A hybrid facial features extraction-based classification framework for typhlotic people

SaiTeja Chopparapu, Beatrice Seventline Joseph
Department of Electrical Electronics and Communication Engineering, GITAM (Deemed to be University), Visakhapatnam, India

ABSTRACT

Facial features play a vital role in the real-time cloud-based applications. Since, most of the conventional models are difficult to detect heterogeneous facial features due to high computational memory and time for the internet of things (IoT) based video surveillance mechanisms. Video based facial features identification and extraction include a large number of candidates features which are difficult to detect the contextual similarity of the facial key points due to noise and computational memory. In order to resolve these issues, a hybrid multiple features extraction measures are implemented on the real-time video dataset to extract key points using the cloud-based classifier. In this work, a hybrid classifier is used to classify the key facial points in the cloud computing environment. Experimental results show that the proposed hybrid multiple feature extraction-based frameworks have better computational efficiency in terms of error rate, recall, precision, and accuracy than the conventional models.

Corresponding Author:

SaiTeja Chopparapu
Department of Electrical Electronics and Communication Engineering, GITAM (Deemed to be University)
Visakhapatnam 530045, India
Email: saiteja.chopparapu960@gmail.com

1. INTRODUCTION

In recent years, large data sets and the computing power offered by graphics processing units (GPUs) have been motivated by research into deep-learning algorithms that have shown excellent performance in various computer vision tasks and achieved a decisive action over traditional method. The fundamental concept of cloud computing is that user data are not recorded locally, but are placed in the data center of the internet. These data centers could be managed and maintained by the companies that provide cloud computing services. Users can access the stored data at any time through any internet-connected terminal equipment using the application programming interface (API) provided by cloud providers. The immense growth of social media, ecommerce traffic and various web services has significantly raised the need for computational services [1]. One of the heuristically approaches of optimization for action extraction is ant colony optimization (ACO) along with particle swarm optimization. Particle swarm optimization (PSO) optimizes a problem by preserving a particle population and transferring these particles into the search area. Both the ACO and PSO algorithms use one-classification rules sequence of covering patterns. The construction in machine learning models was an active research topic. The main learning machine models to learn group classifiers in high-dimensional datasets are boosting, bagging or stacking [2]. Feature extraction is the essential component in the IoT based authentication frameworks. Since, different facial features have different orientations, texture, and intensities, it is difficult to find the essential features in the large datasets. Feature selection algorithms can be classed into adaptive, statistical, and semi-supervised supervised search models, that is, wrapper models, and embedded

models could be used to describe all feature selection techniques. The feature selection model is different from classifier learning, as it does not assume that the learning algorithm is biased.

The degree of uncertainty in this will be dependent on varies according to the nature of the training data, its variability, dependency, and interdependence [3]. A deep learning algorithm designed for the operation of two-dimensional image data, the convolution neural network (CNN). The architecture of CNN is different. The two main components are convolution and pooling layers. Although it seems simple, these elements could be arranged infinitely. The tuning of hyper-parameters is the main process that affects the prediction performance for the CNN model (including its architecture and parameters) [4], [5]. The tradition is that the network's performance is tested manually when the value of each parameter is changed while the other is retained. In particular, when the data set is large and the available resources are limited, this is computationally expensive. The success of using machine teaching methods for every prediction depends on the best architecture to be found and on the hyper parameters to be tailored to the given problem and produce an exact result. This included many proposals, redundant suggestions, many fake positives, and difficult to collect representative semantic information in complex contexts. These limitations included many of the proposals [6]. The rapid development of profound learning resulted in large margins for detecting deep learning algorithms for objects over traditional feature extraction algorithms. Several studies have been performed for the detection of action in the ambient assisted living (AAL) environment [6]. Traditional approaches to object detection are generally based on manufactured properties for the location of objects in each image. Three major steps are often taken in these methods: proposals, extraction and classification. This feature vectors were usually encoded in low-level visual descriptors like scale invariant feature transform (SIFT) [7], Haar [8], histogram of oriented gradient (HOG) [9] or speeded-up robust features (SURF) [10], showing certain scale, light and rotational variance robustness. During the classification phase, categorical labels are assigned to the regions covered. Methods for classification include support vector machine (SVM) and AdaBoost. Although traditional methods have shown good performance in many benchmarks' public datasets, in difficult conditions they still have many restrictions. During classification, this resulted in many fake positives. Secondly, feature designers are hand-made based on low-level visual indicators, making it difficult, in complex conditions, to capture representative semantical information. Finally, each detection pipeline step is separately designed and optimized and therefore the entire system cannot be provided with an optimal overall solution. Following the success of the application of profound CNNs for image classification, object detection has also made significant progress based on deep CNNs. Here, wrappers take advantage of learning techniques that highlight the most attractive features. Supervised feature selection increases classification efficiency while simultaneously by reducing computer processing time. In recent years, filter-based feature selection criteria have been devised, e.g. fisher score, trace ratio feature, and relief correlation feature (CFS) [11]. Feature selection improves the raw patterns. Feature selection algorithms fall into three categories: wrapper, filter, and hybrid, depending on their approaches. While the wrapper technique focuses on features and uses a statistical measure to identify significance. While the filter technique is quicker, the wrapper technique is more accurate. Combination of exploration and exploitation with meta-heuristic techniques leads to an effective result there are diverse methodological approaches suggested to mimic any natural phenomenon or process exploration search balancing with exploitation governs the ability of such techniques to avoid local optimal and global optimal values [12].

2. RELATED WORKS

Feature extraction is a basic step that directly affects the outcome of recognition systems-a poor choice of descriptor can considerably degrade performance and precision. Finding the relevant descriptor is based on trial-and-error method and the large number of features in dataset. One of the primary benefits of feature learning techniques in relation to handcrafted extraction is the generalization of the feature space in the same visual domain. In hexagonal-volume local binary pattern (H-VLBP), the binary pattern histogram encodes local volumes [13]. Despite its simplicity, the number of separate patterns generated in neighbourhoods' regions by VLBP may become overwhelming. The convolution architecture efficiently uses the image structure by "pooling" and "weight-sharing" to reduce the search space of the network. Pooling and weights initialization help to achieve robustness across differences in scale and space. To optimize this issue, they introduce 3D convolution networks. Traditional models are focused on constructing efficient descriptors or characteristics and then classifying them based on matching features. Here, feature selection measures or filters are used to recognize various key features in anomaly detection process. Global features include silhouette-based descriptors, edge-based features, optical flow-based display, and movement history image (MHI), are used in CNN models [14]. Occlusions, changing viewpoints, and noise often create problem in global features. Local characteristics always use image patches separately, and then these patches are combined to create a space-time model such as SURF and HOG [15]. Local descriptors, particularly for noise images and partly occluded images more efficiently. CNNs have proven to be strong feature extraction model for still image recognition. The recent interest in one-stage methods has renewed between the single shot

detector (SSD) and you only look once (YOLO) [16]. These sensors are tailored for speed but their precision is based on two-stage methods. SSD uses multi-layer defaults within ConvNet from boxes of different scales and forces each layer to concentrate on a prediction of objects of a particular scale. MS-CNN applies multi-layer devolution on ConvNet to increase the map resolution before using the layers to learn regional proposals and pool features to improve detection precision on multi-level layers. Most CNN models are designed on the basis of convolution and pooling intakes, then sampling regularly heights and width while increasing the number of function maps [17]. InDenseNet, dense connected in networks. DenseNet is somewhat logical behind reset transmission information of a layer to another, but in densenet, each map of the functionality of every layer is linked to the input of each subsequent layer in a dense building block this is the most commonly used nonlinear down sampling strategy for translation invariance [18]. This enables next layers to directly access previous layer functions and allows reuse of network features. DenseNet's building block is the dense blocks. Each dense block has several overlapping layers.

A dense block is followed by a transition layer and even the next dense block is its output. The advantages of DenseNet are several: solving the problem of the flattening gradient, stimulating the proliferation and reuse of features and reducing parameters because redundant function maps are not necessary. Mobile net: architectural lightweight. It uses profoundly separable convolutions which basically mean it does not merge and flatten all 3, but performs a single convolution on each channel. It permits filtering of input channels. Deep, smart, separable convolution reduces network complexity and model sizes that are suitable for mobile or low-computing devices [19], [20] conduct studies on biometrics using a smartphone gyroscope to detect users. Touches, face patterns and phone alignment on smartphones have been used to present an unrestricted, implicit biometric multimodal system. The 95 subjects were chosen for the collection of various touch and phone movement patterns for mobile multimodal data set. The results have been shown to be accurate and to improve usability and security. Huang *et al*. [21] developed and proposed a biometric recognition multimodal system using two distinct biometric features, that of the back and the palm veins. The binarization technique was initially used for image preprocessing. The researchers then employed the morphological dilation method, together with an average filter for image smoothening, to remove smaller, undecoupled objects. Filtering and thresholding were performed in two steps for the extraction of features. The researchers then used the model predictive control (MPC)-method and K-means for the extraction and matching of the LBP and template processes. The MPC is the MPC method for extracting features. For modeling matching with a relative 1.6d side length the energy efficiency ratio (EER) value was 0.01965%, while there was 0.058% for the matches LBP matching mechanism to a relative 1.5d side length. A CASIA v1.0 dataset was used for research. The competitive valley hand detection (CHVD) ROI removal process was employed. After that the features were extracted using one of 3 different methods: LBP, 2D local binary (2DLBP) pattern or a combination of the features (LPB and 2DLBP). The researcher used principal components analysis (PCA) and selected the featuring reduction technique using the liner discrimination analysis (LDA) approach. The main component features were used by the researchers. The CASIA palm printing database for the determination of ED-like results was used for experiments. After the application of the third process (LBP+2DLBP) they observed 98.55% precision. Although this new approach for combining palm printing was used the scientists felt it could deliver accurate results for venous patterns. Koley *et al*. [22] addresses the issue of users' authentication of biometric facials. Proposed authentication neural network model based on a two-layered perceptron with input neurons 90, hidden neurons 10 and output neurons 4. The specified network architecture parameters have been experimentally calculated. It is noted. Input parameters include the geometry of local features: co-ordinate (X), coordination (Y), vector direction (Q). Local 30 features are commonly used. The neural network classifier of selection deviations is shown experimentally to have the level of mistake of the first type of 5.2% and the level of error of the second kind-0%. On that basis, the model of the built neural network in conjunction with other technologies of biometric authentication is argued. Similar results are described by [23] that further characterize the mathematical system on which the operation of two-layer perceptron is based. The use of the neural networks in the facial identification system is covered in [24]. It is shown that difficulties in identifying are because the image generated by the scanner may differ slightly during each scan during the facial scanning process. For these problems, a multi-layered perceptron with a hidden layer is proposed to use the neuronal network model. Its simplification and approval explain the choice of neural network architecture. A facial image of 188×240 pixels is the source of input information for the neural network model. 12 geometrical moments, each corresponding to one of the input neurons, are calculated. The number of neurons produced is six. In the theorem of Kolmogorov, the numbers of hidden neurons are 25. As an activation feature, Sigmoid is used. The network has been trained using the conventional back spreading algorithm. There were 100 printed samples. It is 1,000 times of training. The accuracy claimed of training recognition examples was 100%, indicating the perspectives of facial recognition networks in the neural network, according to [25] that facials made by different individuals may be equal in global features, but that it is impossible for them to be equal in

local features (minutiae). Consequently, the process of identification usually comprises two steps. Initially, facials should be classified according to global criteria by dividing them into classes using databases. The second step is the identification of the facial on the basis of a structural comparison and the coincidence factor of the detailed points.

The proposed algorithms for facial images classification based on the Gabor filter application, transform the wavelet Haar, Daubechies and multi-scale of the neural network according to type of models. Numerical experiments are performed and the results are presented for the proposed algorithms. A five-level Daubechies transform wavelet and a multi-layer neural network-type double-layer positron algorithm based on the combined application of Gabor filters, is shown to achieve a classification precision of about 75%. Liang *et al*. [26] a research of facial image methods based on neural networks, such as the two-layer architecture for perceptions are used. As input parameters of the neural network, the module and the argument of a vector field in the image gradient are used. The conclusion is drawn about the need to increase the neural network input vector power to 400. Padol and Yadav [27] details the local and global features of the biometric authentication systems facial. It is demonstrated that it depends largely on the quality of the facial image to distinguish the features which can be used in the future in the identification process. The standard facial scanners are indicated as providing a 500-dpi resolution, the image is characterized by a luminosity level of 256, and a maximum vertical rotation angle of 15 degrees is indicated. At the same time, the end points where the papillary lines end "distinctly" and the branch points where the papillary lines are bifurcated, as characteristic features are proposed. Please note that the images toward the surface having a resolution approximately about 1,000 dpi are possible to identify, detect or reap in greater detail the internal composition on papillary lines (glands of the suds), using that finger surface properties and enable a significant improvement with the accuracy of identification. The level of technical support currently available for common biometric authentication systems doesn't however allow images of this type to be obtained. Reza *et al*. [28] take into consideration the technology for the design and operation of the two-tier facial network recognition system. The allocation of features is done at the first level and the analysis of features chosen is carried out at the second level as a result of which a user is identified. The accuracy of the selection of informational features of facials, which in turn depend largely on the quality of the recognized image, has been demonstrated to affect the performance of the recognition system. This is why a module is present in the system to improve the quality of the original (from the scanner) image. The feature of this module is that the quality of the facial structure should not be damaged by minutiae. To that end, we propose Gabor filters that make the gray print a white image and only 1-pixel wide skeletonization. To this end, we propose to use Gabor filters. Neural networks are used in the proposed system at both detection levels. The hidden layer of 200 neurons and the transfer function is a symmetrical Sigmoid, each in which there is a two-layer perception. The neuron output for the two-layer sensor at the 1st level recognition is 5 and the 2nd level for the same sensor. The empirical nature of the determination of the parameter is noted. The numerical results show 92% accuracy in recognition. It is stated that the use of deep neural networks and a parallel of computer learning and reconnaissance processes are part of further research. The effectiveness of facial clusters in relation to probability of fluid grades based on modern neural networks is compared [29]. The main condition for the study is that deeper neural networks can be used for graphic image analysis. A multilayered sensor with 3 hidden neuron layers was used as an underlying model. Pre-training perceptron implemented with sparse auto encoder. The results show that the change in the number of hidden neurons from 200 to 1,250 is not related to the exactness of recognition of around 93%. Simultaneously, some 97% had higher results for a classification using a fuzzy classifier. The authors could conclude that neural network methods are ineffective in facial recognition. At the same time the article does not support the experimental plan to accurately detect the structural parameters of neural network models. It also raises the question of the suitability of the pre-training process, which is critical when labeled examples of training are not enough and the logistical function has been used [30], a multimodal biometric system has been developed with face iris and ear mode. Issues of the traditional approaches: i) traditional binary classification approaches use static feature extraction measures for facial features analysis, ii) traditional approaches require high computational time for large number of candidate facial features, and iii) binary classification uses limited features space in order to predict facial features.

3. PROPOSED MODEL
3.1. Multi-level facial feature extraction framework

Most of the traditional frameworks are used to filter facial features feasibility with convolution kernels of 3×3×3. Here the most essential data security features are found in different convolution layers, maximum pooling and filters. Using soft max activation function as shown in Figure 1, the completely connected layer is used to filter the essential functions in the image. These characteristics are used to classify biometric characteristics with the proposed classification model.

Figure 1 describes the overall framework of the proposed approach for multi-modal facial features extraction and classification process. In this approach, different facial features are taken as input for feature extraction process. In this work, essential key features are extraction using facial key points extraction approach, scalable gradient-based features extraction, facial features using curvature measures, log inverse differential moment, and max correlation measures. These set of feature extraction measures are used to extract different facial key candidate sets in the framework. There are several layers in which the hidden layers between the input and the output layer are completely connected. However, the main issue is the prediction of the key characteristics because of their high size. In order to overcome this challenge, a neural network model based on CNNs architecture is used for larger applications in the field of image processing. Rotation, translation or scaling nodes for group layers are employed in computer image tasks to model an object in a different patch or dimension. With these connections, the network will develop although the input connection is static and the nearby units of each network device are heavily influenced. The proposed C3D network is used to locate the low-level features and filter out the key features of each image. In the framework, these feature extraction measures along with convolution filters are used to filter the essential key feature sets for the fully connected layer. Filtered features in the fully connected layer are given to ensemble classification approach for multi-level classification process. In this work, a novel kernel function optimization based SVM model is proposed in order to predict the multi-level facial features in order to improve the true positive and error rate.

3.2. Facial feature measures

In the facial features extraction, essential key points are extracted using the proposed key point's extraction method. The user's facial expressions are evaluated using the proposed key point's extraction process. The existing dynamic chaotic map used only two parameters α and β. Moreover, the chaotic region can easily be predictable as the weighted parameters are fixed and as rare ranging from 0 to N.

3.2.1. Facial key points extraction

The steps for extracting facial key points are presented as follows:

Step 1: to each frame in the video file.

Step 2: initialize each frame to $V_I(x, y)$ for key features extraction.

Step 3: apply image normalization using (1)-(5):

$$r_i = F(r_{i-1}, \lambda) = \begin{cases} \dfrac{r_{i-1}}{\lambda}, 0 < r_{i-1} < \lambda \\ \dfrac{(r_{i-1} - \lambda)}{(0.5 - \lambda)}, \lambda \le r_{i-1} < 0.5 \\ F(1 - r_{i-1}, \lambda), 0.5 \le r_{i-1} < 1 \end{cases}$$

$$V_N(x_m, y_n) = \frac{1}{|W||H|} \sum_{m=0} \sum_{n=0} V_I(x_m, y_n) \tag{1}$$

$$\sigma^2_{V_I(x_m, y_n)} = \frac{1}{|W||H|} \sum_{m=0}^{|W|-1} \sum_{n=0}^{|H|-1} (V_I(x_m, y_n) - V_N(x_m, y_n))^2 \tag{2}$$

$$N_{V_I(x_m, y_n)} = \mu_1 + r_i \frac{\sqrt{v_1}}{\sigma_{F_I(x_m, y_n)}} (V_I(x_m, y_n) - V_N(x_m, y_n)); if V_I(x_m, y_n) >= V_N(x_m, y_n) \tag{3}$$

$$N_{V_I(x_m, y_n)} = \mu_2 - r_i \frac{\sqrt{v_2}}{\sigma_{F_I(x_m, y_n)}} (V_I(x_m, y_n) - V_N(x_m, y_n)); if V_I(x_m, y_n) < V_N(x_m, y_n) \tag{4}$$

$$NV(x_m, y_n) = round(\frac{N_{V_I(x_m, y_n)} - Nm_{n\,min}}{N_{V_I(x_m, y_n)} - Nm_{n\,max} \cdot \chi}) \tag{5}$$

Where χ=(Max-0)=(255-0)=255, $\mu_1, \mu_2 \in (120, 150) v_1, v_2 \in (250, 255)$.

Step 4: apply the facial pattern scaling filter using (5) as (6):

$$S(x, y) = \frac{1}{2\pi\sigma_X\sigma_Y} log((\frac{NV(x,y)_X^2}{\sigma^2_{NV(x,y)}} + \frac{NV(x,y)_Y^2}{\sigma^2_{NV(x,y)}}))(NV(x,y)_X) \tag{6}$$

Step 5: in this step, different gradient features are identified using the scaled image S(x,y) as (7):

$$D_{S(x,y)} = \sigma^2 \nabla^2 G(S(x,y), \sigma^2) \tag{7}$$

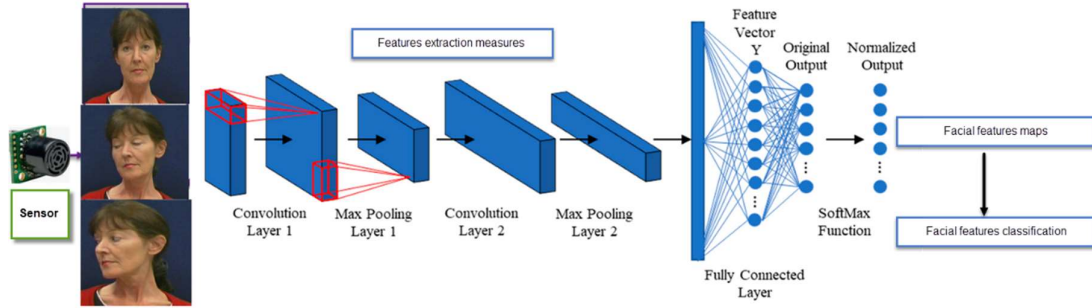Where ∇ is gradient and G is the gaussian function.

Figure 1. Facial feature extraction-based framework

3.2.2. Facial eye and lip curvatures

Let a_x, b_y are the major and minor axis of facial image.

$$\text{Iris gabor filter}(p, q) = \exp\left(-\frac{p'^2 + \gamma^2 q'^2}{2\sigma^2}\right)\cos\left(2\pi \frac{x^2}{\lambda} + \psi\right)$$

where

$$p' = p\cos(\theta) + q\sin(\theta)$$

$$q' = q\cos(\theta) - p\sin(\theta)$$

where λ is the sinusoidal function wavelength in $f = \frac{1}{\lambda}$, θ is Gabor filter orientation, $\psi = 90°$ is a true value of Gabor filter phase offset, $\psi = 0$ is the fanciful value of Gabor filter, σ is the bandwidth and γ is the aspect ratio. Constraints to check the iris curvilinear minimum and maximum structures are show in. Minimum curvature patterns of the image are computed as

$$min\ a_x\ b_y\ min\{\frac{e^{-\sqrt{a_x^2+b_y^2}/2.\sigma}}{a_x.\sqrt{2.\pi}}\ \frac{e^{-\sqrt{a_x^2+b_y^2}/2.\sigma}}{b_y.\sqrt{2.\pi}}\ G$$

Maximum curvature patterns of the image are computed as (8):

$$max\ a_x\ b_y\ max\{\frac{e^{-\sqrt{\frac{a_x^2+b_y^2}{2}}.\sigma}}{a_x.\sqrt{2.\pi}}\ \frac{e^{-\sqrt{\frac{a_x^2+b_y^2}{2}}.\sigma}}{b_y.\sqrt{2.\pi}} \tag{8}$$

Different models are generated using the minimum and maximum curvature of the image in this feature extraction procedure.

3.2.3. Log inverse differential moment

Log inverse differential moment (LIDM) is used to find the homogeneity of the image structures. The normalization factor $(1 + (m1 - m2)^2)^{-1}$ is used to find the small regions from the heterogeneous areas at (m^1 and m^2). Here, the heterogeneous images are used to define low LIDM and for homogeneous images higher LIDM are evaluated using the equation.

$$LIDM = \sum_{m1,m2=0}^{G-1} \log\left(\frac{1}{1 + (m1 - jm2)}P(m1, m2)\right)$$

3.2.4. Max correlation inertia

Max correlation inertia (MCI) is used to find the maximal correlation between the grey level linear dependence among the pixels at the given positions. The maximum correlation and inertia measure describe the linear structure of an image. Also, it describes the distribution of grey scale values in an image.

$$\text{Correlation} = \sum_{m1,m2=0}^{G-1} \max\left\{\frac{\{m1 \times m2\} \times P(m1,m2) - \{\mu_x \times \mu_y\}}{\sigma_x \times \sigma_y}, \sum_{m1,m2=0}^{G-1} \{m - m2\}^2 \times P(m1,m2)\right\}$$

3.3. Bayesian based non-linear SVM classification

In the framework, different feature sets in the biometric images are classified using the hybrid non-linear SVM algorithm. The kernel values are modified at each point to remove the input image functionality. Here, a non-linear kernel function is optimized using the multi-modal facial features with multiple classes. Bayesian estimation in the SVM classifier improves the conditional estimation of each feature in the multi-modal feature space. Thedecision boundary of the proposed multi-class SVM is given as.

$$p = sgn\left(\sum_{i=1}^{n} y_i . \emptyset(x_i, y_i) + b\right)$$

4. EXPERIMENTAL RESULTS

Experimental results are simulated in real-time cloud computing environment. Results are developed using the python environment for feature extraction and classification process. In this work, different IoT captured video frames are taken as input to the proposed model in order to filter the essential key patterns and facial features. In these experimental results, different performance metrics such as number of key features, classification recall, precision, accuracy, F-measure, error rate, and runtime are computed and compared to the conventional models. The sample of in going facial expression images of the dataset for the training data preparation and landmark feature points detection process.

Real-time face detection using the noisy frame. Here, proposed feature extraction measures are used to find the key features in the real-time videos. As shown in the sample video frame, human face is detection with high probability using the proposed feature selection measures for the classification problem. Figure 2 illustrates the performance of proposed multiple features extraction count to the conventional facial feature extraction measures in the framework when the threshold is 0.3. Here, threshold is used to filter the essential key features among the large number of facial feature space. Figure 3 illustrates the performance of proposed multiple features extraction count to the conventional facial feature extraction measures in the framework when the threshold is 0.5. Here, threshold is used to filter the essential key features among the large number of facial feature space. Table 1 illustrates the performance of proposed multiple features extraction count to the conventional facial feature extraction measures in the framework when the threshold is 0.7. Here, threshold is used to filter the essential key features among the large number of facial feature space.
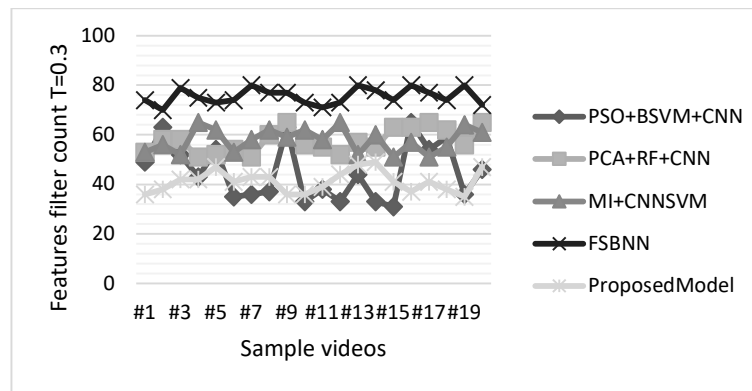

Figure 2. Performance analysis of proposed multiple feature extraction measures to the traditional measures for essential key features filtering when threshold T=0.3 in the framework

Table 2 describes the performance of proposed multiple facial feature extraction-based classifier to the conventional models for accuracy measure. In this table, the average accuracy value of all the test videos is taken as accuracy comparison between the proposed and existing models. Table 3 describes the performance of proposed multiple facial feature extraction-based classifier to the conventional models for recall measure. In this table, the average accuracy value of all the test videos is taken as recall comparison between the proposed and existing models.
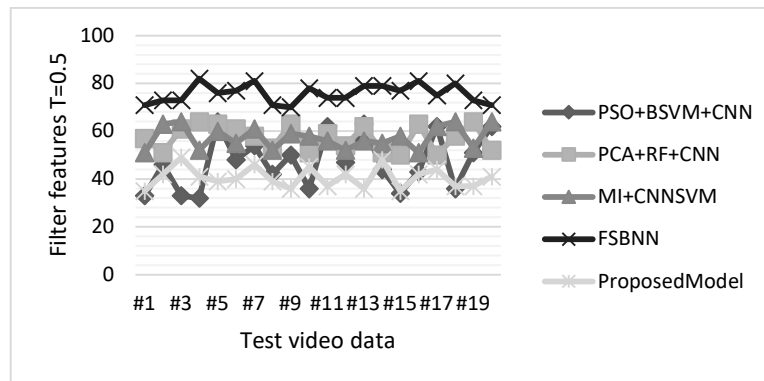

Figure 3. Performance analysis of proposed multiple feature extraction measures to the traditional measures for essential key features filtering when threshold T=0.3 in the framework

Table 1. Performance analysis of proposed multiple feature extraction measures to the traditional measures for essential key features filtering when threshold T=0.7 in the framework

Video data	PSO+BSVM+CNN	PCA+RF+CNN	MI+CNNSVM	FSBNN	Proposed model
#1	50	56	65	72	43
#2	55	54	51	77	46
#3	31	59	56	80	45
#4	58	61	51	75	38
#5	49	62	50	77	35
#6	31	64	55	73	43
#7	38	63	51	83	38
#8	32	52	58	76	38
#9	46	53	62	71	36
#10	39	55	59	73	40
#11	50	55	65	74	35
#12	46	60	54	74	45
#13	63	51	60	71	39
#14	57	62	57	75	41
#15	40	62	63	77	35
#16	44	54	52	71	44
#17	65	52	64	76	46
#18	60	53	59	76	36
#19	60	51	60	78	43
#20	35	63	62	72	37

Figure 4 describes the performance of proposed multiple facial feature extraction-based classifier to the conventional models for recall measure. In this table, the average accuracy value of all the test videos is taken as precision comparison between the proposed and existing models. Table 4 describes the performance of proposed multiple facial feature extraction-based classifier to the conventional models for F-measure. In this table, the average accuracy value of all the test videos is taken as F-measure comparison between the proposed and existing models.

Figure 5 describes the performance of proposed multiple facial feature extraction-based classifier to the conventional models for area under curve (AUC) measure. In this table, the average accuracy value of all the test videos is taken as AUC comparison between the proposed and existing models. Table 5, illustrates the performance of proposed multiple facial feature extraction-based classifier to the conventional models for runtime (ms) computation and Table 6, illustrates the error rate analysis of proposed classifier on different test facial features. From the table, it is noted that the proposed model has better minimization of error rate on the test data.

Table 2. Performance of proposed multiple facial feature extraction-based classifier to the conventional models for accuracy measure

Video data	PSO+BSVM+CNN	PCA+RF+CNN	MI+CNNSVM	FSBNN	Proposed model
#1	0.89	0.9	0.91	0.96	0.98
#2	0.89	0.89	0.92	0.96	0.98
#3	0.92	0.89	0.89	0.96	0.98
#4	0.91	0.92	0.9	0.95	0.98
#5	0.89	0.9	0.89	0.96	0.98
#6	0.91	0.91	0.93	0.97	0.99
#7	0.89	0.89	0.88	0.96	0.98
#8	0.88	0.93	0.91	0.97	0.98
#9	0.9	0.91	0.89	0.96	0.98
#10	0.9	0.89	0.92	0.95	0.98
#11	0.88	0.92	0.9	0.95	0.98
#12	0.91	0.91	0.93	0.96	0.98
#13	0.9	0.9	0.89	0.95	0.98
#14	0.92	0.91	0.92	0.97	0.97
#15	0.89	0.91	0.94	0.97	0.99
#16	0.91	0.89	0.93	0.95	0.97
#17	0.91	0.92	0.88	0.96	0.98
#18	0.9	0.89	0.94	0.96	0.99
#19	0.9	0.9	0.89	0.96	0.97
#20	0.91	0.9	0.9	0.97	0.98

Table 3. Performance of proposed multiple facial feature extraction-based classifier to the conventional models for recall measure

| Video data	PSO+BSVM+CNN	PCA+RF+CNN	MI+CNNSVM	FSBNN	Proposed model
#1	0.9	0.91	0.93	0.96	0.98
#2	0.9	0.89	0.89	0.96	0.97
#3	0.9	0.91	0.94	0.95	0.98
#4	0.9	0.91	0.92	0.96	0.98
#5	0.91	0.92	0.93	0.95	0.97
#6	0.92	0.91	0.92	0.96	0.97
#7	0.89	0.9	0.88	0.95	0.98
#8	0.89	0.91	0.93	0.97	0.98
#9	0.89	0.89	0.91	0.95	0.98
#10	0.9	0.9	0.91	0.97	0.98
#11	0.89	0.93	0.89	0.96	0.98
#12	0.91	0.89	0.92	0.95	0.98
#13	0.88	0.92	0.93	0.96	0.98
#14	0.88	0.88	0.9	0.97	0.98
#15	0.89	0.9	0.93	0.95	0.98
#16	0.9	0.91	0.92	0.96	0.98
#17	0.89	0.9	0.88	0.95	0.98
#18	0.9	0.9	0.93	0.97	0.98
#19	0.89	0.9	0.92	0.95	0.98
#20	0.9	0.91	0.88	0.95	0.98
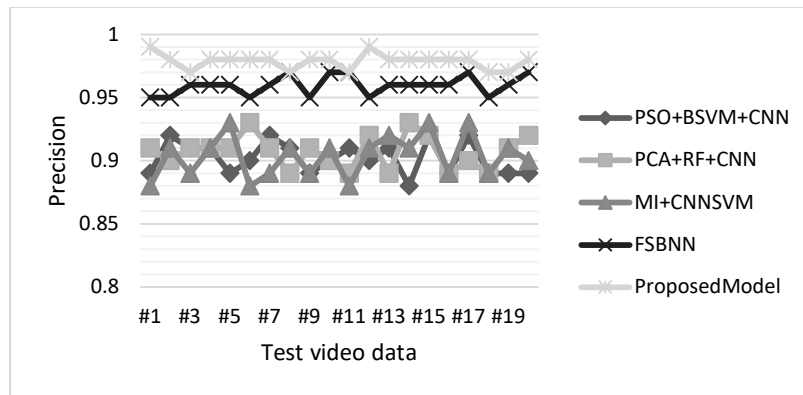
Figure 4. Performance of proposed multiple facial feature extraction-based classifier to the conventional models for precision measure

Table 4. Performance of proposed multiple facial feature extraction-based classifier to the conventional models for F-measure

| Video data	PSO+BSVM+CNN	PCA+RF+CNN	MI+CNNSVM	FSBNN	Proposed model
#1	0.89	0.91	0.9	0.95	0.97
#2	0.91	0.93	0.92	0.95	0.97
#3	0.9	0.91	0.89	0.96	0.98
#4	0.89	0.91	0.91	0.96	0.98
#5	0.9	0.9	0.88	0.96	0.98
#6	0.9	0.93	0.89	0.96	0.98
#7	0.88	0.91	0.9	0.97	0.98
#8	0.91	0.89	0.92	0.96	0.98
#9	0.92	0.92	0.93	0.97	0.98
#10	0.91	0.89	0.89	0.96	0.98
#11	0.88	0.88	0.9	0.96	0.98
#12	0.88	0.89	0.91	0.96	0.98
#13	0.89	0.9	0.91	0.96	0.98
#14	0.9	0.88	0.9	0.95	0.97
#15	0.88	0.89	0.92	0.96	0.97
#16	0.92	0.9	0.9	0.96	0.97
#17	0.91	0.89	0.93	0.96	0.98
#18	0.92	0.88	0.92	0.95	0.98
#19	0.9	0.91	0.91	0.96	0.99
#20	0.9	0.93	0.9	0.96	0.98
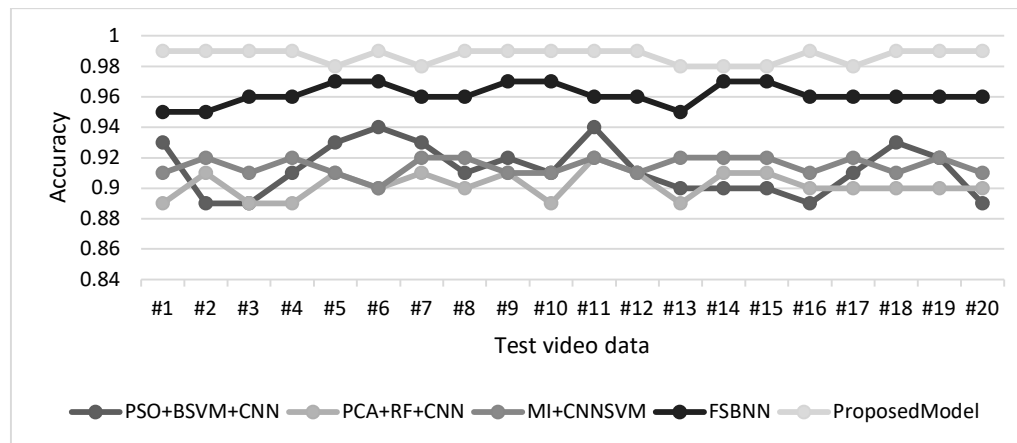
Figure 5. Performance of proposed multiple facial feature extraction-based classifier to the conventional models for AUC measure

Table 5. Performance of proposed multiple facial feature extraction-based classifier to the conventional models for runtime(ms) computation

Video data	PSO+BSVM+CNN	PCA+RF+CNN	MI+CNNSVM	FSBNN	Proposed model
#1	21371	15628	14262	13727	29420
#2	12593	10512	12180	10986	27313
#3	13549	15159	19308	17961	28118
#4	11597	10412	12980	14252	26480
#5	10297	10000	21110	12406	27668
#6	16104	10660	12389	12655	26352
#7	10856	12590	23287	17152	26121
#8	20845	21257	12072	20528	26576
#9	13673	21257	13509	19571	26616
#10	22784	20585	12350	15084	26396
#11	17215	16677	10627	19181	28166
#12	17059	15573	21037	17840	28673
#13	10502	20049	11688	16981	25314
#14	19436	11415	19972	23709	26520
#15	10005	12830	18455	22857	27066
#16	15327	18831	15419	15934	27792
#17	18514	22045	20755	21038	26331
#18	11959	22853	16387	23680	25600
#19	18775	19509	18247	20986	27172
#20	13098	18927	16431	11453	27691

Table 6. Performance of proposed multiple facial feature extraction-based classifier to the conventional models for error rate

Video data	PSO+BSVM+CNN	PCA+RF+CNN	MI+CNNSVM	FSBNN	Proposed model
#1	0.1	0.1	0.08	0.04	0.02
#2	0.06	0.08	0.08	0.03	0.01
#3	0.09	0.08	0.08	0.05	0.02
#4	0.12	0.11	0.09	0.03	0.01
#5	0.06	0.1	0.09	0.03	0.01
#6	0.1	0.08	0.09	0.04	0.02
#7	0.08	0.1	0.08	0.04	0.01
#8	0.07	0.1	0.09	0.03	0.01
#9	0.09	0.09	0.09	0.05	0.01
#10	0.08	0.1	0.1	0.04	0.02
#11	0.1	0.09	0.09	0.05	0.01
#12	0.08	0.09	0.09	0.03	0.02
#13	0.12	0.09	0.1	0.03	0.02
#14	0.1	0.1	0.09	0.04	0.01
#15	0.09	0.1	0.1	0.04	0.01
#16	0.1	0.09	0.1	0.04	0.02
#17	0.06	0.08	0.09	0.04	0.02
#18	0.08	0.09	0.09	0.03	0.02
#19	0.06	0.1	0.08	0.04	0.01
#20	0.07	0.09	0.08	0.03	0.01

4.1. Result analysis

In this work, a multi-level facial feature extraction-based ensemble classification framework is implemented on different facial expression datasets. As discussed in the experimental section, proposed model has better accuracy, precision, recall, F-measure, AUC and runtime than the traditional approaches such as PSO+BSVM+CNN, PCA+RF+CNN, MI+CNNSVM and FSBNN. Also, proposed model has better error rate (~10%) than the conventional models.

5.   CONCLUSION

In this paper, an efficient homogenous facial features extraction and classification framework is proposed to extract different features for the classification problem. Since, most of the traditional single modal facial features have limited features space for the classification problem. In this work, a hybrid classifier is used to classify the key facial points in the cloud computing environment. Experimental results show that the proposed hybrid multiple feature extraction-based framework has better computational efficiency in terms of accuracy, error rate, recall, precision and AUC than the conventional models. In future work, this modal is extended to implement a novel feature extraction and segmentation based multi-class classification framework on different multi-modal biometric features with a large features space and data size.

REFERENCES

[1]   M. Ahmady, S. S. Mirkamali, B. Pahlevanzadeh, E. Pashaei, A. A. R. Hosseinabadi, and A. Slowik, "Facial expression recognition using fuzzified Pseudo Zernike Moments and structural features," *Fuzzy Sets and Systems*, vol. 443, pp. 155–172, Aug. 2022, doi: 10.1016/j.fss.2022.03.013.

[2]   B. Allaert, J. Mennesson, I. M. Bilasco, and C. Djeraba, "Impact of the face registration techniques on facial expressions recognition," *Signal Processing: Image Communication*, vol. 61, pp. 44–53, Feb. 2018, doi: 10.1016/j.image.2017.11.002.

[3]   F. H. Almukhtar, "Facial emotions recognition using local monotonic pattern and gray level co-occurrence matrices plant leaf images aided agriculture development," *Optik*, vol. 271, p. 170161, Dec. 2022, doi: 10.1016/j.ijleo.2022.170161.

[4]   R. A. Virrey, C. De S. Liyanage, M. I. bin Pg Hj Petra, and P. E. Abas, "Visual data of facial expressions for automatic pain detection," *Journal of Visual Communication and Image Representation*, vol. 61, pp. 209–217, May 2019, doi: 10.1016/j.jvcir.2019.03.023.

[5]   S. Barra, S. Hossain, C. Pero, and S. Umer, "A Facial Expression Recognition Approach for Social IoT Frameworks," *Big Data Research*, vol. 30, p. 100353, Nov. 2022, doi: 10.1016/j.bdr.2022.100353.

[6]   R. Biswas, V. González-Castro, E. Fidalgo, and E. Alegre, "A new perceptual hashing method for verification and identity classification of occluded faces," *Image and Vision Computing*, vol. 113, p. 104245, 2021, doi: 10.1016/j.imavis.2021.104245.

[7]   G. B. Pujol *et al.*, "Anticipating Facial Nerve Position Using Three-Dimensional Tractography During the Preoperative Assessment of Cerebellopontine Angle Tumors," *World Neurosurgery*, vol. 168, pp. e317–e327, Dec. 2022, doi: 10.1016/j.wneu.2022.09.119.

[8]   C. Căleanu, D.-S. Huang, V. Gui, V. Tiponuţ, and V. Maranescu, "Interest Operator versus Gabor filtering for facial imagery classification," *Pattern Recognition Letters*, vol. 28, no. 8, pp. 950–956, Jun. 2007, doi: 10.1016/j.patrec.2006.12.013.

[9]   C.-D. Căleanu, X. Mao, G. Pradel, S. Moga, and Y. Xue, "Combined pattern search optimization of feature extraction and classification parameters in facial recognition," *Pattern Recognition Letters*, vol. 32, no. 9, pp. 1250–1255, 2011, doi: 10.1016/j.patrec.2011.03.019.

[10] D. Cassi, G. Battistoni, M. Magnifico, C. Di Blasio, G. Pedrazzi, and A. Di Blasio, "Three-dimensional evaluation of facial asymmetry in patients with hemifacial microsomia using stereophotogrammetry," *Journal of Cranio-Maxillofacial Surgery*, vol. 47, no. 1, pp. 179–184, Jan. 2019, doi: 10.1016/j.jcms.2018.11.011.

[11] W.-C. Chiang, H.-H. Lin, C.-S. Huang, L.-J. Lo, and S.-Y. Wan, "The cluster assessment of facial attractiveness using fuzzy neural network classifier based on 3D Moiré features," *Pattern Recognition*, vol. 47, no. 3, pp. 1249–1260, 2014, doi: 10.1016/j.patcog.2013.09.007.

[12] S. Coulibaly, B. Kamsu-Foguem, D. Kamissoko, and D. Traore, "Deep Convolution Neural Network sharing for the multi-label images classification," *Machine Learning with Applications*, vol. 10, p. 100422, Dec. 2022, doi: 10.1016/j.mlwa.2022.100422.

[13] N. Dagnes *et al.*, "Optimal marker set assessment for motion capture of 3D mimic facial movements," *Journal of Biomechanics*, vol. 93, pp. 86–93, Aug. 2019, doi: 10.1016/j.jbiomech.2019.06.012.

[14] C. R. De Silva, S. Ranganath, and L. C. De Silva, "Cloud basis function neural network: A modified RBF network architecture for holistic facial expression recognition," *Pattern Recognition*, vol. 41, no. 4, pp. 1241–1253, Apr. 2008, doi: 10.1016/j.patcog.2007.09.015.

[15] M. M. Dehshibi and A. Bastanfard, "A new algorithm for age recognition from facial images," *Signal Processing*, vol. 90, no. 8, pp. 2431–2444, Aug. 2010, doi: 10.1016/j.sigpro.2010.02.015.

[16] A. Elmahmudi and H. Ugail, "Deep face recognition using imperfect facial data," *Future Generation Computer Systems*, vol. 99, pp. 213–225, Oct. 2019, doi: 10.1016/j.future.2019.04.025.

[17] J. L. Gómez-Sirvent, F. López de la Rosa, R. Sánchez-Reolid, R. Morales, and A. Fernández-Caballero, "Defect classification on semiconductor wafers using Fisher vector and visual vocabularies coding," *Measurement*, vol. 202, p. 111872, Oct. 2022, doi: 10.1016/j.measurement.2022.111872.

[18] X. Guo, Y. Tie, L. Ye, and J. Yan, "Identifying facial expression using adaptive sub-layer compensation based feature extraction," *Journal of Visual Communication and Image Representation*, vol. 50, pp. 65–73, Jan. 2018, doi: 10.1016/j.jvcir.2017.11.007.

[19] H. He and S. Chen, "Identification of facial expression using a multiple impression feedback recognition model," *Applied Soft Computing*, vol. 113, p. 107930, Dec. 2021, doi: 10.1016/j.asoc.2021.107930.

[20] B. Hernández, G. Olague, R. Hammoud, L. Trujillo, and E. Romero, "Visual learning of texture descriptors for facial expression recognition in thermal imagery," *Computer Vision and Image Understanding*, vol. 106, no. 2–3, pp. 258–269, 2007, doi: 10.1016/j.cviu.2006.08.012.

[21] J. Huang *et al.*, "Automatic recognition of schizophrenia from facial videos using 3D convolutional neural network," *Asian Journal of Psychiatry*, vol. 77, p. 103263, Nov. 2022, doi: 10.1016/j.ajp.2022.103263.

[22] S. Koley, H. Roy, S. Dhar, and D. Bhattacharjee, "Illumination invariant face recognition using Fused Cross Lattice Pattern of Phase Congruency (FCLPPC)," *Information Sciences*, vol. 584, pp. 633–648, Jan. 2022, doi: 10.1016/j.ins.2021.10.059.

[23] P. Kumar, P. P. Roy, and D. P. Dogra, "Independent Bayesian classifier combination based sign language recognition using facial expression," *Information Sciences*, vol. 428, pp. 30–48, Feb. 2018, doi: 10.1016/j.ins.2017.10.046.

[24] H. Li, F. Chung, and S. Wang, "A SVM based classification method for homogeneous data," *Applied Soft Computing*, vol. 36, pp. 228–235, Nov. 2015, doi: 10.1016/j.asoc.2015.07.027.

[25] Z. Li and J. Xu, "Feature extraction via prototype margin distance maximizing criterion for subspace learning," *Optik*, vol. 231, p. 166491, Apr. 2021, doi: 10.1016/j.ijleo.2021.166491.

[26] D. Liang, J. Yang, Z. Zheng, and Y. Chang, "A facial expression recognition system based on supervised locally linear embedding," *Pattern Recognition Letters*, vol. 26, no. 15, pp. 2374–2389, Nov. 2005, doi: 10.1016/j.patrec.2005.04.011.

[27] P. B. Padol and A. A. Yadav, "SVM classifier based grape leaf disease detection," in *2016 Conference on Advances in Signal Processing (CASP)*, Jun. 2016, pp. 175–179, doi: 10.1109/CASP.2016.7746160.

[28] Z. N. Reza, F. Nuzhat, N. A. Mahsa, and M. H. Ali, "Detecting jute plant disease using image processing and machine learning," in *2016 3rd International Conference on Electrical Engineering and Information Communication Technology (ICEEICT)*, Sep. 2016, pp. 1–6, doi: 10.1109/CEEICT.2016.7873147.

[29] M. Geetha, R. S. Latha, S. K. Nivetha, S. Hariprasath, S. Gowtham, and C. S. Deepak, "Design of face detection and recognition system to monitor students during online examinations using Machine Learning algorithms," in *2021 International Conference on Computer Communication and Informatics (ICCCI)*, Jan. 2021, pp. 1–4, doi: 10.1109/ICCCI50826.2021.9402553.

[30] A. Costache, D. Popescu, and L. Ichim, "Facial Expression Detection by Combining Deep Learning Neural Networks," in *2021 12th International Symposium on Advanced Topics in Electrical Engineering (ATEE)*, 2021, pp. 1–5, doi: 10.1109/ATEE52255.2021.9425340.

BIOGRAPHIES OF AUTHORS

SaiTeja Chopparapu [ID] [GS] [SC] [◈] is a Ph.D. student in the Department of Electronics and Communication Engineering, GITAM (Deemed to be University). He received his B.Tech. degree from Dhanekula Engineering College, Ganguru and M.Tech. degree from VIT (Deemed to be University). His area of interests includes wireless sensor networks, image processing, and linear integrated circuits. He can be contacted at email: saiteja.chopparapu960@gmail.com.

Beatrice Seventline Joseph [ID] [GS] [SC] [◈] is currently working as a HoD in the Electrical Electronics and Communication Engineering, GITAM (Deemed to be University). She obtained her Ph.D. degree from Andhra University in 2012. She has more than 25 years of teaching experience. More than 30 research papers have been published in various National/International journal/conferences. She's a lifetime member of IETE. Her area of interest includes radar signal processing digital, image processing and VLSI signal processing. She can be contacted at email: sjoseph@gitam.edu.