TPM Vol. 32, No. S4, 2025
ISSN: 1972-6325
https://www.tpmap.org/

AN EFFICIENT HYBRID LEARNING APPROACH FOR EMOTION RECOGNITION USING FACIAL EXPRESSION

C.SHAKILA[1,*], T.KAMALA KANNAN[2]
[1,*]RESEARCH SCHOLAR, DEPARTMENT OF COMPUTING SCIENCE AND TECHNOLOGY, VELS INSTITUTE OF SCIENCE TECHNOLOGY AND ADVANCED STUDIES (VISTAS) – PALLAVARAM CHENNAI,600117, TN, INDIA.
ORCID: 0009-0008-4883-7439
[2]PROFESSOR, DEPARTMENT OF INFORMATION TECHNOLOGY, SCHOOL OF COMPUTING SCIENCE, VELS INSTITUTE OF SCIENCE TECHNOLOGY AND ADVANCED STUDIES (VISTAS) – PALLAVARAM CHENNAI,600117, TN, INDIA .kamalakannan98@gmail.com
ORCID : 0000-0002-5982-8983
***[1]Corresponding Author: C.Shakila Email: shakilac752@gmail.com**

Abstract- Face-based emotion identification is an important area of study in man-machine interaction research. Face accessories, uneven light, shifting settings, and other factors are some of the difficulties in the field of emotion recognition. The drawback of traditional emotion detection techniques is that feature extraction and categorization are mutually optimized. Researchers are paying more attention to deep learning (DL) techniques in an attempt to solve this problem. In classification tasks, DL approaches are becoming more and more crucial. This study addresses emotion recognition through transfer learning approaches. Nasnet Mobile Network Features with GRU-CNN (NMGC) classifier is used in this work. Finally, updating the weights is the only method available to train the newly added layers. An accuracy of 98.63% was achieved in the experiment when assigning emotions based on the CK database.

Keywords- emotion recognition, face identification, deep learning, feature representation, accuracy

1. INTRODUCTION

In communication, emotions play a critical role. A variety of uses benefit from facial expression recognition, such as smart card applications, social robots, e-learning, criminal justice systems, security monitoring, and customer satisfaction identification [1]. Emotion recognition, Face detection, and feature extraction are the three main parts of the typical emotion identification system. Local binary patterns, Principal component analysis, Bezier curves, feature level fusion techniques, Principal component analysis [2], two-directional two-dimensional Modified Fisher principal component analysis, clustering methods, two-directional two-dimensional Fisher principal component analysis, Independent Component analysis, etc. are the most widely used feature extraction techniques, according to the literature. The features are then supplied into classifiers to be classified, like Decision trees, Naïve Bayes, k-nearest neighbors, Hidden Markov models, Support vector machines (SVM), etc. Conventional systems suffer from independent feature extraction and classification processes [3]. Therefore, improving the system's performance is difficult. DL networks address conventional methodologies through an end-to-end learning process. For DL, the larger the dataset, the more critical factor is dataset size. To enhance DL performance, researchers are employing techniques such as translations, data augmentation, cropping, normalizations, scaling and noise addition methodologies to supplement the volume of data [4]. Regarding segmentation and classification tasks, CNNs are the most effective algorithms. One of the key benefits of this CNN is the automatic feature extraction. One DL technique is transfer learning, which involves using knowledge transferred from one task to another to retrain a model for that task. Time-saving and accuracy-boosting are two of transfer learning's primary benefits [5].

We present some of the most present findings in the convolutional neural network (CNN) expression recognition field. The multi-region ensemble CNN technique for facial expression identification was first presented by [6]. The features derived from the three regions of the mouth, nose, and eyes are obtained by three sub-networks. Subsequently, three subnetworks' weights are combined to predict those emotions. This work makes use of the

databases RAF-DB and AFEW 7.0. The author in [7] suggested using the auxiliary model to recognize emotions. The information from the three main sub-regions of the mouth, nose, and eyes is integrated with the overall face image using the weighting technique in this work to capture the maximum amount of information. The four databases listed below evaluate the model: CK, SFEW, JAFFE, and FER2013 [8]. The author demonstrated face emotion recognition using VGG16 and Resnet50. This work makes use of the databases FER2013 and JAFFE. The results of the experiment demonstrate that, when compared to other advanced methods, Resnet50 achieves the best classification accuracy [9].

Deep CNN-based features were proposed by [10] as a method for emotion recognition. In this work, a multi-class SVM is used for classification, while VGG16 is used for feature extraction. With the CK database, the suggested algorithm produced an accuracy of 86.04% with the face detection algorithm and 81.36% without. The author in [11] Inception V3 model was used for emotion recognition. The work received a test accuracy of 39% after being assessed on the KDEF database. A system for identifying emotions that managed position variations and occlusions using the ALEC V2 architecture was described by [12]. The model achieves 92.5% accuracy when tested on real-time hidden images. Based on facial expression detection, the author proposed pre-trained CNN features in [13]. In this work, the features are extracted using a pre-trained VGG19 network, and the expressions will be predicted using SVM. The experiment produced 92.86% and 92.26% accuracy levels, respectively, using the databases JAFFE and CK.

An SVM classifier-based transfer learning strategy was presented by [14]. To extract the features for this study, CNNs and AlexNet are used. The features are then fed into SVM for classification. The work was completed with good precision utilizing the CK and NVIE databases. CNNs for facial emotion recognition were demonstrated. Several models, including VGG 16, ResNet50, and VGG 19, were used in the experiment with the fer2013 dataset. With an accuracy of 63.07%, VGG 16 outperformed the other three models. The author demonstrated a LeNet-based system for understanding emotions. A composite dataset (JAFFE, KDEF, and own proprietary data) is used in this work. In this study, undesirable pixels not needed for expression detection are removed using the Haar cascade library. An accuracy of 96.43% was attained in this attempt. For facial expression identification, the VGG16 and Resnet50 architectures were first presented by [15]. This paper proposes a hybrid Nasnet Mobile Network Features with GRU-CNN (NMGC) classifier models. In this work, the NMGC model outperformed the baseline CNN, the individual GRU and MobileNet models, and their respective levels of accuracy. The researchers looked into the face expression identification process using learning. The combined datasets of CK and JAFFE and the pre-trained networks of NMGC architectures obtained an average accuracy of 98.6% in this work.

The work is provided as: section 2 gives the comprehensive analysis of diverse approaches investigated by the researchers. The integrated learning model is explained in section 3 with experimental outcomes in section 4. The conclusion is provided in section 5.

2. RELATED WORKS

Effective computing is one of the areas of research that is most active at the moment. Effective computing is developing technologies to understand better and replicate human effects [16]. Efficient computing aims to make computers smarter so they can communicate with humans. Effective computing is used in various disciplines, including virtual sales assistants, Internet banking, neurology, medicine, and security [17]. Emotion recognition through verbal cues, body language, and facial expressions is the initial step in affective computing. The model illustrates the division of emotions theories into three categories: neurological (Facial feedback theory), cognitive (Lazarus theory), and physiological (James–Lange and Cannon-Bard theories). According to the James-Lange paradigm, the interpretation of the physiological reaction is what causes emotion to occur. Walter Cannon then challenged the James-Lange theory and proposed the Cannon-Bard hypothesis, which maintains that bodily responses and emotions occur simultaneously. The physiological response happens first in the Lazarus hypothesis, also known as the cognitive appraisal theory, and the individual then considers why they are experiencing the feeling [18]. Lastly, the theory of facial feedback describes how facial expressions convey emotion.

The author in [19] use a novel DL-based spoof face identification algorithm. Two DL models are adapted to carry out this process, known as receptive fields CNN and LRF-ELM models. It is provided with an associated layer and a pooling layer arranged before the associated layer. CNN model is included with the continuation layers of associated and pooling layers. The CNN architecture is also provided with fully associated layers. The experiments show that this approach is run on two mainstream parody face recognition databases, CASIA and NUAA. The efficiency of the approach LRF-ELM, which is run on these databases, is highly improved. The author in [20] established an IA-gen process to reduce the image variations by recreating the expressions from the

inputted face images. The conditional generative models are initially utilized to create six prototypic facial expressions from inputted face images while maintaining the identity information as unchangeable. Generative Adversarial Networks are introduced to make the conditional generative models create prototypic facial expressions from the inputted images. Then, regular CNN (FER-Net) is utilized to classify the expressions. This approach is implemented on Oulu-CASIA, BU-4DFE, CK+ and BU-3DFE databases, and the performance of this method is enhanced effectively. The author in [21] used a new method that recognizes the expressions of the face image with low computational complexity. This approach builds a model of facial attributes based on a weighted selection scheme. It uses Hidden Markov Models to divide an input video into one of these six expressions: anger, happiness, fear, surprise, and sadness. The permanent segments like apex, onset, and neutral expression elements are acquired using a variable-point detecting unit. The calculations on subject-independent analysis are performed using Cohn- Kanade and Beihang University facial expression datasets. The intensity of the estimation of the expressions and the performance of the face identification process are greatly enhanced [22].

The author in [22] control multi-dimensional data using LDA and three-fold SVM to minimize false labeling and complexity. Specific FER is utilized, and six basic expressions are considered multi-class data. The images are split into seven triangles using two focal points. An integrated global and local feature descriptor is formed. The discriminant attributes are acquired by applying and processing discrete Fourier transforms with LDA and correctly mapping the input feature space to the specific output space. The Japanese Female Facial Expression, Cohn-Kanade DFAT, and FER 2013 datasets compute the system's performance [23]. The experimental outputs show that the multi-dimensional data using SVM and LDA approach is efficient and straightforward for data evaluation (quadratic). In [24] formed an expression recognition process based on cognition and mapped binary patterns. An LBP operator is primarily utilized in this approach to acquire facial contours. Later, a pseudo-3D model divides the face regions into the six basic expression sub-areas. The extraction of features is performed by mapping the sub-areas and global facial expression images to LBP operators. The two classification models utilized in this method, SVM and softmax, and two emotion models are utilized, known as the circumflex emotion model and basic emotions model. The astounding factors in the image can be erased efficiently. Because of the circumflex emotion model, the efficiency of this method is enhanced than the existing emotion techniques [25].

The author in [26] extract the expressions from a single-face image by integrating geometric and appearance attributes with SVM classification. Generally, the appearance attributes are calculated by splitting the face area into a regular framework such as a holistic format. However, in this approach, the appearance attributes are evaluated by splitting the entire face area into domain-specific local areas. The geometric attributes can also be removed from the domain-specific areas [27]. The major local areas were also discovered using an incremental searching method that reduces the dimensions of features and increases identification accuracy [28]. The facial expression recognition outcomes that utilize the attributes from domain-specific areas are compared with the outcome acquired by holistic representation. FER performance is computed on publicly available advanced CK+ facial expression datasets [28] – [30].

3. METHODOLOGY

This section gives the detailed analysis of proposed NMGC model for emotion recognition with facial expression. The input images from online available dataset is provided for pre-processing, feature representation and classification. The experimental outcomes are provided in the consecutive section. Fig 1 demonstrates the work flow of the anticipated NMGC model.

TPM Vol. 32, No. S4, 2025
ISSN: 1972-6325
https://www.tpmap.org/

Open Access

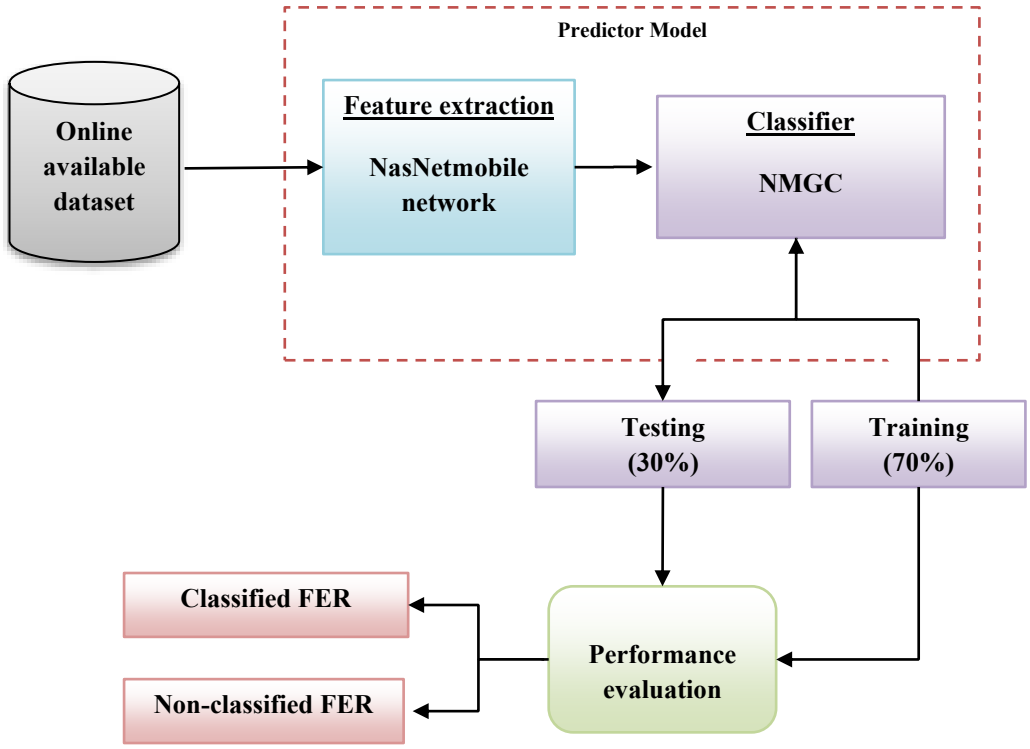


Fig 1 Overall block diagram

3.1. Pre-processing

The dataset was processed using a variety of experimental image processing techniques. CLAHE, on the other hand, significantly increased prediction accuracy. To prevent excessive contrast enhancement, which could produce processed images with strange appearances and unwanted artifacts, the CLAHE cut histogram at a predetermined clipping value. In addition, mishandling the application of contrast enhancement methods may result in an unsatisfactory appearance in disappearing areas, usually the small veins. By adding a global threshold value to the filter's fixed contrast point, the improved CLAHE could solve the issue. Subsequently, the pre-defined global threshold value would adaptively improve the targeting image's histogram. This section demonstrate the better efficiency of the enhanced CLAHE, which surpasses the fixed clipping feature provided by the traditional CLAHE. Notably, the suggested model enhanced the image's distinctiveness, particularly the little veins. To widen the width of the intensity distribution within a suitable range and enable the entry of more valuable data into the CNN model, we then performed normalization. The images could cause undesired distortion. The region (light ring) could be swapped for the original black background to fix the issue. It was possible to run the correction procedure in the "LAB" color space before switching it back to the RGB color system.

$$CLIP\ Limit\ = \left[\frac{\varphi}{L}\right] + \left[\beta\ .\ \left(\varphi - \left[\frac{\varphi}{L}\right]\right)\right] \tag{1}$$

$$CLIP\ Limit\ = \frac{T}{80} \tag{2}$$

Here, T represents global threshold, φ represents pixel block population, β represents clip factor and L represents gray scale.

3.2. NasNet

NASNetmobile design uses reinforcement learning techniques to investigate the best CNN architectures. The Google Brain team has achieved substantial progress in Neural Architecture Search (NAS). Although there are

differences in the sizes of NAS designs, it should be noted that NasNetMobile is a reduced version. There are over 4.5 million parameters in NASNetMobile. 224 × 224 pixels is the size of the approved input image.

3.3. GRU

Gated Recurrent Unit (GRU) networks are improved versions of recurrent neural networks (RNNs). RNN cannot provide a long-term nonlinear relationship when the input sequence is extended because long-term dependencies emerge. This suggests that the learning sequence has instances of gradient explosion and vanishing. Numerous optimization theories and enhanced methods have been put forth to address this issue, including long short-term memory networks, independent RNNs, bidirectional long short-term memory, GRU networks, and echo state networks (Cao, 2020). The two main goals of the GRU network are the long-term reliance and gradient disappearance issues with RNNs. With fewer parameters than an LSTM, a GRU network functions similarly to long short-term memory networks that use forget gates. Fig 2 illustrates how the GRU network optimizes the learning mechanism by utilizing reset and update gates. The model is assisted in deciding how much past data (from earlier time steps) should be carried over into the future by the update gate and how much previous data should be ignored by the reset gate. The GRU network's update gate model is computed using the following formula.

$$z(t) = \sigma(W(z).[h(t-1), x(t)]) \tag{3}$$

The update gate function is represented by $z(t)$, the sigmoid function is represented by σ, $W(z)$ represents the update gate's weight, $h(t-1)$ indicates the output of the preceding neuron and $x(t)$ indicates the input of the current neuron. The GRU neural network's reset gate model is computed using the formula below:

$$r(t) = \sigma(W(r).[h(t-1), x(t)]) \tag{4}$$

Here, $x(t)$ is the input of the current neuron; σ is the sigmoid function; $r(t)$ is the reset gate function; and $W(r)$ is the reset gate weight. The Equation below displays the value of the GRU hidden layer's output.

$$h(t) = tanh(Wh.[rt * h(t-1), x(t)]) \tag{5}$$

The variables $h(t), h(t-1), x(t), W_h$ (the update gate weight) and tanh (the hyperbolic tangent function) represent the input and output values of the current neuron, as well as the previous and subsequent neurons, respectively. The hyperbolic tangent function $tanh()$, the input of the current neuron, W_h is the weight of the update gate, the output of the previous neuron, $x(t)$, the amount that $h(t)$, and the output value to be decided in this neuron, $h(t-1)$ are all controlled by R_t. It regulates the amount of memory that must be kept up to date. The following Eq. (6) displays the data for the final result from the hidden layer.

$$h(t) = (1 - z(t).h(t-1) + z(t) * h(t)) \tag{6}$$

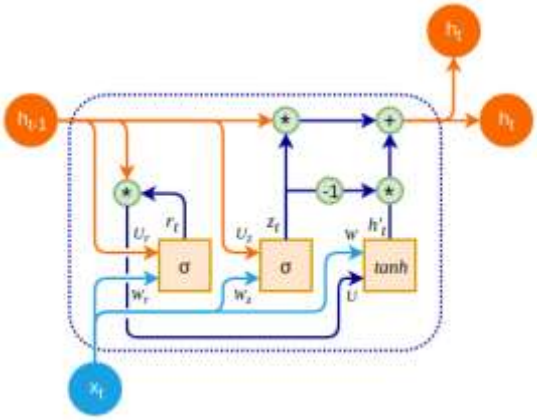


Fig 2 Generic GRU

TPM Vol. 32, No. S4, 2025
ISSN: 1972-6325
https://www.tpmap.org/

Open Access

3.4. CNN

Here, Feed-forward neural networks with a known, grid-like structure processing data are called convolutional neural networks, or CNNs. Both supervised and unsupervised learning can be used. Despite its original primary aim being field image processing, CNN has shown outstanding success in various real-world applications, such as natural language processing and speech recognition. The typical CNN architecture used for image categorization is the foundation for the CNN model. As seen in Fig 3, it comprises a classification section and a feature extraction. These components include maximum merge layers, batch normalization, and convolution. The architecture's hidden layer is made up of these layers. Convolution operations are carried out by the convolutional layer using the given filter and kernel parameters. As the maximum pooling layer reduces the feature space's dimension, it computes the network weights for the subsequent layer. For each training mini-batch, batch normalization is utilized to lessen the impact of various input distributions, improving training. Training with CNN models is fast and accurate due to their activation functions. Activation functions used by CNN include the Rectified Linear Unit (ReLU), hyperbolic tangent (Tanh), and Sigmoid. As indicated by the equations below, we employed two activation functions in this model: the input and hidden layers' ReLU function and the output layer's Sigmoid function.

$$h_i^m = ReLU\ (W_i^{m-1} * V_i^{m-1} + b^{m-1}) \tag{7}$$

Where, the convolutional layer is represented by hi m, the nodes are represented by V_i^{m-1}, the neurons' weights are represented by W_i^{m-1} and the bias layer is represented by b^{m-1}).

$$S(x) = \frac{1}{1 + e^{-\sum_k W_i + x_i + b}} \tag{8}$$

Where, b is the bias, e is Euler's number = 2.781, W_i is the input weight and X_i is the input.

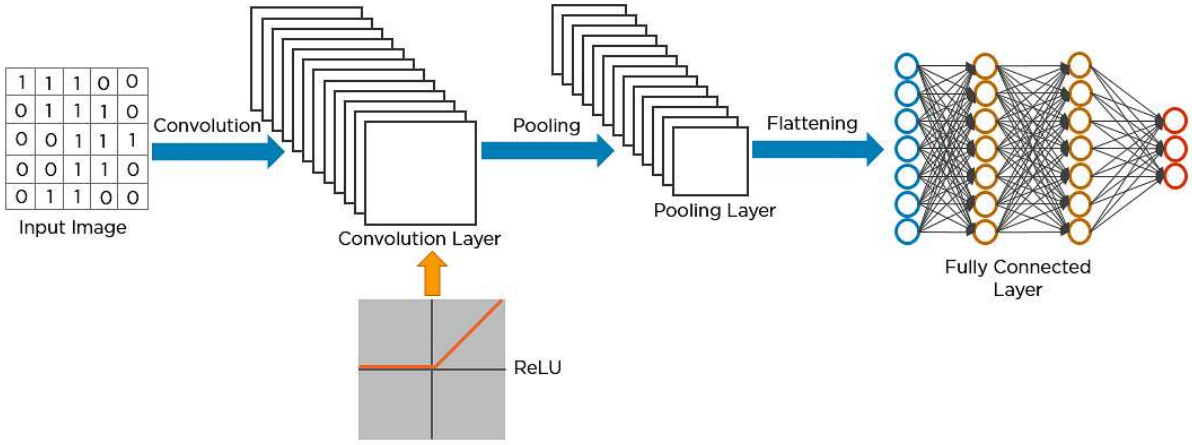


Fig 3 Generic CNN

4. NUMERICAL RESULTS AND DISCUSSION

The proposed outcomes of the NMGC model are covered along with the corresponding explanations and numerical results. The machine is set up for the simulation and runs in the MATLAB 2020a environment, with 8 GB of RAM, a 64-bit OS, and an Intel Core I5 CPU. Here, two experiments are modeled to validate anticipated model performance. Initial work examines the performance of anticipated algorithms and validates NMGC training time, which is lower than the conventional CNN model. Facial emotion recognition data were acquired from the expression dataset, as in Table 1. This dataset comprises training images and testing images. It has 48 ∗ 48 grayscale images. Faces are located in the middle of every image. Henceforth, experimentation data are directly provided as inputs to the network for training purposes without any preprocessing. Here, training and testing are done with an 80:20 ratio is done here. Generally, all entries are from the original dataset. This is utilized for both validation and training.

Table 1 Dataset samples

Dataset	Facial expression dataset
Neutral	50
Disgust	53
Anger	51
Fear	58
Sad	59
Happy	66
Surprise	63

Then, to validate NMGC effectiveness, three factors are measured. They are feature extraction without or with weighted sharing, and the fusion-based learning model and NMGC optimality prediction are validated. Here, NMGC is shared among diverse kinds of attribute mapping in FE subsets. However, diverse attribute mapping is provided separately to diverse CNN for fusion analysis. Then, it is essential to integrate prediction expression and feature fusion. It is essential to perform learning and parameter computation. Some factors like accuracy, computation time and quantity are measured. The anticipated NMGC model executes slowly with 4.1 Hz. Moreover, NMGC acquires slightly superior results than conventional models. This is owing to an extremely restricted number of samples used to train NMGC. Therefore, if there are some appropriate training samples, NMGC still holds a higher potential to outperform traditional models. However, it has to learn huge parameters and consumes huge training time. Table 2 depicts the accuracy computation of the proposed NMGC with the CNN model. Here, the period is set as 7.4 Hz, and the accuracy of the NMGC is 98.63% which is 12.63% higher than the CNN model.

Table 2 Accuracy evaluation

Methods	Parameter	Period	Accuracy
CNN	≅ 40 MB	7.4 Hz	86
NMGC	≅ 30 MB	7.4 Hz	**98.63**

The NMGC model is compared with GRU, CNN and ALEC learning to depict the significance of learning-based fusion. Here, NMGC features are extracted from a subset of GRU-CNN feature extraction. From Table 3, network fusion acquires superior outcomes than SVM, whereas NMGC integrated fusion and learning process in end-to-end training acquires superior results than GRU and CNN. To validate the consequences of the fusion model, every initialized GRU-CNN feature extraction parameter is learned from fusion subset parameters. This is equal to learning hierarchical weighting merged with high-dimensional pre-trained features. This NMGC acquires better results than CNN, GRU and ALEC. This model depicts that merging training frameworks is essential for anticipated NMGC.

Table 3 Classification results

Models	Classification results
GRU	84.17
ALEC	85.73

CNN	86.86
NMGC	**98.63**

From Table 3, four different models are considered for attaining classification accuracy. They are CNN, GRU, ALEC and NMGC. The predicted NMGC has a better classification result with 98.63% based on the suggested model which is greater than other methods. To confirm that the predicted NMGC model is optimal for predicting expression, some essential classifiers are considered. Experiments were performed in an expression dataset with dimensionally fused deep features offered by NMGC. The classifier parameters are chosen carefully by a training set of all training phases as in Table 3. In contrast, CNN is set to have default values from 1-1000 times for training. From the table provided below, it is known that every classifier can attain more effectual results than a regression model. Thus, this proves that dimensionality fusion is extremely discriminative. Regarding speed and accuracy, each classifier has advantages. Now, with deep features provided by NMGC, GRU is utilized for the best classifier-based expression prediction. The GRU, CNN and ALEC models are compared with the proposed NMGC model. The suggested NMGC model has a prediction accuracy of 98.63%, which is 1.03%, 2.16%, and 3.72% greater than previous methods. In the same way, the suggested model is contrasted with several machine learning techniques, including the regression model, k-NN, NB, RF, and k-SVM (Kernel SVM). As seen in Table 4, the regression, k-NN, NB, RF, and k-SVM models are 6.86%, 2.05%, 1.97%, 2.71%, and 1.13% lower in prediction accuracy than the suggested NMGC model which has an accuracy of 95%.

Table 4 Accuracy computation of NMGC with existing models

Classifier	Accuracy
Regression	81.03
K-NN	85.84
NB	85.92
RF	85.18
Kernel SVM	86.76
NMGC	**98.63**

At last, it is considered that NMGC offers issues encountered in the optimal prediction of fused deep features. The batch size of the proposed NMGC is 32, epochs is 100 and activation is STEM_BN1. Experimental outcomes request features that can acquire effectual results than convolutional deep features. Here, the generalized competency of the NMGC model under cross-dataset validation is provided. In this experiment, NMGC was trained and evaluated. From the evaluated experiments, NMGC provides an average cross-dataset recognition accuracy. This work shows the highest outcomes that are trained and evaluated. NMGC acquires better performance than other techniques with some expectations. For trained modeling, NMGC shows the least facial expression dataset correspondingly. It recommends NMGC training over an expression dataset owing to the huge training data. This expression dataset is only sometimes consistent with other models. This is because facial images vary in huge posing factors. This causes certain misalignments in landmark points. Here, all training images are combined with facial expression datasets for training and computed performance of NMGC with other CNN models over manually collected expression data. Some poor performances are attributed to two factors: It is extremely complex for certain models to recover missing facial parts with real-time models. Next is a weaker ability to concentrate on distinctive facial regions. NMGC acquires the lowest and highest classification accuracy, corresponding with disgust and happy categories. Some confused categories are sad, surprise, fear and disgust. Some failed samples are also encountered with the facial expression dataset. Even though NMGC is more robust with classification, it experiences higher facial occlusions. It is validated as the inevitable cause of huge misalignment in facial expressions. Therefore, NMGC needs to be more competent to concentrate on previously planned facial patches. It constructs weight, which is not adaptive when occlusion occurs. Only probable approaches to deal with these complex examples will be explored.

4.2. Performance metrics

Performance measures for the proposed NMGC model include precision, accuracy, specificity, sensitivity, recall, F-measure, ROC, confusion matrix, and MCC. These are assessed using True Positive (TP), False Positive (FP), True Negative (TN), and False Negative (FN).

a) Accuracy

An overall indicator of classification effectiveness is accuracy. It is stated as follows in Eq. (9):

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN} \tag{9}$$

b) Sensitivity

It is depicted as a classifier measure to identify positive class patterns. As shown in Eq. (10), it is represented as:

$$Sensitivity = \frac{TP}{TP + FN} \tag{10}$$

c) Specificity

It is depicted to measure classifier competency to identify negative class patterns as in Eq. (11):

$$Specificity = \frac{TN}{TN + FP} \tag{11}$$

d) ROC curves

The prediction accuracy is measured using the ROC curve. Generally, this curve is plotted using the sensitivity and specificity measure with various threshold values. It is also determined as TPR (recall) and FPR (fall out) rates. The curve is plotted among the sensitivity and fall out of the classifier model. Fallout is considered the value of the specificity rate.

e) Mathew's correlation coefficient (MCC)

MCC is the classification rate in binary class problems ranging from -1 to +1. Here, -1 specifies the mistake or error, and +1 specifies the appropriate label. However, '0' relies on random prediction, and it is expressed as in Eq. (12):

$$MCC = \frac{(TP * TN) - (FP * FN)}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \tag{12}$$

The model's functionality is used to assess NMGC. Phases of training, testing, and validation are applied to the samples. During simulating, only 10% of the samples are used for testing. The slices are rotated and cropped in $90^0, 180^0, and\ 270^0$) angle degrees to improve the testing and training process. The nodules are vertically and horizontally flipped. Various control metrics are modeled to confirm that the NMGC performance is significant. Some conditions need to be maintained during the experimentation process for certain factors and diverse effects to be evaluated. MCC completely relies on specificity and sensitivity values. Therefore, it is directly processed while performing the computation with the MATLAB 2020a simulator.

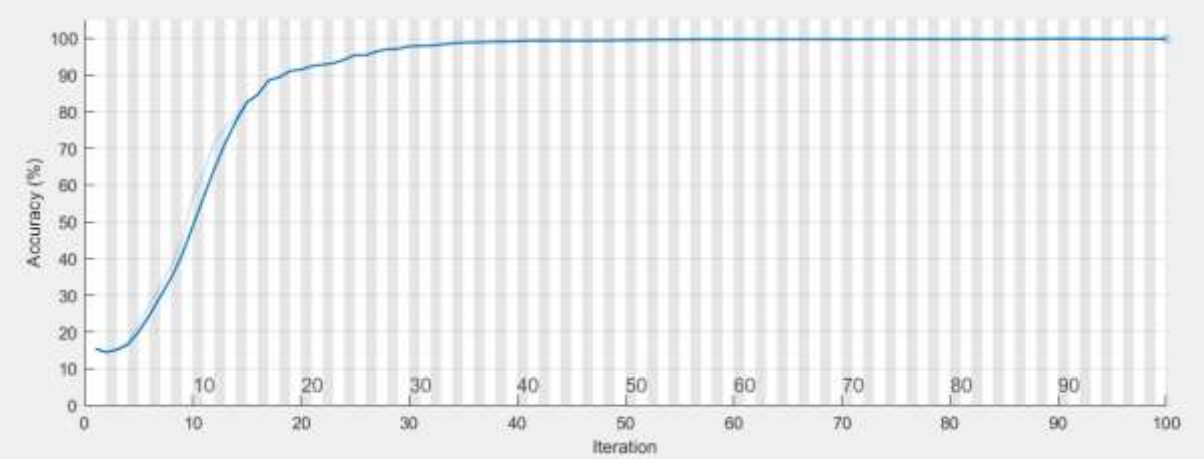


Fig 4 Accuracy comparison

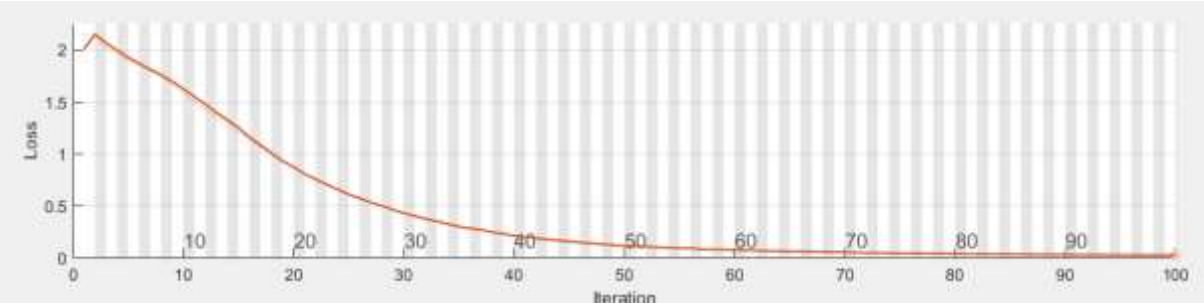


Fig 5 Loss comparison

Epoch	Iteration	Time Elapsed (hh:mm:ss)	Mini-batch Accuracy	Mini-batch Loss	Base Learning Rate
1	1	00:00:09	15.36%	2.0101	0.0010
50	50	00:03:07	99.49%	0.1207	0.0010
100	100	00:06:23	99.85%	0.0286	0.0010

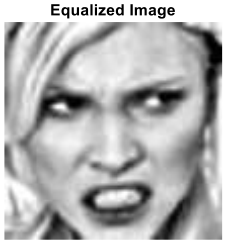
Fig 6 Loss and accuracy based on 100 epochs

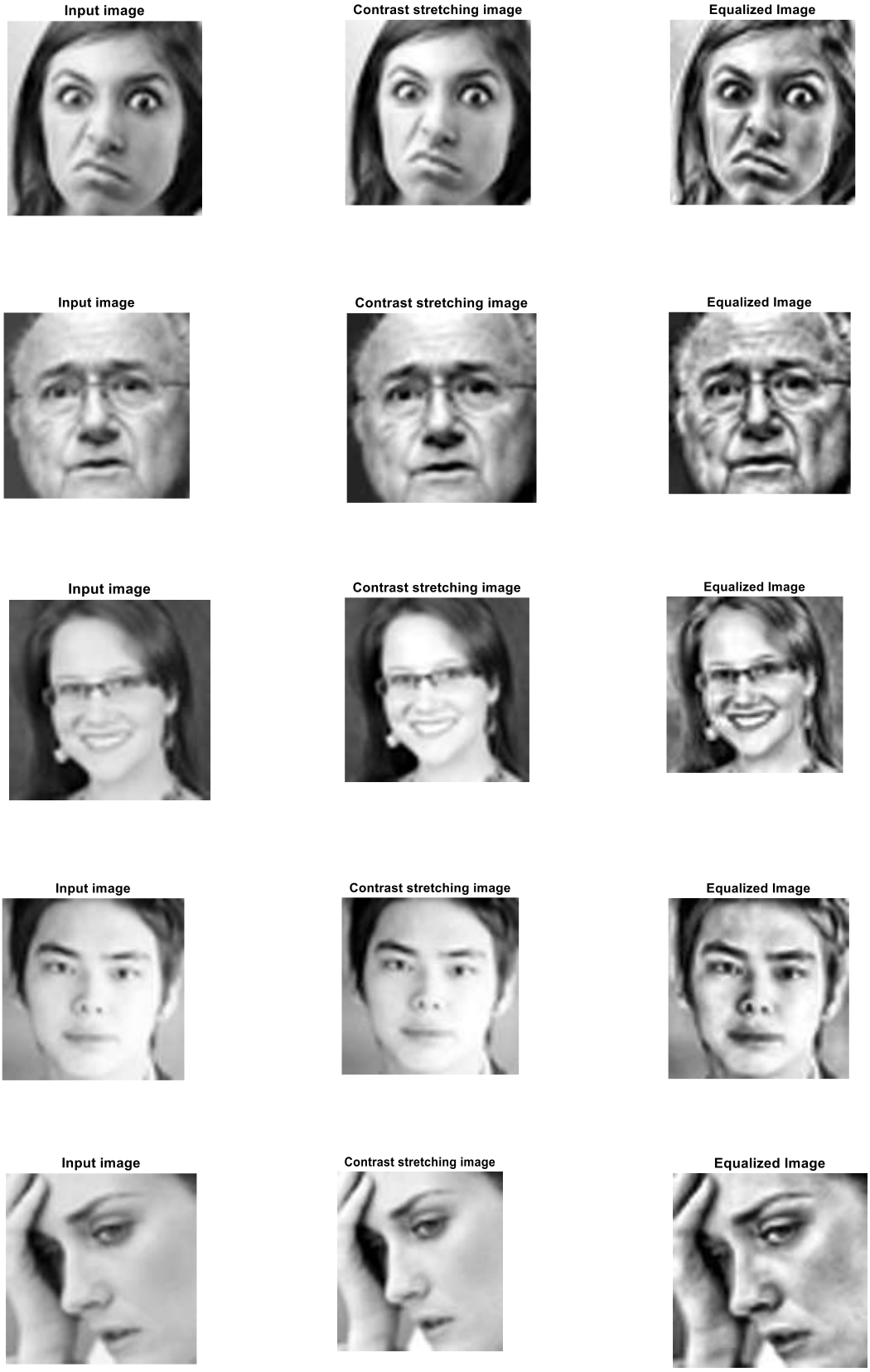
Fig 7 Facial emotion expression using sample data

Fig 7a Input images

Fig 7b Contrast stretching image

Fig 7c Equalized image




Input image
Contrast stretching image
Equalized Image
Input image
Contrast stretching image
Equalized Image
Input image
Contrast stretching image
Equalized Image
Input image
Contrast stretching image
Equalized Image
Input image
Contrast stretching image
Equalized Image

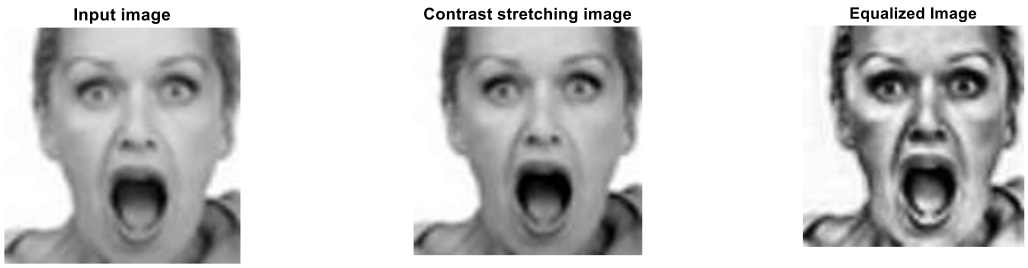


Table 5 Accuracy comparison of NMGC with other approaches

Approaches	Iteration 20	Iteration 40	Iteration 60	Iteration 80	Iteration 100
NMGC	98.6	96.72	95.94	95.94	93.24
CNN	70	72	68	70	66
GRU	68	70	72	72	68
ALEC	71	70	71	58	55

Table 6 Recall comparison of NMGC with other approaches

Approaches	Iteration 20	Iteration 40	Iteration 60	Iteration 80	Iteration 100
NMGC	98.33	98	97.3	97	97.5
CNN	92	96	92	90	94
GRU	88	87	87	89	90
ALEC	88	88	84	90	95

Table 7 Precision comparison of NMGC with other approaches

Approaches	Iteration 20	Iteration 40	Iteration 60	Iteration 80	Iteration 100
NMGC	99.72	98.3	99.8	98.2	98.99
CNN	60	63	61	60	60.8
GRU	56	55	61	55	62
ALEC	57	51	59	53	50

Table 8 F-measure comparison of NMGC with other approaches

Approaches	Iteration 20	Iteration 40	Iteration 60	Iteration 80	Iteration 100
NMGC	99.02	99.3	99.6	99.2	99.5
CNN	74	75	76	71	70
GRU	69	69	75	66	72
ALEC	64	69	73	63	65

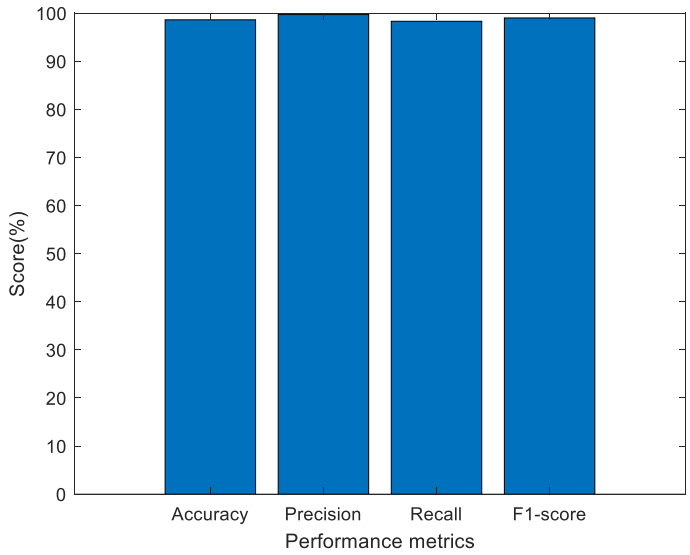


Fig 8 Performance comparison

Tables 4 to 8 show validation accuracy of anticipated NMGC with diverse performance metrics. For the first iteration, the NMGC's prediction accuracy is 98.63%, recall is 98.33%, precision is 99.72%, and F-measure is 99.02%. Fig 8 compares several measures including accuracy, recall, precision, and F-measure between NMGC, CNN, GRU and ALEC. The prediction accuracy is measured for five different iterations. In the initial iteration, the NMGC model shows 98.63% accuracy, 16.7967%, 19.1716%, and 16.1739% higher than CNN, GRU and ALEC models. For the first iteration, the suggested NMGC model's recall is 98.33% which is 0.832%, 4.6474%, and 5.0498% higher than CNN, GRU and ALEC models. The suggested NMGC model has a precision of 99.72%, which is higher than previous models by 15.2435%, 19.1634%, and 17.8588%. The suggested NMGC model has an F-measure of 990.2%, greater than previous models by 9.6842%, 14.5561%, and 19.9665%. Fig 7a to Fig 7c depicts the facial expression validation with the sample data and the appropriate classification of emotions using the NMGC model, respectively. Then, NMGC acquires the finest outcomes as convolution layer operation gives shape and texture characteristics of two diverse dimensions. Fig 8 shows the sample facial expression images evaluated with the proposed NMGC. The anticipated NMGC model is compared with various prevailing methods like CNN, GRU and ALEC giving a superior trade-off among the prevailing techniques.

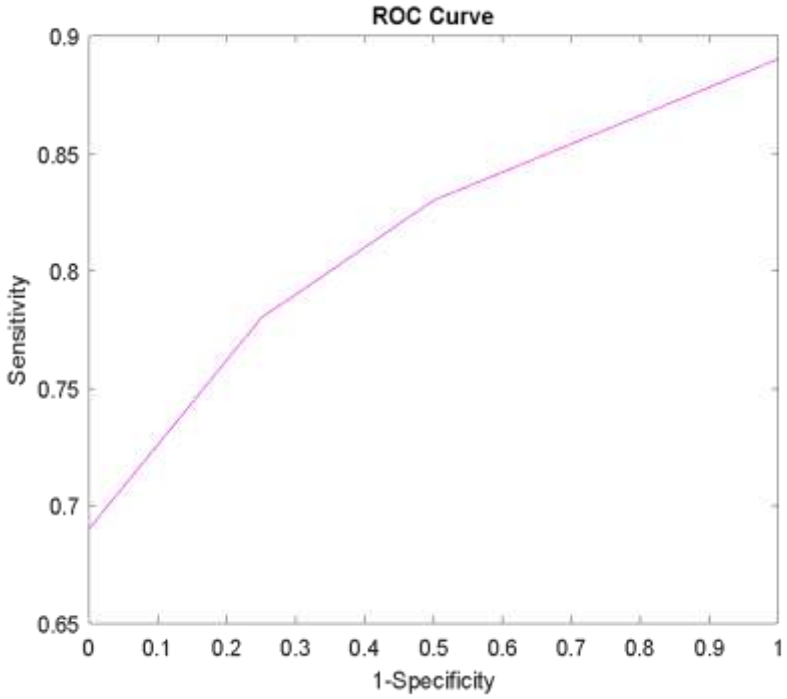


Fig 9 ROC curve

The suggested model's ROC computation with the sensitivity and specificity measures is displayed in Fig 9. First, NMGC performs the classification of facial expressions and classification outcomes compared to the anticipated model. The anticipated model works better than CNN in completing classification accurately. Henceforth, NMGCs can be satisfactory in the recognition process. An experimental study in the prevailing model shows that CNN has nine inputs and one output to show better results. NMGC model proves better specificity, accuracy, sensitivity, and AUROC results for emotional recognition datasets. However, it undergoes a manual feature extraction process, which is time-consuming.

5. CONCLUSION

This work presents a framework known as NMGC which is a hierarchically supervised model for recognizing human emotion. More specifically, anticipated models cast off deep NMGC to provide outputs of individual emotions. Here, some samples are trained with NMGC under the assistance of certain baseline architecture with effectual configurations. It includes numerous parameters and diverse initializations from trained models of emotion recognition databases. With the class probabilities of these individual networks, a hierarchical framework is formed dependent on NMGC. Next, this process is repeated with a set of NMGC as baseline architecture. The anticipated model is evaluated with a facial expression dataset, and its corresponding performance was measured with some strategies like CNN, GRU, ALEC and DF-CNN for computing the accuracy of NMGC over them where NMGC outperforms CNN, GRU and ALEC in some cases of experimental validation and proves to be superior. The future direction of this work is to develop and analyze a structural model that can internally merge classification NMGC with crafted features in the training and testing process. Here, both steps are performed separately, which provides advancements in offering feedback to one another during the training procedure. As this work employs CNN, which is a DL concept, thousands and thousands of images have to be provided for training. Therefore, online datasets are considered here. Real-time data will be collected and processed as a future research extension to attain superior outcomes.

REFERENCES

[1] Atmaja and M. Akagi, ‘‘Speech emotion recognition based on speech segment using LSTM with attention model,’’ in Proc. IEEE Int. Conf. Signals Syst. (ICSigSys), Jul. 2019, pp. 40–44.
[2] Schuller, ‘‘Speech emotion recognition: Two decades in a nutshell, benchmarks, and ongoing trends,’’ Commun. ACM, vol. 61, no. 5, pp. 90–99, Apr. 2018.
[3] Song, ‘‘Transfer linear subspace learning for cross-corpus speech emotion recognition,’’ IEEE Trans. Affect. Comput., vol. 10, no. 2, pp. 265–275, Apr. 2019.
[4] Pan, P. Luo, J. Shi, and X. Tang, ‘‘Two at once: Enhancing learning and generalization capacities via IBN-Net,’’ in Proc. Eur. Conf. Comput. Vis. (ECCV), Sep. 2018, pp. 464–479.
[5] Wei and Y. Zhao, "A novel speech emotion recognition algorithm based on wavelet kernel sparse classifier in the stacked deep auto-encoder model," Pers. Ubiquitous Comput., vol. 23, nos. 3–4, pp. 521–529, Jul. 2019
[6] Takaki, H. Kameoka, and J. Yamagishi, ‘‘Direct modelling of frequency spectra and waveform generation based on phase recovery for DNN-based speech synthesis,’’ in Proc. Interspeech, Aug. 2017, pp. 1128–1132
[7] Park, W. Chan, Y. Zhang, C.-C. Chiu, B. Zoph, E. D. Cubuk, and Q. V. Le, ‘‘SpecAugment: A simple data augmentation method for automatic speech recognition,’’ 2019, arXiv:1904.08779. [Online]. Available: http//arxiv.org/abs/1904.08779
[8] Neumann and N. T. Vu, ‘‘Improving speech emotion recognition with unsupervised representation learning on unlabeled speech,’’ in Proc. ICASSP - IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP), May 2019, pp. 7390–7394.
[9] Zeng, L. Dong, G. Chen, and Q. Dong, "Multi-feature fusion speech emotion recognition based on SVM," Proc. IEEE 10th Int. Conf. Electron. Inf. Emergency Commun. (ICEIEC), Jul. 2020, pp. 77–80.
[10] Assuncao and P. Menezes, ‘‘Intermediary fuzzification in speech emotion recognition,’’ in Proc. IEEE Int. Conf. Fuzzy Syst. (FUZZ-IEEE), Jul. 2020, pp. 1–6.
[11] W. Liu, W. L. Zheng, and B. L. Lu, “Emotion recognition using multimodal deep learning,” in Proc. Int. Conf. Neural Inf. Process., Kyoto, Japan, 2016, pp. 521–529.

[12] Sariyanidi, H. Gunes, and A. Cavallaro, “Learning bases of activity for facial expression recognition,” IEEE Trans. Image Process., vol. 26, no. 4, pp. 1965–1978, Apr. 2017.

[13] Liu, S. Shan, R. Wang, and X. Chen, “Learning expressionlets on spatio-temporal manifold for dynamic facial expression recognition,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., Columbus, OH, USA, 2014, pp. 1749–1756.

[14] S. Dinesh, K. Maheswari, B. Arthi, P. Sherubha, A. Vijay et al., “Investigations on Brain Tumor Classification Using Hybrid Machine Learning Algorithms,” Journal of Healthcare Engineering, Vol. 2, pp. 1-9, 2022.

[15] Keren and B. Schuller, “Convolutional RNN: An enhanced model for extracting features from sequential data,” in Proc. IEEE Int. Joint Conf. Neural Netw., Vancouver, BC, Canada, 2016, pp. 3412–3419.

[16] Mayya, R. M. Pai, and M. M. Pai, “Automatic facial expression recognition using dcnn,” Procedia Computer Science, vol. 93, pp. 453–461, 2016.

[17] Kim, H. Lee, J. Roh, and S.-Y. Lee, “Hierarchical committee of deep cnns with exponentially-weighted decision fusion for staticfacial expression recognition,” in Proceedings of the 2015 ACM on International Conference on Multimodal Interaction. ACM, 2015.

[18] Dhall, O. Ramana Murthy, R. Goecke, J. Joshi, and T. Gedeon, “Video and image based emotion recognition challenges in the wild: Emotiw 2015,” in Proceedings of the 2015 ACM on International Conference on Multimodal Interaction. ACM, 2015.

[19] Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” CoRR, vol. abs/1409.1556, 2014.

[20] Ciregan, U. Meier, and J. Schmidhuber, “Multi-column deep neural networks for image classification,” in Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on. IEEE, 2012.

[21] Krisshna, V. K. Deepak, K. Manikantan, and S. Ramachandran, “Face recognition using transform domain feature extraction and PSObased
feature selection,” Appl. Soft Comput., vol. 22, pp. 141–161, Sep. 2014.
[22] Liu, S. Li, S. Shan, and X. Chen, “AU-inspired deep networks for facial expression feature learning,” Neurocomputing, vol. 159, pp. 126–136, Jul. 2015.
[23] Zavaschi, A. S. Britto, Jr., L. E. S. Oliveira, and A. L. Koerich, “Fusion of feature sets and classifiers for facial expression recognition,” Expert Syst. Appl., vol. 40, no. 2, pp. 646–655, 2013.
[24] Diao, F. Chao, T. Peng, N. Snooke, and Q. Shen, “Feature selection inspired classifier ensemble reduction,” IEEE Trans. Cybern., vol. 44, no. 8, pp. 1259–1268, Aug. 2014.
[25] Zeng et al., “One-class classification for spontaneous facial expression analysis,” in Proc. 7th Int. Conf. Autom. Face Gesture Recognit., Southampton, U.K., 2006, pp. 281–286.
[26] Zeng et al., “Audio-visual spontaneous emotion recognition,” in Artificial Intelligence for Human Computing (LNCS 4451). Heidelberg, Germany: Springer, 2007.
[27] Cootes, C. J. Taylor, D. H. Cooper, and J. Graham, “Active shape models-their training and application,” Comput. Vis. Image Understand., vol. 61, no. 1, pp. 38–59, 1995.
[28] Edwards, and C. J. Taylor, “Active appearance
models,” in Computer Vision—ECCV98. Heidelberg, Germany: Springer, 1998, pp. 484–498
[29] Ojala, M. Pietikäinen, and T. Maenpaa, “Multiresolution gray-scale and rotation invariant texture classification with local binary patterns,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 24, no. 7, pp. 971–987, Jul. 2002.
[30] Liu et al., “Non-manual grammatical marker recognition based on multi-scale, spatio-temporal analysis of head pose and facial expressions,” Image Vis. Comput., vol. 32, no. 10, pp. 671–681, 2014.