International Journal of Engineering & Technology, 7 (1.1) (2018) 125-132

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET

Research paper

Simultaneous evolutionary neural network based automated video based facial expression analysis

S. Jeyalaksshmi [1*], S. Prasanna [2]

[1]*Assistant Professor, Dept of IT, Vels University, Pallavaram, Chennai.*
[2]*Associate Professor, Dept. of Computer Application, Vels University, Pallavaram, Chennai*
**Corresponding author E-mail: pravija.lakshmi@gmail.com*

Abstract

In real life scenario, facial expressions and emotions are nothing but responses to the external and internal events of human being. In Human Computer Interaction (HCI), recognition of end user's expressions and emotions from the video streaming plays very important role. In such systems it is required to track the dynamic changes in human face movements quickly in order to deliver the required response system. In real time applications, this Facial Expression Recognition (FER) is very helpful like physical fatigue detection based on facial detection and expressions such as driver fatigue detection in order to prevent the accidents on road. Face expression based physical fatigue analysis or detection is out of scope of this work, but this work proposed a Simultaneous Evolutionary Neural Network (SENN) classification scheme is proposed for recognising human emotion or expression. In this work, at first, automatically detects and tracks facial landmarks in videos, and face is detected by using enhanced adaboost algorithm with haar features. Then, in order to describe facial expression modifications, geometric features are taken out and the Local Binary Pattern (LBP) is extracted to improve the detection accuracy and it has a much lower-dimensional size. With the aim of examining the temporal facial expression modifications, we apply SENN probabilistic classifiers, which examine the facial expressions in individual frames, and after that promulgate the likelihoods during the course of the video to take the temporal features of facial expressions such as glad, sad, anger, and fear feelings. The experimental results show that the performance of proposed SENN scheme is attained better results compared than existing recognition schemes like Time-Delay Neural Network with Support Vector Regression (TDNN-SVR) and SVR.

Keywords: *Human computer interaction (HCI), facial expressions and emotions, simultaneous evolutionary neural network (SENN), adaboost, geometric features, local binary pattern (LBP), classification.*

1. Introduction

Now days in many real time applications HCI based systems are used to immediately and accurately track the human activities from the videos. One such area is realizing and tracking the human face expression and emotions recognitions from the video streaming with an objective of different purpose such as physical fatigue detection. In human-to-human conversation, the sound of mental, emotional, and even physical state is used in conversations about important information in addition to pronounce a communication channel and facial expressions is the notion of a human's facial expressions in its modest form is a more elusive happy or angry thoughts [1], feelings or absorption of all speaker expectation from listeners, sympathy, or even what the speaker is saying no signal can provide to computing background, bring our routine human user to persist at the lead in the fabric would move to absorb that predict a generally establishment [2]. It pervasive computing and ambient intelligence such as want to achieves for future computing. it's easy to naturally occurring multimodal human-human communication-focused result to the end user will need to developed to identify such interfaces and purposes and as articulated by emotional state of social as well as emotional signals would need to have the ability to sense future nonverbal actions and expressions. The automatic recognition Research inspired.

Facial expression recognition, computer vision, pattern identification and human and computer interaction research has attracted towards notices in communities. Mechanized identification of facial expressions affective computing technologies, comprising intelligent tutoring systems supply the essence of the next generation computing device a forms, patient monitoring systems, and so on. Individual wellness profiled. Various age groups, Human face, genders and other physical features of a person varies from the reason [3].

Emotions are basic to human beings in day to day interactions, and it use in everyday life. Emotion recognition has become an important and interesting field of study in HCI, Human Robot Interaction (HRI), and so on. The six basic emotions are, sickening, happy, fear, anger, sad and surprise. Computer graphics, automatic driver fatigue detection, 3D/4D avatars animation in the entertainment industries, psychology, video & text chat and gaming applications are include in diverse applications. Recognition of emotions from facial expressions using videos consists of preprocessing, feature pulling out and division. Importance of facial expression system is widely recognized in social interaction and social intelligence system analysis is an active research topic since the 19th century. Suwa ET was introduced facial expression recognition in 1978 Al. creating a facial expression recognition system the main point of face detection and alignment Feature extraction and classification, image standardization [4].

There are techniques that we use to identify facial expressions to speed up an efficient algorithm number. Efficient algorithm faces motion detection by using optical flow proposed for facial it must be based on either recognition detection technology that is based on the optical flow vector speed infusion technology. Optical flow speed during intervals of time reflects the image changes the algorithm works on segmented image frames and we offer vector depends on their results the strongest degree of equality gives facial emotions determines. Algorithms of Unit (AU) coded facial expression based on database operation unite. Using this method can identify facial expressions match there to recognize that there are four types of expression [5].

Facial expression to recognize the first type uses emotion s speed. The second type of optical flow using facial expression to identify an image frame is the third Type facial expression to recognize the active shape model to use. The fourth type neural networks [3] using facial expression to recognize face a complex multidimensional view. Model and to develop a model for face recognition is hard work. Face detection there are available several types of different condition database (expression, Lights, etc.) with a different face [6]. There are many ways to facial expressions, such as non-monotonic illumination variation, random noise and changes in age, pretending as though each suffered from the limitations of this method, and to identify conditions of expression. Although some methods, such as Gradient face, a High discrimination power illumination variation, they are still recognized for expression and age variation conditions capabilities [7].

Since from last decade there are number of methods presented for video based or image based human facial expression and emotions recognition. But, the detection accuracy of facial expression is considered as a critical issue. To improve the facial expression detection accuracy, here, SENN based probabilistic classifiers is proposed to detect the facial expression from video. In this process, first, the facial landmarks and face are detected. Then, two types of features are extracted to improve the detection accuracy. Finally, the probabilities of each frame expressions are classified by using SENN. The experimental results show that the proposed scheme attained better results. This paper consists of five sections. Section 2 describes the related work for facial expression by using video, section 3 presents the proposed scheme, section 4 predicts the performance results of proposed scheme and section 5 concludes the paper.

2. Related work

In this segment, the video based facial expression detection or recognition existing schemes has been discussed. Hajati et al., (2017) [8] presented a novel derivative sparse representation method for face as well as texture recognition by means of short-length videos. Initially, it constructs local linear subspaces of dynamic texture segments by calculating spatiotemporal directional derivatives in a cylinder neighborhood inside dynamic textures. Nothing like conventional techniques, a nonbinary texture coding method is presented to excerpt high-order derivatives by means of continuous circular as well as cylinder areas to evade aliasing effects. After that, these local linear subspaces of texture segments are plotted onto a Grassmann manifold through sparse representation. So as to institute the correspondences of subspace points on the manifold for gauging the resemblance amid two dynamic textures, a novel joint sparse representation technique is implemented. Wide-ranging experimentations on the Honda/UCSD, the YouTube, the CMU motion of body, and the DynTex datasets prove that the research technique reliably outdoes the hi-tech approaches in dynamic texture recognition. On the other hand, it contains greater computational complexity.

Chen et al., (2016) [9] studied audio modalities (speech) as well as visual modalities (face images) are used. A novel feature descriptor known as Histogram of Oriented Gradients from Three Orthogonal Planes (HOG-TOP) is introduced to excerpt dynamic textures from video sequences to describe facial appearance modifications. So as to take facial configuration modifications, a novel geometric feature derived from the warp transformation of facial landmarks is presented. Furthermore, the part of audio modalities on recognition is as well opened up in our research. We used the multiple feature fusion for handling the video-based facial expression recognition problem underneath lab-controlled environment and in the wild, correspondingly. Experimentations carried out on the expanded Kohn-Kanada (CK+) database as well as the Acted Facial Expression in Wild (AFEW) 4.0 database prove that our method is vigorous in handling video-based facial expression recognition problem underneath lab-controlled environment and in the wild matched up with the other high-tech techniques. On the other hand, it should enhance accurateness.

Le Nguyen Bao et la., (2016) [6] presented a new face recognition as well as face tracking approach from video streaming. Video frame doesn't display any expression of former localization of a face nor create any recommend regarding the pose. Window like rectangle is drawn by computing top-left, top-right, bottom left, and bottom-right point in face image outline of a video frame. Certain pre-processing mathematical tasks is needed for a video frame for eliminating error. In keeping with counter boundary images production by edges is needed. Then, scalar as well as vector distance amid each and every corner points of two successive frames are identified to track the face location. Corner point's displacement signifies modifications the location as well as face position into the subsequent frame. Not yet assessed underneath real time situations.

Mohammad et al., (2016) [7] was introduced Online video contextual advertisement user-oriented system. Actually this approach was union of networking streaming structure by means of Meta-data structure to video data storage as well as video-based face identification by means of machine learning models from the camera with multimedia communications. The suitable object classes are determining from the real captured images. Those images will be study on predefined set of conditions. Dependent upon the defined object class, the system utilize the multimedia advertising contents database as well as suitable contents was chosen and play mechanically. Furthermore, this method examined previous face identification in video streaming in addition age assessment from face images methods. It may not be accurate under sudden pose variations.

Meng et al., (2016) [10] presented a new two-stage automatic system is presented to uninterruptedly foresee affective dimension values from facial expression videos. In the initial phase, conventional regression approaches are utilized to categorize every individual video frame, in the next phase, a time-delay neural network (TDNN) is presented to design the temporal associations amid successive likelihoods. The two-stage method isolates the emotional state dynamics modeling from a discrete emotional state prediction step dependent upon input features. The temporal information utilized by the TDNN is not prejudiced by the greater variability amid features of successive frames and lets the network to effortlessly exploit the gently varying dynamics amid emotional states. The system was completely assessed on three diverse facial expression video datasets. Our experimentation outcomes prove that the usage of a two-stage method united with the TDNN to consider formerly categorized frames meaningfully enhances the complete performance of unremitting emotional state estimation in naturalistic facial expressions.

Pojala Chiranjeevi et.al (2015) [11] presented different approach based on video facial tracking for the detection of physical fatigue. It was introduced the effective non-contact system for identifying non-localized physical tiredness from utmost muscle activity by means of facial videos attained in a real-world environment with natural lighting where subjects were let to willingly turns their head, modify their facial expression, and change their pose. This technique used a facial feature point identifying technique by gathering a 'good feature to track' as well as a 'supervised descent method' to state the dares that start from real-world scenario. A face quality valuation system was as well integrated in this system for decreasing incorrect outcomes by removing poor quality faces that happened in a video sequence because of problems in head motion, genuine lighting, and pose deviation. But, higher computational cost for both feature extraction and classification.

Hayat & Bennamoun (2014) [12] presented a completely automatic structure that opens up the dynamics of textured 3D videos for identification of six distinct facial expressions. Local video-patches of variable lengths are taken out from plentiful places of the training videos and signified as points on the Grassmannian manifold. An effectual graph-based spectral clustering technique is utilized to group these points for each expression class. By means of a valid Grassmannian kernel function, the ensuing cluster centers are embedded into a Reproducing Kernel Hilbert Space (RKHS) in which six binary SVM models are studied. Specified a query video, we excerpt video-patches from it, signify them as points on the manifold and compare these points with the learnt SVM models subsequent to a voting based approach to choose regarding the class of the query video. The experimentation outcomes on BU4DFE data set prove that the system attains extremely greater classification accurateness for facial expression recognition from 3D videos. Classification Framework is Basic and not evaluated under complex head movements.

Munawar Hayat et al., (2014) introduced a completely automatic system which uses the elements of dynamics of textured 3D videos for distinguishing of six distinct facial expressions. From numerous regions of the training videos, Local video-patches of variable lengths are taken out and provide attentions on the Grassmannian manifold. An effective graphbased spectral clustering algorithm was utilized to independently group these focuses for every expression class. But, it’s not suitable for real time data.

Mostafa et al., (2014) [13] introduced an additional functional structure for the organization of facial expressions produced by human in an impulsive environment seen by video cameras. The important purpose of this method was it organized an alternative database from the databases used for analysis. Training data sets were impulsive and not postured, the databases were publically accessible and sufficiently extensive and machine system was totally robotized and required no supervision. The classifier got was taught on the BU-3DFE image database and assessed on different databases. The grouping precision was discovered to be close in performance to the practically identical stage. But, it has high computational complexity.

Nicolaou et al. (2012) [14] introduced the temporal dependencies over a dimensional domain by prolonging the relevance vector machine regression structure to take the output structure and the covariance inside a preset time window. Eyben et al. (2011) [15] utilized it for audio-visual classification of vocal upsurges in human chat and the outcomes proved noteworthy enhancements over a static method dependent upon SVM. Nicolaou et al. (2010) [16] utilized LSTM networks to outdo SVR because of their capability of learning previous and forthcoming contexts.

In summary, video based approaches could take elusive modifications and temporal trends of facial expression that could not be attained by static image based approaches. Because of the huge number of data in videos, a completely automated technique for analysis is needed. Consequently, in this work, automatic video based facial expression is concentrated.

3. Methodology

In this section, the proposed SENN based automatic detection of video based facial expression or emotion has been discussed. As well as step by step process of proposed scheme is presented in given below section.

1.1. System overview

The overall process of proposed scheme is illustrated in figure 1. SENN support for automated facial expression examination of video data encompasses the subsequent components: 1) identifying landmarks, which describe the facial shape, and tracking landmarks and therefore the facial modifications because of expressions; 2) feature extraction dependent upon these landmarks; 3) formation of classifiers dependent upon excerpted features, and probabilistic classification at every frame of the video sequence; and 4) probabilistic propagation of facial expressions all through the video.

Initially, a face detector as well as a landmark detector is used to locate landmarks in videos. Dependent upon these identified landmarks, the technique additionally excerpts the geometric as well as LBP features to describe the face shape modifications produced by facial expressions. Geometric as well as LBP features are normalized that are illustrated to be vigorous to skin colour as well as enlightenment differences, and are input to facial expression classifiers for examination. For that reason, the third part of the technique is the formation of probabilistic classifiers by means of the mined features. Offline–trained SENN are used for getting the probability of every facial expression. As the probabilistic classifiers define the facial expressions at discrete frames, our framework promulgates the measurements at distinct frames all through videos by means of a sequential Bayesian inference method, to get a depiction of facial expression modifications in the complete video in the form of a temporal probabilistic profile of facial expressions.

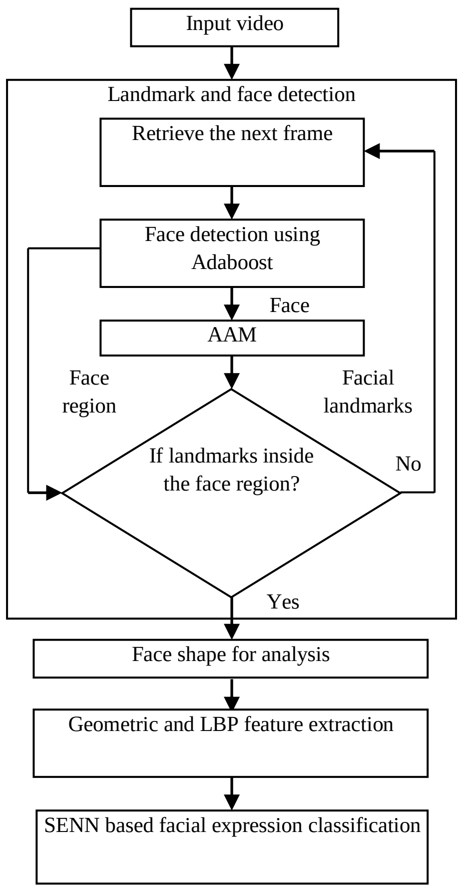


Figure 1: Overview of SENN framework for facial expression analysis using video data

1.2. Landmark detection and tracking in video

Landmark detection as well as tracking technique is provided in this segment. In the research of [17], the face region is physically outlined for getting the deformation amid faces with expression as well as neutral faces for examination. On the other hand, manual labeling is time-taking, and subjected to the person who tags the face. Particularly in our analysis, the video of every participant might encompass diverse facial expressions, up to 10,000 to 20,000 frames. Accordingly, it is a difficult job to physically spot all the face shapes in the videos. An automatic system is anticipated to do the landmark points identification and tracking with least human interference. So as to mechanize the process, we identify the face as well as facial landmarks in the initial frame of the video by means of a face detector as well as an Active Appearance Model (AAM) [18], and after that track the landmark points in the residual frames. Meanwhile, the face detector is running via the video to observe the tracking, as well as re-initializes the tracker while participants' faces are beyond the front view or obstructed while the facial expression analysis could not be carried out. In Figure 1, the complete method is demonstrated.

Face detection

In our method, the face is automatically detected in the first frame of the video. Many face detection methods have been recently developed [19]. Among current methods, the AdaBoost based approaches attain outstanding detection accurateness in addition to real- time speed [20–22]. We used AdaBoost algorithm with Haar features, to detect frontal and near-frontal faces [20]. But, to select the poor hypotheses, AdaBoost must analyse the complete features separately that would dramatically increase the computational time of classification, especially for large scale datasets. With the intention of handling this issue, we present Latent Dirichlet Allocation (LDA) model for enhancing the efficacy of AdaBoost and it's termed as improved adaboost. Latent Dirichlet Allocation (LDA) [23] is a generative probabilistic topic model for groups of textual data and it is extensively utilized for feature reduction.

In this method, critical Haar features are sequentially selected from an over-complete feature set, which may contain more than 45,000 Haar wavelet features. Threshold classifiers are learned from the selected features, and are combined by AdaBoost. With a cascade structure [20], enhanced AdaBoost-based frontal face identification approaches could attain real-time speed (to be exact greater than 15 frames for each second) with accurateness equivalent to other approaches. Note that our face detector aims at detecting only frontal faces, since our facial expression analysis is only applied to frontal faces.

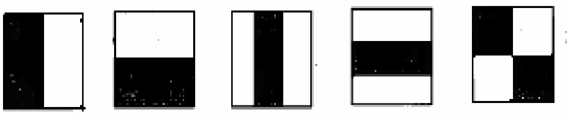


Figure 2: The set of Haar-like features for adaboost

An integral image is the summation of each and every pixel values (in an image) beyond and to the left, comprising itself. Dependent upon the calculated integral image, the Haarlike features are proficiently computed. The pseudocode of the adaboost is specified in the following algorithm 1.

Algorithm 1: Enhanced adaboost based face detection

Input: Number of frames: $(x_1, y_1), \ldots, (x_L, y_L)$, $g_j(x_i)$ is the jth Haar-Like feature of ith example x_i.

Output: Best feature values are selected

1. Initialized the weights

$$W_{l,i} = \begin{cases} \frac{0.5}{m} & i \leq m \\ \frac{0.5}{n} & otherwise \end{cases}$$, where m, n are the number of positive or negative examples respectively

$$for\ t = 1 \ldots T$$

2. Normalize the weights

$$W_{t,i} = \frac{W_{t,i}}{\sum_{j=1}^{L} W_{t,j}}$$

3. For each feature j, train a weak classifier h_j and evaluate its error ε_j with respect to w_t

$$\varepsilon_j = \sum_{l=1}^{L} W_{t,j} \left| h_j(x_i) - y_i \right|$$

$h_j(x_i) = \begin{cases} 1 & p_j g_j(x) < p_j \theta_j \\ 0 & otherwise \end{cases}$, where $p_j \in \{1, -1\}$ is a parity bit and θ_j is a threshold.

4. Choose the feature h_k^j ($t_k as\ the\ pivot\ term$) with the lower error ε_t // LDA step, Pivot term defined found harder to classified by previous classifiers in each iteration

$$H_t^j = \sum_{i=1}^{L} w_t(x_i, g_j) \exp\left(-\alpha_t \phi(x_i, g_j) h^j(x_i)\right)$$

5. Select the feature for which H_t^j is minimum error
6. Update the weights $w_{t+1,i} = w_{t,i} \beta_t^{1-e_i}$, where $e_i = 0$ if feature x_i is classified correctly, $e_i = 1$ otherwise, and $\beta_t = \frac{\varepsilon_t}{1-\varepsilon_t}$.
7. Final selected feature

$H(x) = \begin{cases} 1 & \sum_{t=1}^{T} \alpha_t h_t(x) \geq 0.5 \sum_{t=1}^{T} \alpha_t \\ 0\ otherwise & \end{cases}$, where $\alpha_t = \log(1/\beta_t)$

Landmark detection and tracking

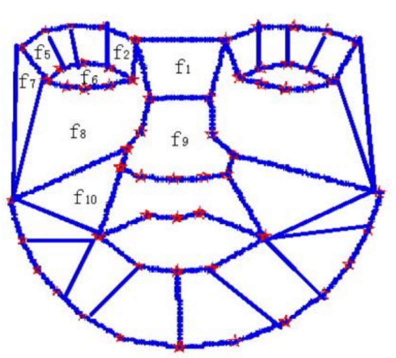
Inside every identified face, our technique additionally finds significant landmarks to describe facial expression modifications. Active appearance model (AAM) [18] situates these landmark points. AAM is a statistical technique to design face appearance as well as face shape. According to AAM, the face shape is denoted by a collection of landmarks, as well as the face texture is the image intensity or colour of the complete face region. AAM face model unites the principal components from face texture as well as shape to validate a vector, and after that employ an added Principal Component Analysis (PCA) to additionally decrease the feature dimensionality. AAM models are learned offline from gathered annotated training samples. In order to situate landmarks in a specified image with unknown faces, an effectual technique is designed in [18] to recognize landmarks in images by reducing the error amid real face and its PCA renovation.

Combination of face detection and landmark detection

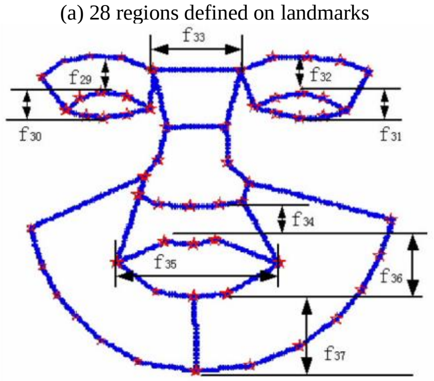
Although participants are instructed to restrict their head movement in the course of data capture, the faces of participants can be beyond the front view occasionally. These cases would be unsuccessful for the period of face tracking and in the facial expression examination. In order to state this issue, face detection is combined with landmark tracking such that landmarks detected can be monitored. The frontal face detector will lose detection when the faces are out of the frontal view or are occluded. Then the AAM tracking will be stopped. The face detector would looking for frames till the face is returned to its front view, or the obstruction is over. After that the AAM tracker is reset within the identified face region. In the experimentations, almost 1.4% of frames in each and every participants have proved non-frontal faces. The faces beyond front view would be omitted from the consequent facial expression analysis.

1.3. Facial expression feature extraction

Geometric features are extracted from landmarks to characterize facial expression modifications. The primary kind of geometric features are the region modifications of 28 regions described by 58 landmark points, as explained in Figure 3(a). These regions define the global modifications produced by facial expressions. There are as well certain facial actions, which are nearly associated to expression modifications. These facial actions include eye opening, mouth opening, mouth corner movement and eyebrow movement. To specifically describe such actions, we define another type of geometric features, which measure distances between some landmark points, as illustrated in Figure 3(b).



(a) 28 regions defined on landmarks



(b) Distance features characterizing expression changes

Figure 3: Geometric features defined on landmarks for expression analysis

LBP based feature extraction

LBP could attain impressive accurateness in pattern recognition areas with a robust texture discrimination ability. The fundamental LBP algorithm encrypts the signs of the pixel variances amid the central pixel and its adjacent pixels in a texture unit (TU). Fig. 4(a) depicts a TU with 3×3 pixels in the LBP algorithm and Fig. 4(b) depicts the LBP coding process in the equivalent TU. g_c is known as the central pixel of the TU and $gi(i = 1, \ldots, 8)$ is its neighbor.

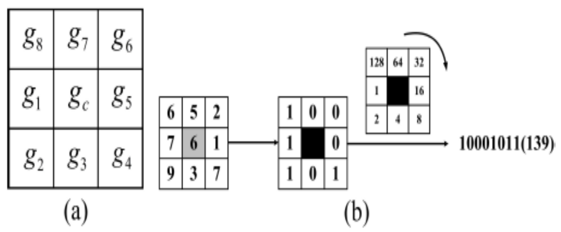


Figure 4: (a) TU with 3×3 pixels, (b) LBP process in the corresponding TU

There are three steps for computing the LBP value of the central pixel g_c. Stage by stage process of LBP is provided along these lines:

Step 1: Encrypt $g_i(i = 1, \ldots 8)$ into a binary value 1 or 0 in keeping with the thresholding function (1).

$$s(g_c, g_i) = \begin{cases} 1, & if\ g_c \geq g_i \\ 0, & if\ g_c < g_i \end{cases}$$

Step 2: weight of every neighbour is defined

Step 3: LBP value is calculated by adding all weighted neighbours

With the aim of removing the consequences of distinct dissimilarities in facial expressions, excerpted features are normalized in numerous means. Initially, every face shape is normalized to the similar scale. We utilize the face width to normalize faces, as it doesn't differ with facial expression modifications. Next, geometric features as well as LBP values are normalized by every subject's neutral faces. E.g., every 2D geometric feature is divided by its equivalent value at the neutral expression of the same person. Consequently, the geometric features replicate the percentage modifications of 2D face geometry, and discrete topological variances are negated. Lastly, each and every feature values are normalized to z-scores for consequent examination.

1.4. SENN based facial expression classifier

With the aim of measuring facial expressions, taken out features are utilized to train facial expression classifiers. We accept SENN as a pattern classification method to train classifiers. The MLP classifier involved in this work being the feature of Artificial Neural Network (ANN), the classifier performance is in turn improved due to the influence on the network size. In case of smaller network size, the MLP classifier tends to possess lesser learning capabilities however if larger results in increased loss due to generalization. The improved and optimized MLP design is therefore presented in this work as given by Evolutionary Computation (EC). The scheme representing the combination of EC and MLP classifier based neural network is termed as Evolutionary Neural Networks (ENNs). To detect the facial expression, the scheme called Simultaneous ENN (SENN) [24] is proposed to optimally design the MLP networks for improved classification. This improvement comprises of the MLP weight concepts to avoid evaluation due to the noisy fitness.

The typical MLP network executes the framework of SENN classification. In MLP neural networks, there are two evolutions, Topology-optimization Evolutionary Neural Network (TENN) and SENN. TENN aims to progress the ideal structure of a MLP to a specified problem. However the problems occur in requirement of a priori architecture must be set beforehand training and as well the non-existence of weight information will be difficult for the fitness evaluation. Hence the SENN classifier has been introduced. In case of SENN classification, optimization of the connection weights is done using Lamarckian evolution by employing the method of random mutation and local search is done by employing the resilient back propagation (RPROP) which is a gradient-based algorithm [25].The classification of data into training and validation sets (TS and VS) is performed by evaluating each individual using the optimal weights set with respect to it random connections carrying the value that trains the MLP in its corresponding genotype. Setting each individual with regard to the initialized population is done by selecting unseen nodes randomly amid 0 and H. Later, every probable connection is made by respective likelihood of p. The creation of new individuals is done by applying the structural mutation operator that operates either by adding or deleting a generated random number of nodes or connections between 1 and M. The evolution of the population consisting of P individuals result in G generations and as election scheme that include changing the fitness values into rankings followed by using the application based on roulette wheel scheme. During each generation, partial of the individuals are stored whereas the restore bred by means of the genetic operator. Among generated individuals, the best one generally saved as the finest one having needed MLP features. The steps involved in the process of classification are given as follows:

Algorithm 2: SENN classification process

1. BEGIN
2. Split the data into the TS, Fitness Set (FS) and Model Selection sets
3. $i \leftarrow 0$
4. $\psi_i \leftarrow$ Initialize the population (i.e. features) and weights of the P MLPs
5. While (i< G-generation)
6. Train the MLPs from ψ_i on TS with L number of RPROP epochs
7. Evaluate the current population ψ_i
8. $B_i \leftarrow$ Select P/2 ancestors from ψ_i for reproduction
9. $O_i \leftarrow$ Apply the macro (50%) or structural (50%) mutations on all elements of A_i
10. $S_i \leftarrow$ Select P/2 − 1 survivors from ψ_i
11. Set the next generation ($\psi_{i+1} \leftarrow best(\psi_i) \cup S_i \cup O_i$)
12. $i \leftarrow i + 1$
13. END

In this procedure, concerning the fitness function, a diverse policy was cast-off here; meanwhile the concurrent development of weights and topologies is actually delicate to over fitting. The training set is signified as TS. Therefore, the validation set is divided into: a Fitness Set (FS), cast-off for the fitness assessment (the RMSE is also used), and a Model Selection set (MS), utilized to first-rate the finest individual at the final part of the procedure. This assessment function is implemented by means of the pseudo-code:

1. BEGIN
2. $j \leftarrow 1$
3. While ($j \leq P_\psi$)
4. $MLP_j \leftarrow$ Select the j-th MLP from the population
5. Set the MLP_j fitness as the RMSE on FS
6. $j \leftarrow j + 1$
7. END

From the above procedure the facial expression is detected for individual frame. By presuming the distance output by SENN as a Gaussian likelihood model, the posterior probabilities are directly fit with the sigmoid function as depicted in Eqn (1):

$$p(Z|x = i) = N(\mu_i, \sigma_i)$$

$$p(x = i|Z) = \frac{1}{1 + \exp(A_i z + B_i)}$$

Here x is known as the class label, Z is called the yield of SENN. The parameters μ_i, σ_i, A_i B_i are guesstimated from training data. We are fascinated in the probability $p(Z|x)$ that is utilized for the later Bayesian probability propagation. The label x denotes the facial expression, and Z is known as the SENN outcome of excerpted features. The class label x considers discrete values, that is to say, $x = i$ specifies the presence of i-th facial expression. The probabilistic outcomes of facial expression classifiers, $p(z\,|x_i)$, model facial expressions at distinct frames, conversely have not completely utilized the temporal information of facial expressions in videos. In order to promulgate the posterior probabilities of facial expressions all through the whole video, we utilize a sequential Bayesian estimation. The sequential Bayesian estimation as well as its Monte Carlo derivations are extensively utilized in visual tracking [27, 28], since they could manage sequential inference problems efficiently. Based on the above principle, the facial expressions are classified as happy, sad, anger, and fear emotions.

4. Results and discussion

In this section, the proposed SENN performance has been evaluated by using the widely used YouTube face video database [28], and compared with existing recognition schemes like TDNN-SVR [10] and SVR [15]. The performance is measured in terms of recognition accuracy, sensitivity, specificity, precision, recall, and F-measure.

Performance evaluation criteria for facial expression classification

In this experimental study, the performance is mentioned by the values of accuracy, sensitivity and specificity. These are measured the values of true positive rate(tr_p), true negative rate (tr_n), false positive rate (fa_p)and false negative rate (fa_n). The sensitivity measured by using the formula of $sen = \left(\frac{tr_p}{tr_p+fa_n}\right)$.

The sensitivity measured by using the formula of $spec = \left(\frac{tr_n}{tr_n+fa_p}\right)$.

The accuracy of proposed SENN classifier has been calculated by using ROC analysis. In this ROC, the x and y- axis is denoted as the given values

$$x = false\ positive\ rate(1 - spec)$$

$$y = true\ positive\ rate\ (sen)$$

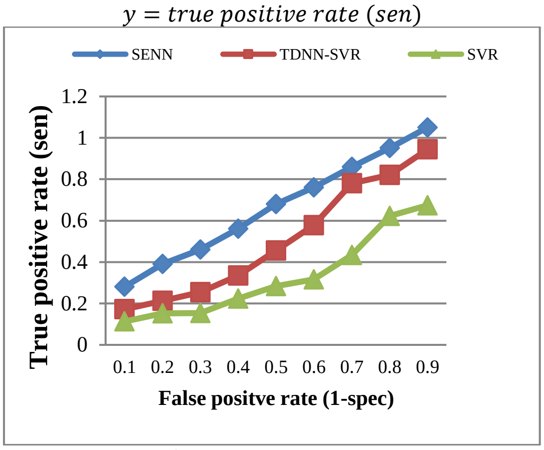


Figure 5: ROC curve performance comparison among various recognition schemes

The graphical representation of ROC performance comparison for proposed SENN and existing TDNN-SVR and SVR is showed in figure 5. It illustrates, the performance of SENN attained high accuracy compared than existing algorithms. Due to the efficient face detection and tracking with effectual feature extraction, the proposed scheme has best results like sensitivity value of 90.11% and specificity of 96.12%. Numerical evaluation results are showed in table 1.

Table 1: ROC Numerical Evaluation Results among Various Classifiers

(1-spec)	**SENN**	TDNN-SVR	**SVR**
0.1	0.28	0.172	0.113
0.2	0.39	0.213	0.152
0.3	0.46	0.254	0.153
0.4	0.56	0.335	0.224
0.5	0.68	0.456	0.284
0.6	0.76	0.577	0.316
0.7	0.86	0.781	0.434
0.8	0.95	0.821	0.623
0.9	1.05	0.944	0.674

Performance Comparison of processing time

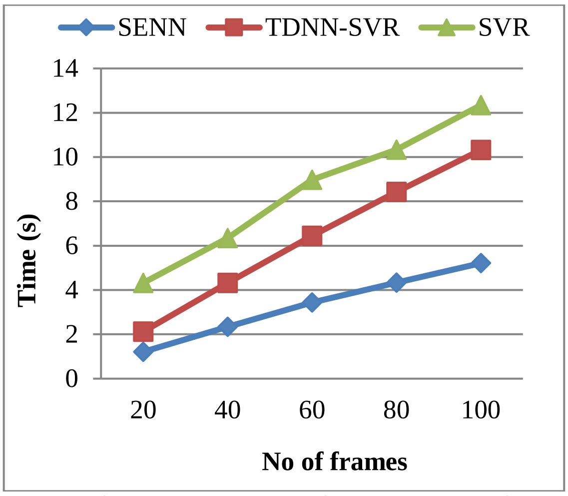


Figure 6: Performance comparison of processing time for various classifiers

The graphical representation of processing time performance comparison for proposed SENN and existing TDNN-SVR and SVR is showed in figure 6. It illustrates, the proposed SENN has take less processing time to classify the FER compared than existing algorithms. Due to the less training time and efficient features, the proposed scheme attained less time. Numerical evaluation results are showed in table 2.

Table 2: Processing Time Numerical Evaluation Results for All Classifiers

No of frames	SENN	TDNN-SVR	SVR
20	1.21	2.12	4.32
40	2.34	4.32	6.34
60	3.43	6.45	8.98
80	4.34	8.43	10.34
100	5.21	10.32	12.34

Performance comparison of accuracy, sensitivity, specificity

The graphical representation of Accuracy, sensitivity, specificity performance comparison for proposed SENN and existing TDNN-SVR and SVR is showed in figure 7. It illustrates, the proposed SENN produces high accuracy, sensitivity and specificity results compared than existing algorithms. Due to the high true positive rate of proposed scheme and the efficient classification, the proposed scheme attained high accuracy, sensitivity and specificity performance result of 95.84%, 90.12% and 96.12%. Numerical evaluation results are showed in table 3.

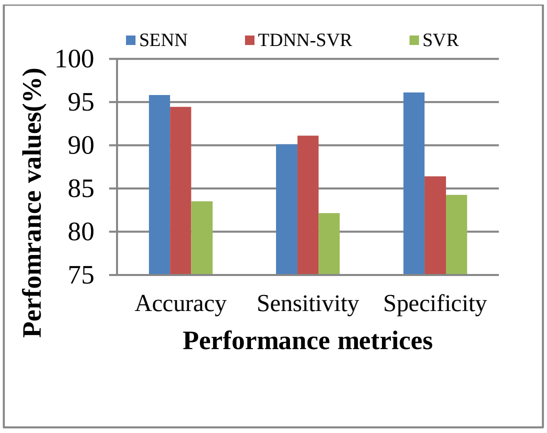


Figure 7: Performance comparison of Accuracy, sensitivity, specificity for various classifiers

Table 3: Accuracy, Sensitivity, Specificity Numerical Evaluation Results for All Classifiers

Performance matrices	SENN	TDNN-SVR	SVR
Accuracy	95.82	94.43	83.54
Sensitivity	90.11	91.12	82.15
Specificity	96.12	86.43	84.26

Performance comparison of precision, recall and F-measure

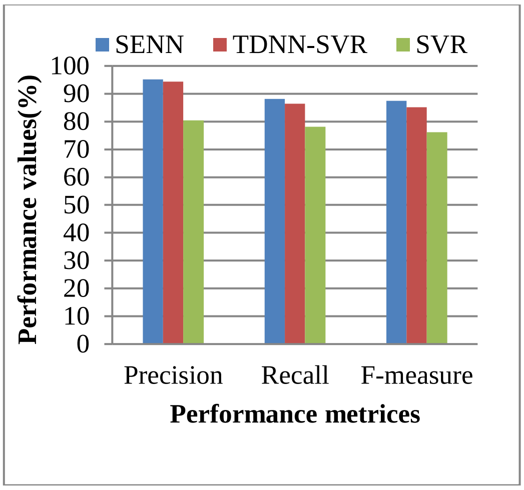


Figure 8: Performance comparison of precision, recall and F-measure for various classifiers

The graphical representation of precision, recall and F-measure performance comparison for proposed SENN and existing TDNN-SVR and SVR is showed in figure 8. It illustrates, the proposed SENN produces high precision, recall and F-measure results compared than existing algorithms. Due to the less negative rate and high positive prediction rate, the proposed SENN attained precision, recall and F-measure performance results of 95.1%, 88.14 % and 87.45%. Numerical evaluation results are showed in table 4.

Table 4: Precision, recall and F-measure numerical evaluation results for all classifiers

Performance matrices	SENN	TDNN-SVR	SVR
Precision	95.12	94.34	80.42
Recall	88.13	86.36	78.13
F-measure	87.46	85.18	76.14

5. Conclusion

In this research, we introduced an automated computational structure for examining facial expressions by means of video data, creating a probabilistic profile of expression change via SENN. The structure discovers gorgeous info presented in the video, by giving a probabilistic composition of every frame of the sequence, thus emphasizing elusive changes in addition to the likelihood of a combination of emotions. In this process, initially, the face as well as landmarks are detected and tracked from video data by using enhance adaboost and AAM model. Then, the effectual geographical and LBP features are extracted from face landmarks to detect the facial expression. Finally, the SENN is used to classify the facial expression for individual frames and to analyse throughout video; the Bayesian interface scheme is presented. The experimental results show that the performance of proposed SENN attained high detection accuracy rate of 95.82%, sensitivity rate of 90.11%, specificity rate of 96.12%, precision rate of 95.12%, recall rate of 88.13% and F-measure rate of 87.46% compared than existing scheme like TDNN-SVR and SVR. In future, an attempt can be made to develop hybrid approach for facial feature extraction and recognition accuracy can be further improved using same NN approach and hybrid approaches with other medical databases.

References

[1] Samad R & Sawada H, "Edge-based Facial Feature Extraction Using Gabor Wavelet and Convolution Filters", *MVA*, pp.430-433, (2011).

[2] Thai LH, Nguyen NDT & Hai TS, "A facial expression classification system integrating canny, principal component analysis and artificial neural network", *International Journal of Machine Learning and Computing*, Vol.1, No.4, pp.388-393, (2011).

[3] Sisodia P, Akhilesh V & Sachin K, "Human Facial Expression Recognition using Gabor Filter Bank with Minimum Number of Feature Vectors", *International Journal of Applied Information Systems*, Vol.5, No. 9, pp.9-13, (2013).

[4] Jun Y & Zengfu W, "A Video-Based Facial Motion Tracking and Expression Recognition System", *Springer Science Business Media New York*, (2016).

[5] Dey A, "Contour based Procedure for Face Detection and Tracking from Video", *3rd Int'I Conf. on Recent Advances in Information Technology I RAIT*, (2016).

[6] Le DN, Van Chung L & Nguyen GN, "Performance evaluation of video-based face recognition approaches for online video contextual advertisement user-oriented system", *Information Systems Design and Intelligent Applications*, pp.287-295, (2016).

[7] Haque MA, Irani R, Nasrollahi K & Moeslund TB, "Facial video-based detection of physical fatigue for maximal muscle activity", *IET Computer Vision*, (2016).

[8] Hajati F, Tavakolian M, Gheisari S, Gao Y & Mian AS, "Dynamic Texture Comparison Using Derivative Sparse Representation: Application to Video-Based Face Recognition", *IEEE Transactions on Human-Machine Systems*, (2017).

[9] Chen J, Chen Z, Chi Z & Fu H, "Facial expression recognition in video with multiple feature fusion", I*EEE Transactions on Affective Computing*, (2016).

[10] Meng H, Bianchi-Berthouze N, Deng Y, Cheng J & Cosmas JP, "Time-delay neural network for continuous emotional dimension prediction from facial expression sequences", I*EEE transactions on cybernetics*, Vol.46, No.4, pp.916-929, (2016).

[11] Chiranjeevi P, Gopalakrishnan V & Moogi P, "Neutral face classification using personalized appearance models for fast and robust emotion detection", *IEEE Transactions on Image Processing*, Vol.24, No.9, pp.2701-2711, (2015).

[12] Hayat M & Bennamoun M, "An automatic framework for textured 3D video-based facial expression recognition", *IEEE Transactions on Affective Computing*, Vol.5, No.3, pp.301-313, (2014).

[13] El Meguid MKA & Levine MD, "Fully Automated Recognition of Spontaneous Facial Expressions in Videos Using Random Forest Classifiers", *Affective Computing*, Vol.5, pp.418-431, (2014).

[14] Nicolaou MA, Gunes H & Pantic M, "Output-associative RVM regression for dimensional and continuous emotion prediction", *Image Vis. Comput. Best Autom. Face Gesture Recognit.*, Vol.30, No.3, pp.186–196, (2012).

[15] Eyben F, Petridis S, Schuller B, Tzimiropoulos G, Zafeiriou S & Pantic M, "Audiovisual classification of vocal outbursts in human conversation using long-short-term memory networks", *IEEE International Conference on Acoustics, Speech and Signal Processing*, pp.5844-5847, (2011).

[16] Nicolaou MA, Gunes H & Pantic M, "Continuous prediction of spontaneous affect from multiple cues and modalities in valencearousal space", *IEEE Trans. Affect. Comput.*, Vol.2, No2, pp. 92–105, (2011).

[17] Christopher ACK, Barrett F, Gur RE, Gur RC & Verma R, "Computerized Measurement of Facial Expression of Emotions in Schizophrenia", *Journal of Neuroscience Methods,* (2007).

[18] Cootes TF, Edwards GJ & Taylor CJ, "Active Appearance Models", *IEEE Trans on PAMI*, Vol.23, No.6, pp.681–685, (2001).

[19] Yang MH, Kriegman DJ & Ahuja N, "Detecting Faces in Images: A Survey", *IEEE Trans on PAMI*, Vol.24, No.1, pp.34–58, (2002).

[20] Viola P & Jones M, "Robust Real-time Object Detection", *International Journal of Computer Vision*, Vol.57, No.2, pp.137–154, (2004).

[21] Li SZ & Zhang Z, "FloatBoost Learning and Statistical Face Detection", *IEEE Trans on PAMI*, Vol.26, No.9, pp.1112–1123, (2004).

[22] Wang P & Ji Q, "Learning Discriminant Features for Multi-View Face and Eye Detection", *Computer Vision and Image Understanding*, Vol.105, No.2, pp.99–111, (2007).

[23] Blei DM, Ng AY & Jordan MI, "Latent dirichlet allocation", J. *Mach. Learn.Res*., Vol.3, pp.993–1022, (2003).

[24] Rocha M, Cortez P & Neves J, "Simultaneous evolution of neural network topologies and weights for classification and regression", *International Work-Conference on Artificial Neural Networks*, pp. 59-66, (2005).

[25] Riedmiller M, "Advanced supervised learning in multi-layer perceptrons-from backpropagation to adaptive learning algorithms", *Computer Standards & Interfaces*, Vol.16, No.3, pp. 265-278, (1994).

[26] Andreasen, NC, *Scale for the Assessment of Negative Symptoms (SANS)*, Iowa City: University of Iowa, (1984).

[27] Yeasin M, Bullot B & Sharma R, "From facial expression to level of interest: a spatio-temporal approach", I*EEE Conference on Computer Vision and Pattern Recognition*, (2004).

[28] Wolf L, Hassner T & Maoz I, "Face recognition in unconstrained videos with matched background similarity", *Proc. IEEE Comput. Soc.Conf. Comput. Vis. Pattern Recognit.*, pp.529–534, (2011).