAlreshidi and MohibUllah, “Facial emotion recognition using hybrid features”, Informatics, vol. 7, no. 1, pp. 1-13, 2020.
21. AkashSaravanan, GuruduttPerichetla and Gayathri K. S, “Facial emotion recognition using convolutional neural networks”, SN Applied Sciences, vol. 2, pp. 1-8, 2019.
22. ChirraVenkata Rami Reddy, UyyalaSrinivasulu Reddy and KolliVenkata Krishna Kishore, “Facial emotion recognition using NLPCA and SVM”, International Information and Engineering Technology Association, vol. 36, no. 1, pp. 13-22, 2019.
23. Kaustubh Kulkarni, Ciprian Adrian Corneanu, IkechukwuOfodile, Sergio Escalera, Xavier Baro, SylwiaHyniewska, JuriAllik and GholamrezaAnbarjafari, “Automatic recognition of facial displays of unfelt emotions”, Journal of IEEE Transactions on Affective Computing, vol. 12, no. 2, pp. 377-390, 2018.
24. Xiao Liu, Xiangyi Cheng and Kiju Lee, “GA-SVM based facial emotion recognition using facial geometric features”, IEEE Sensors Journal, vol. 21, no. 10, pp. 11532-11542, 2021.