International Journal of Computational and Experimental Science and ENgineering
(IJCESEN)

Vol. 10-No.4 (2024) pp. 918-927
http://www.ijcesen.com

ISSN: 2149-9144

Research Article

Elevating Facial Expression Detection: Empowered by VGG-19 and Weight-Normalized Gradient Boost Algorithm

M. Rajeshwari[1]*, K. Krishna Prasad[2]

[1]Research scholar, Institute of Computer Science and Information Science, Srinivas University, Mangalore
* **Corresponding Author Email:** rajimuraleedhar@gmail.com - **ORCID:** 0000-0001-9613-4967

[2]Professor &HOD, Cyber security and Cyber Forensics, Institute of Engineering and Technology, Srinivas University, Mukka, Mangaluru, Karnataka, India.
Email: krishnaprasadkcci@srininasuniversity.edu.in- **ORCID:** 0000-0001-5282-9038

Article Info:

DOI: 10.22399/ijcesen.521
Received : 17 October 2024
Accepted : 29 October 2024

Keywords :

FER2013 dataset,
multi-level feature extraction,
weight normalised Gradient boost algorithm,
multi-modal feature extraction.

Abstract:

Facial expression recognition termed as FER, is one of the vital tasks which is able in mimicking the human ability and wellness. Face expression and other gestures delivered via face are some of the important aspects of conveying non-verbal communications which plays a vital role in interpersonal relations. This domain has a higher scope of research due to enhancing computer vision and human-computer interaction. With respect and consideration to these domain interest, DL has gained a reasonable approach in performing FER. Some of the state-of-art approaches encompassing the ML approaches have faced several cases of laybacks such as overfitting, computational complexity, less adaptable to higher and big datasets. Thus, considering these laybacks and in achieving accurate FER in aspects of bringing proximal levels of FER, the current approach deployed DL methods for both feature extraction from the input image and in classifying them. Deep multi-level feature extraction is performed using the VGG-19 model and Weight Normalized gradient boosting algorithm is adapted for classifying these expressions using the FER13 dataset. This dataset constitutes the input images, which are in ranges of 28,709 sample image data and about 3,589 image data for test. These input images are initially pre-processed for obtaining better accuracy rates when performing feature extraction and classifications. The complete model in effective FER is evaluated using the performance metrics comprising Accuracy (98%), Recall (99%), F1-score (98%) and the precision rates (97%). This analysis of the performance will aid in affirming the overall efficacy of the proposed system.

1. Introduction

Humans, on a daily basis are instinct to interact among one another, in aspect of achieving a goal. The emotional state of a human can be reflected by their words, gestures and via facial expression. Among these actions, face is one of the complex element which are difficult to understand and in providing information regarding their moods. Several biometric methods is categorised into physiological and behavioural. Whereas, physiological methods are stable and non-alterable but behavioural patterns are sensible towards the human conditions like illness, stress and fatigue. Hence it should be detected through robust techniques [1]. FER, are one of the essential tasks in analysing the emotion, mental health and human interaction and their state of well-being. Several feature extraction model have been initiated, but are associated with some of the common limitations such as the template-based models are of with higher accuracy rates. But, these are complex and the images provided as input should be of same in size, orientation and in the levels of illuminations. Similarly, colour based models provoking segmentation, appearance-based models vitally consider the colour, quality and the illumination ranges of the image. Thus adapting the suitable form of feature selection model is vital in recognising the suitable facial expression. Since, Artificial Intelligence, simply termed as AI, are vital part in providing explicit solutions in domains of NLP, speech recognition, computer vision. Whereas, Computer vision is one of a vital aspects in making a proper FER as one of its vital application. Concurrently, there are several algorithms which are adapted in recognising the facial expression. Which range from Machine Learning [2] to Neural Network (NN) models. ML models such as KNN, SVM are widely adapted in case of FER issues. But NN models are more capable of achieving higher accuracy rates than the other such ML models. Moreover, it is utilized in the face recognition, where the learning corrects the classification coefficients by Eigen face procedure [3]. Several such studies in case of FER, are performed. This suggested study [4] used CNN as a primary model in performing FER. This model used a smaller dataset, as an aspect of considering that Deep CNN has an issue of overfitting in case of adapting big dataset. The aspect of using a smaller dataset is that these images do not have a clear features, such that a deeper model can learn limited range of features, and cause an issue of overfitting later. Thus, this recommended study considered using FER2013 dataset, to make the model reliable and works well in case of face images having lower resolution levels. This model has achieved an accurate rate of 69.3% and 85.3% for FER2013 and FER plus dataset correspondingly. Similarly, this suggested study [5] adapted Emotion Detection (ED) as a primary motive by implying UNED dataset, based on urudu language. This is present in paragraph and as sentence, containing six different primary emotions. The quality of corpus is evaluated in this study. The further class of classification are deliberated using the ML and DL approaches. Whereby, the DL models over performed the ML models in achieving the F1-score rate of about 85% using the UNED dataset and 50% on the paragraph-based corpus. Likewise, this advised approach [6] makes use of FER especially in drivers driving the automated cars. These cars are associated with some of liable issues such as ethics, social responsibility and in terms of cyber-security. Currently, to overcome these tasks, the automobile industries adapt a driver assistance modules and systems which can aid in careful driving an in handing their emotions. These can be used in the evaluation of monitoring the drivers in aspects of avoiding accidents, thus, this suggested approach makes use of Squirrel Search Optimization (SSO) along with FER containing DL. This is termed as SSO-DLFER. This approach targets in performing two major actions such as emotion and face detection. Retina Net model is adapted in aspects of performing face detection. Whereas, for emotion detection prospects, DLFER with NASNet is applied with the GRU, having large feature extractor. Hyper parameter tuning is performed based on SSO algorithm. Benchmark dataset is adapted, in the study where, comparative analysis were performed showing the efficacy of SSL-DLFER approach. Consecutively, this suggested system implies, FER in aspects of adapting the features which are derived from the pre-trained self-supervised models which are combined of text, audio and with visual forms of data modalities. This suggested approach makes use of a unique form of transformer. Which is attention-based, incorporated with multi-model semi-supervised learning model. This case achieved an accuracy rate of 86.4% of accuracy in aspects of performing multi-modal classification of emotion. Though these approaches are well adapted in case of performing FER, but at some instances, these approaches lay back in case of some limitations associated with confine using of larger dataset, overfitting issues, computational complexity and in cases of reading invariant features. To overcome these issues and laybacks, the proposed approach have adapted a DL model in aspects of performing both feature extraction and in performing classification of the emotion detected. In resolving these tasks, deep- multi level feature extraction is performed using VGG-19 model and classification is performed using the weight normalised gradient boosting model. The overall evaluation of the model are validated using the probabilistic performance metrics comprising the accuracy rates, F1-score ranges, precision and the ranges of recall values. The main aim and objectives of the proposed study are,

- To imply Deep Multi-level Feature extraction using VGG-19 model for accurate and high level feature extraction mechanism from the image dataset.
- To make use of Weight Normalized gradient boost algorithm in case of performing high end classification of the Facial Expression Recognition images.
- To evaluate the performance of the model using the performance metrics comprising accuracy rates, Precision, recall and the F1-score rates

1.2 Paper Organization

The following sections are categorized as, Section 2 discusses the review of conventional works with the suitable problems identified. Following this, section 3 expounds the proposed procedures with suitable flow, algorithms and mathematical derivations. Section 4 presents the results attained after simulating the proposed work and concluded in section 5 with future suggestions.

2. Literature Review

In this section, various state-of-art approaches are listed with their core concern used in the emotion detection by various facial expression.

This suggested approach makes use of Gaussian Mixture Model, which is known as GMM. This model is computed from each of sub bands, and from co-occurrence of features, which are specific to the sub bands. For the case of feature extraction, curvelet transform is applied, and each if the coefficients are applied in each individual scale of orientation. The invariance in rotation is obtained using the cycle-shift which is only around GMM features. This suggested study makes use of several and well-known dataset. The feature vector is analysed in terms of size, where it is observed to be dimensions are smaller and are better than the exiting rotation-invariant shoeing better performance. This study has better performance and are successful in cases of retrieving precision and recall rates [7].

The extraction of the context amid the neighbouring words and considering are vital in case of assessing the inter-modal notes prior the multimodal fusion. This tends to be one of the vital aspects in case of performing the research currently. This suggested approach handles both the analysis of inter-modal utterances in case of both sentiment and in performing emotion classification. These were performed on both CMU-MOSI dataset for performing sentiment analysis and IEMOCAP for emotion classification. The suggested model outperformed the standard forms of baselines over a range of 3% in accuracy for classification [8].

This recommended approach adapts curriculum learning approach, in detecting the facial expression at the period of training. Initially, clustering approach encompassing the unsupervised density-distance method used in determining the clustering centre of each of the category. The primary dataset used in classified to three various subsections, where, various complexity relying to distance from each of the sample to the centre is feature space is determined. In case of adding complexity, multi-stage training process is adapted. This model has indicated an accuracy rate of 72.1% of accuracy on FER2013 dataset and similarly 98.1% of accuracy for CK+ dataset classifying 7 different facial emotions [9].

The suggested system aims in extracting the facial landmarks of the particular subject. This is used in performing the training task which is suitable for the, designing of Deep RNN structure. These are used in evaluating two various continuous coefficients of the emotion which are to be stimulation and valence. This is done using the Russell’s model. At a final stage, user-friendly dashboard is created, which is used in representing both the momentary and the long-term fluctuations of the emotional state delivered by subject [10].

The recommended approach adapts multi-level feature extraction and fusion using the FER dataset. A single convolutional kernel is used which is adapted for performing the whole feature extraction process from the network. Feature fusion model connecting the global and the local attention. Various features are paired in the top-down way, aims in construction of the facial expression feature. Label smoothing and L2 regularization are also carried out in aspects of overcoming the issues related to data imbalance and overfitting issues. These have achieved an accuracy rate of 59.3% on AffectNet and varying accuracy rates under varies datasets [11].

This recommended study imparts Self-Supervised learning approach on making a representation learning using various data modalities. Initially, input modalities are passed as input. The features are extracted independently from the pre trained SSL models. Transformer and attention based fusion mechanism are introduced in aspects of performing multi-model emotion recognition. Benchmark dataset is used and are evaluated for four different datasets [12].

Micro-Expression Detection (MED) is performed in this state-of-art approach. This is accomplished using Extreme ML for MED. This is done because of having faster ability of learning the data and are higher in performance. SVM is used as the baseline model and the training time is compared with the ELM time for training. Feature extraction is done by analysing the Local Binary Pattern. Known as LBP. These are performed on the Chinese academy of science by collecting various expression samples. ELM had a better ranges of performance when compared to the SVM. Where, 0.34 seconds is taken by SVM for training and .04 seconds by ELM model [13].

Furthermore, ResNet-50 is used for efficient feature extraction to enhance facial expression recognition accuracy. At the same time, a convolutional attention mechanism is integrated to lessen the impact of unimportant areas in the feature map, resulting in an accuracy of 87.62% on the RAF-DB dataset and 88.13% on the FER2013 expression dataset [14].

RFER-EADL utilizes histogram equalization to standardize the intensity and contrast levels of images showing the same individuals and facial expressions. Then, the Chimp Optimization Algorithm (COA) is utilized as a hyperparameter-tuning method for the DenseNet-169 model.

Ultimately, expression recognition and classification utilize teaching and learning-based optimization (TLBO) in conjunction with a LSTM model. The COA and TLBO algorithms were used to help in selecting the best parameters for the DenseNet and LSTM models [15].

2.1 Problem Identification

Some of the core concerns from existing approaches are emphasized as explored below,

- SVM are less adaptable to the case of performing other forms of activation and for learning tasks which are less deployed [13] .
- Only low resolution images are capable of achieving higher accuracy rates, the model being less capable for adapting big datasets and images procured with higher resolution [16,17].
- Quality of multi-model features are less, which though enhances the accuracy of classifications but lay backs in enhancing the accuracy range of overall multi-modal system [8].

3. Proposed Methodology

Facial expression are the vital signs of expressing the emotion of inner well-being. FER, in the humans poses noteworthy trials due to their realistic occlusions, illumination, and head pose variations of the facial images. The proposed approach makes use of the FER13 dataset, in performing Face Expression Recognition. This is carried out in the insight of examining and in detecting the varying forms of emotions which are expressed via facial expressions. With a vital aspects of resolving the prevailing drawbacks which are procured with less accuracy rates, low convergence rates and redundant data levels. the aim of this proposed technique aims in predicting the appropriate expression obtained using the facial expression by adapting the some of the DL approaches encompassing Deep multi-level feature extraction technique using VGG-19 and the classifying these emotions delivered using Weight normalized Gradient Boost algorithm.

Initially, the input images from the FER-13 dataset are obtained using appropriate image pre-processing procedures which aims in proper feature extraction mechanism. These pre-processing techniques are able to enhance the performance of the face expression recognition levels. This image pre-processing techniques procured in the proposed study makes the image resizing technique, which is used in making the image to proper in the pixel level, clear in image size, and are more suitable for further image data processing. The complete overview of the proposed approach using flow representation is presented in figure 1.

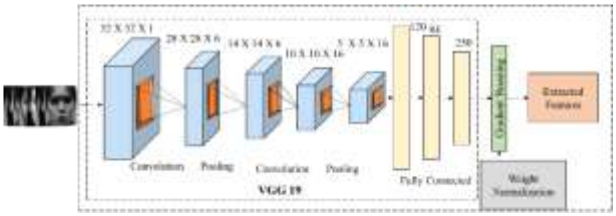

Figure 1. *Overall Flow of the proposed approach*

This procedure is followed by the technique called feature extraction, where the proposed approach makes use of deep multi-level Feature extraction which are opted for a more relevant and optimal feature extraction process which aims in by reducing the other redundant data additionally bringing advantages such as by preventing over fitting issues, leveraging pre-trained knowledge and in bringing optimal outcomes. Concurrently, Weight normalized Gradient boost mechanism is adapted as a specific approach for efficient classification techniques, which aims in bringing advantages by stabilizing the learning phase of the model for bringing accurate classification results. Moreover, it is flexible, interpretable and have less computational power and have the capability to perform normalisation over the individual features. This procedure is followed by effective data split into train test data in ratios of 80:20 respectively. These train data are taken for the process of classification (training phase) which are done using the weight normalized Gradient boost model. The final validity of the model is examined using the probabilistic performance metrics, embracing the accuracy levels, precision rates, F1-score and the recall rates.

3.1 Deep multi-level feature extraction using VGG-19

VGG19 is learned with operative features for standard image recognition that frequently simplify fine towards definite tasks such as facial emotion recognition. It bring about better efficiency on the FER13 dataset associated towards training a method against scratch. Original images of FER13 can be exactly high-dimensional. Besides, the model is trained from the beginning on such data can cause overfitting, whereas the model learns precise facts of the training images relatively than learning generalizable structures. Feature extraction serves as a method to reduce dimensionality. The pre-trained features of VGG19 capture general image characteristics that are potentially important for different facial expressions, despite small differences in FER13. This assists the following model in concentrating on understanding patterns related to emotions in these features, resulting in improved generalization on different data. This model is capable of extracting both the low-level and high-level features of images from layer by layer, and providing a realised form of feature extraction. In case of VGG-19 model, comprising 2*64 + 2*128 + 4*256 + 2*4*512 a total of 5504 neurons usually one are able to measure the size of the network in aspects of number of parameters (i.e. weights) and number of layers encompassed. VGG-19 has about 144 million parameters which is notably large. The attributes in the VGG-19 model are concentrated for 3 FC (Fully Connected) layers. These parameters of the network are initially designed for performing 1000 classification. The complete Pseudocode explaining the VGG-19 is presented in Pseudocode-I.

Pseudocode-1 *Deep* $multi\ level\ feature\ extraction\ VGG-19$
1. Pre − process the images in the dataset (e.g., crop, resize, normalize).
2. Split the dataset into training, validation, and test sets.
3. Intialize the input images with shape (height, width, channels) and assign it to the variable y
4. For block 1:
5. Reapet step 2 for i = 1 to 2:
6. Apply ReLU activation on dot product of y and We bias[i], assign the output to y
7. Apply max pooling operation with stride 2 on y, as
8. For block2:
9. Repeat step 3 for i = 3 to 4:
10. Apply ReLU activation on dot product of y and W bias[i], assign the output to y.
11. Apply max pooling operation with stride 2 on y, a
12. Repeat step 3 for blocks 3,4,5
13. Apply flatten operation on y to convert it to ID ar
14. For i = 5to 7, repeat step 7:
15. Apply ReLU activation on dot ptoduct of y and W bias[i], assign the output to y
16. Apply softmax activation on dot product of y and bias[8], assign the output to y.

3.2 Weight Normalized Gradient boosting algorithm

Gradient Boosting is one of the framework used in boosting. The main aim of this GB model is to perform a sequent build of individual Decision Tree (DT) model on the direction of gradient descent of a particular Loss Function (LF). This loss function is based on the residual, which is meant to be the difference among the predicted and the true value at each of the instance occurred from previous models. If the LF value decreases with respect to the addition of new models, the prediction power of these model enhances. Concurrently, the LF value is declined in the direction of their respective gradient descent. M is considered to be the hyper parameter of the present algorithm. The increase in value of M diminishes the error on occurring in the training set, but at higher range can result in overfitting of the model. If the $m-1$ models learned previously, then mth the prediction model is built. Whereas, in $h(xi;\ am)$, are used in obtaining the parameters (am) by fitting it to the pseudo-residuals, by making use of least square method to make sure the new model, $h(xi;\ am)$, wchih aims in achieving a minimum value of the gradient direction. Next, it computes the coefficient/multiplier of the new model using the loss function by solving the one-dimensional optimisation problem. Finally, with the decision model h(x; am), GB updates the model $(Fm(x))$ via linear superposition. Small learning rates typically improve the model's generalisation ability over gradient boosting without shrinking at the rates of increasing the time of training and querying.

Pseudocode-2 *Weight Normalization based Gradient boosting algorithm*
Intialize for all i: $\varphi_i(i) = \frac{1}{n}$ initialize the weights
for $l = 1$ to D do
for all i: $z_l(i) = \frac{\varphi_i(i)}{\sum_i \varphi_i(i)}$ compute normalized weights
$h_1 :=$ Learn(V, z_1) learn with normalized weights
$E_l = \sum_i z_i(i)[h_1(a_I \neq b_I)]$ call Learn with normalized weights
if $\varepsilon_l > then$ calculate the error of h_i
$D = l - 1$
go to 12
$\beta_l = \frac{\varepsilon_l}{1-}\ \frac{\varepsilon_l}{1-\varepsilon_l}$
for all i: $\varphi_{l+1}(i) = \varphi_i(i)\beta_l^{1-[h_l(a_i \neq b_i]}$
end for
output: $h_{final(a)=\underset{b\in B}{argmax}}.\ \sum_{l=1}^{D}(\log\frac{1}{\beta_l}[h_l(a)=b]$

4. Results and Discussions

The particular section deals with the overall results which are obtained using the FER13 dataset in case of FER detection using the DL techniques adapting optimal feature extraction and efficient classification methods.

4.1 Dataset Description

The FER13 dataset consists about 48*48 pixel levels of greyscale images of the faces. These faces are registered in the automated form which is more or less centres and occupies the same amount of space in each of the image. The main aim of the task used via these dataset are in categorizing the face based emotion which shows several face expression into 7 different classes or categories. For each expression the values are set from 0 to 6.The training data consists of 28, 709 sample image data and the test data consists of about 3, 589 image data.

4.2 Exploratory Data Analysis

EDA is adapted in aspects of validating the data using various visualization methodologies. For case, EDA is applied in defining the patterns or to legalise the assumptions. These are done using either a graphical representations or using a statistical summaries. Furthermore, EDA compromises the data which is used in understanding the entire dataset. This section represents, the confusion matrix of the proposed model, which is used in depicting the correct and the incorrect predictions made by the model. Further, the section represents the number of emotion and their count regarding various classes in Figure 2, as happiness, sad, anger and many. Concurrently, various emotions depicted by the individuals such in aspects of being anger, sad, happy, neutral, cry are also represented for further case of reference.

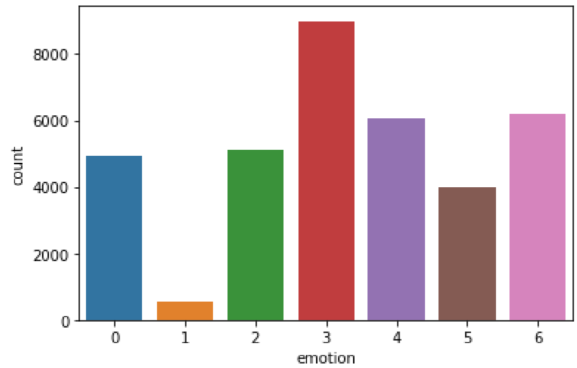

Figure 2 *Graph representing various emotions and their count*

Figure 3 *representing varying emotions*

Figure 4 *representing varying emotions*

Similarly, figure 3 and 4 showing the various emotions which are expressed by the various facial expressions by the individuals encompassing numerous human emotions.

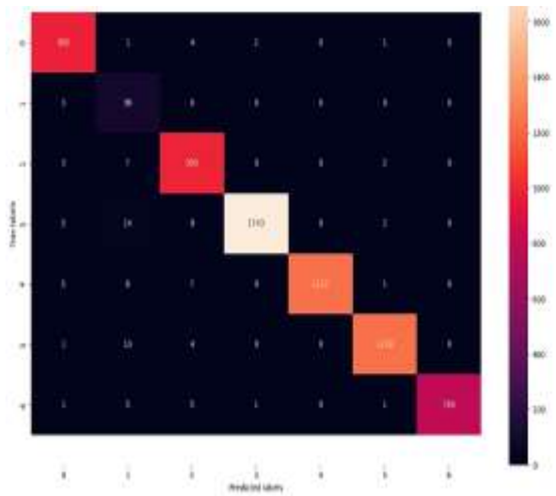

Figure 5 *Confusion Matrix of proposed model*

The confusion matrix presented in figure 5, is used in representing the correlation coefficients among the variables. This is also used in showing the possible pairs of the values which can be made from the table. This is one of a powerful tool in identifying and in visualising the patterns in the given data.

4.3 Performance Metrics

4.3.1. Accuracy

This metric is used in giving the measure of the model proposed across all the classes.

$$Accuracy = (TP + TN)/(TP + TN + FP + FN) \quad (1)$$

4.3.2 Precision

This metric is used in the retrieval of the information, and the instances considered vital for the model performance. These are the ratio of the true positives to the total true positive and the total false positive rates.

$$Precision = TruePositives / (TruePositives + FalsePositives) \quad (2)$$

4.3.3 Recall

This metric is used in defining the model detecting the positive instances to the total number of true and the false incidents of the proposed approach.

$$\text{Recall} = \text{TruePositives} / (\text{TruePositives} + \text{FalseNegatives}) \quad (3)$$

4.3.4 F1-Score

F1 score is denoted as the weighted harmonic-mean value of precision and recall, the F1 score is estimated with the following equation

$$F1 - score = 2 \times (Rc \times Pc)/(Rc + Pc) \quad (4)$$

where, P is denoted as precision and R is denoted as recall.

4.3.5 Time Complexity

This metrics is defined to be the quantified time of an algorithm to run as a function. This is calculated with respect to the length of input. This is given by,

$$time\ complexity = O(tdxlog_n) \quad (5)$$

Here, t is the number of trees, x represents the number of missing entries found in the training data.
Similarly, the prediction of new samples is given using$o(td)$

4.4 Comparative Analysis

FER13 dataset are adapted and are used in analysing various Facial Expression and are analysed for their performance which are compared using various DL techniques used in the FER, such as EmNet, CNN, CNN with de-noising techniques and threshing machine techniques.

Table 1 *Model comparison among proposed and existing method in aspects of accuracy*

MODEL	Accuracy
DCNN Model1	72
DCNN Model2	72.02
EmNet (average fusion)	74.11
EmNet (weighted maximum fusion)	74.06
Proposed	98

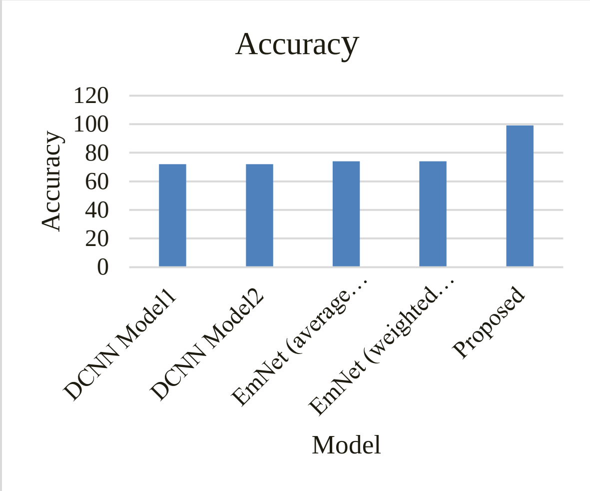

Figure 6 *Model comparison among proposed and existing method in aspects of accuracy*

This section will make an easier method of understanding the enhanced viability of the proposed model than the other existing models in terms of Accuracy, Precision, recall and on the basis of F1-score rates. Table 1 and Figure 6 clearly states that the proposed model have outperformed the existing approaches in terms of detecting various emotions expressed by humans. Whereas, the existing models such as DCNN 1 and 2, EmNet (average fusion and weighted maximum fusion) is compared with the proposed model that uses VGG-19 with weighted normalization and it shows the accuracy of 98%. While the other models have achieved the accuracy rate not as that of proposed model. From these results it is found that the emotions of the humans are identified accurately through the proposed model.

Table 2 *comparison among existing and proposed model in terms of accuracy rates [18]*

MODEL	Accuracy
Existing model	65.97
Proposed	98

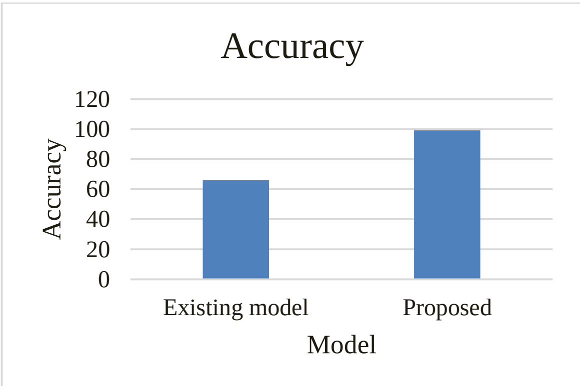

Figure 7 *comparison among existing and proposed model in terms of accuracy rates*

Table 2 and Figure 7 showing the existing model using CNN and Raspberry Pi, in terms of detecting the emotion has achieved an accuracy rate only in ranges of 65.6% whereas using the proposed model, the model achieved 98% which is explicitly greeter than the existing approach in ranges of 34% showing the efficacy of the proposed model. Moreover, the VGG-19 with weighted normalization has showed the higher results when compared with the CNN and Rasberry Pi because in the model the implements a multiple feature extraction, hence it achieves greater accuracy rate.

Table 3 *Comparison of proposed with state-of-art approaches in terms of train and test accuracy [16]*

MODEL	Accuracy	Precision	Recall	f1_score
SGD	76.17	63.0118	61.0729	61.0932
Adam	77.17	66.6236	66.8845	66.6779
Proposed	98	97	99	98

Figure 8 and table 3, it is clear that the prosed model have attained higher rates of both train and test accuracy rate of 99%. Whereas, the existing approaches making use of convolutional net models have attained an accuracy ranges only in rates of 58-61%. This shows that the proposed model have greater accuracy rate in scales of being 38% greater than the state-of-art approaches.

Table 4 *Comparison of Proposed Model with Other Approaches In Terms of Various Performance Metrics [19]*

MODEL	Train Accuracy	Test Accuracy
FERConvNet_Gaussian	0.98	0.58
FERConvNet_Bilateral	0.98	0.63
FERConvNet_Nonlocal Means	0.93	0.61
FERConvNet_HDM	0.98	0.95
Proposed	0.98	0.98

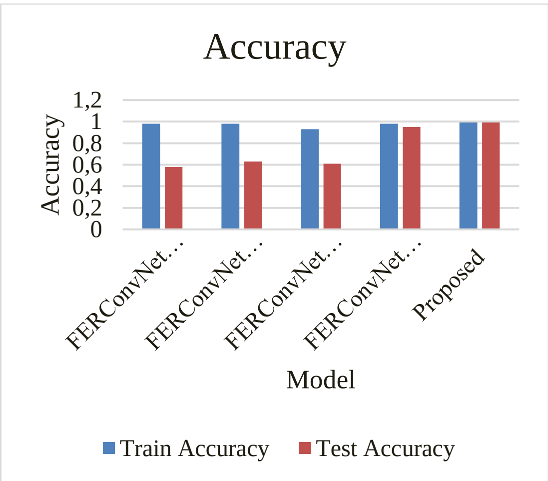

Figure 8 *Comparison of proposed with state-of-art approaches in terms of train and test accuracy*

During the training and testing of data, the dataset of FER with Convolution Net along with Gaussian,Bilateral,Nonlocal means and HDM is compared with the proposed model, where the FER with proposed model has performed well with the accuracy rate of 0.98 in train and test data.

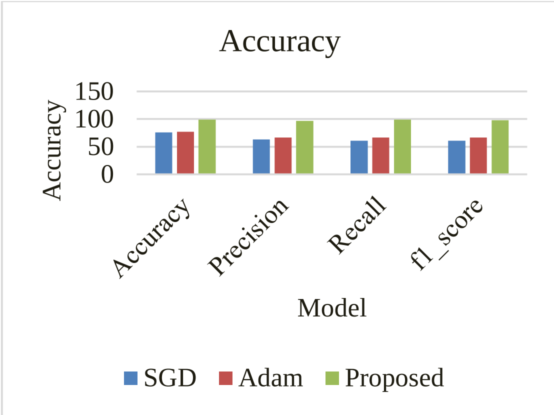

Figure 9. *Comparison of Proposed Model with Other Approaches In Terms Of Various Performance Metrics*

Figure 9 and table 4 showing the proposed model performance in various aspects encompassing the accuracy rates, precision rates, recall and the F1-score rates in comparison to other sorts of approaches used in face expression or emotion detection. The proposed model have achieved an accuracy rates in ranges if 98%, followed by the precision and the recall rates in scales of 97 and 99% respectively, Finally the proposed model attained an F1-score range in scales of 98% depicting the efficacy of the model in all sorts of performance parameters. Hence, the result obtained by the proposed model is compared with the SGD and Adam model and it shows that it outperforms well. Thus, as an impact of adapting these effective feature extraction mechanism and the classification protocol the proposed study have achieved higher ranges of performance metrics and had implicated its efficacy in FER.

5. Discussion

This sections explains about the proposed model has attained a reliable outcome when compared with the traditional methods. The VGG model is mainly focussed on the face recognition and emotions in the FER-13 data. Besides, the gradient boosting with weight normalization is addressed to reduce over-fitting. But the conventional methods such as SVM may need complex approach to avoid over-fitting for specific limited data. Moreover, proposed method is compared with other DL methods for interpretability and efficiency. Hence, it is defined that proposed method has attained greater results when compared with the other conventional methods. Moreover, the proposed method is fine-tuned for the particular features of FER-13 along with the recognition of emotion tasks. Hence it is used in the various applications such as student attendance, security purposes and in hospital environment.

6. Conclusion

The research aimed in performing FER using two vital and primary concepts encompassing the feature extraction mechanism and the classification procedures. These are adapted for analysing the various facial expression expressed by the human beings in aspects of expressing their inner being. For performing these procedures, VGG-19 is used in case of achieving Multi-level deep feature extraction. Similarly, Weight normalised Gradient boost algorithm is adapted in performing effective classification of these emotions. Confusion matrix for the proposed method is also presented in case of establishing the efficacy of the model by their rates of correct predictions made. Though, to attain knowledge over the better performance of the proposed model than state-of-art approaches, comparative analysis were carried out with regards to four metrics probabilistic metrics including (accuracy, recall, F1-score and precision). From the comparative results, it is exemplary in case that, the proposed method discovered outstanding performance than existing approaches with an accuracy rate of 98%, 99% in recall, 98% f1-score and 97% precision. In aspects of future work, the proposed model can be implied with various DL models and approached for performing huge classes of human emotion detection. However, the proposed method is advantageous it has some limitations such as it takes certain time for learning process. In future, it will be implemented for enhanced security options, contactless verification features along with faster procedure for amplified productivity.

Author Statements:

- **Ethical approval:** The conducted research is not related to either human or animal use.
- **Conflict of interest:** The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper

- **Acknowledgement:** The authors declare that they have nobody or no-company to acknowledge.

- **Author contributions:** The authors declare that they have equal right on this paper.

- **Funding information:** The authors declare that there is no funding to be acknowledged.

- **Data availability statement:** The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

References

[1] J.-H. Kim, N. Kim, and C. S. J. a. p. a. Won, "Multi-Modal Facial Expression Recognition with Transformer-Based Fusion Networks and Dynamic Sampling," 2023.

[2] C. Caglio, R. M. Darst, and S. Kalemli-Ozcan, (2022). Collateral heterogeneity and monetary policy transmission: Evidence from loans to smes and large firms. Working Paper. NATIONAL BUREAU OF ECONOMIC RESEARCH Cambridge, MA 02138 DOI 10.3386/w28685

[3] M. Wafi, F. A. Bachtiar, F. J. I. J. o. E. Utaminingrum, and C. Engineering, (2023). Feature extraction comparison for facial expression recognition using adaptive extreme learning machine. *International Journal of Electrical and Computer Engineering (IJECE)* 13;1113-1122, DOI: 10.11591/ijece.v13i1.pp1113-1122

[4] B. B. Mamatkulovich, M. S. Shuhrat o'g'li, and B. J. J. O. A. R. Jasurjonovich, (2023). SPECIAL DEEP CNN DESIGN FOR FACIAL EXPRESSION CLASSIFICATION WITH A SMALL AMOUNT OF DATA. *Open Access Repository* 4 (3), 472-478.

[5] M. F. Bashir, A. R. Javed, M. U. Arshad, T. R. Gadekallu, W. Shahzad, M. O. J. A. T. o. A. Beg, *et al.*, "Context-aware Emotion Detection from Low-resource Urdu Language Using Deep Neural Network," *ACM Transactions on Asian and Low-Resource Language Information Processing,* 22(5);1 -30 https://doi.org/10.1145/352857vol. 22, pp. 1-30, 2023.

[6] D. K. Jain, A. K. Dutta, E. Verdú, S. Alsubai, A. R. W. J. I. Sait, and V. Computing, (2023). An automated hyperparameter tuned deep learning model enabled facial emotion recognition for autonomous vehicle drivers. *Image and Vision Computing* 133;104659, https://doi.org/10.1016/j.imavis.2023.104659

[7] Engin, M.A., Cavusoglu, B. (2019). Rotation invariant curvelet based image retrieval & classification via Gaussian mixture model and co-occurrence features. *Multimed Tools Appl* 78, 6581–6605. https://doi.org/10.1007/s11042-018-6368-8

[8] Huddar, M.G., Sannakki, S.S. & Rajpurohit, V.S. (2020) Multi-level context extraction and attention-based contextual inter-modal fusion for multimodal sentiment analysis and emotion classification. *Int J Multimed Info Retr* 9;103–112. https://doi.org/10.1007/s13735-019-00185-8

[9] Liu, X., Zhou, F. (2020). Improved curriculum learning using SSM for facial expression recognition. *Vis Comput* 36, 1635–1649. https://doi.org/10.1007/s00371-019-01759-7

[10] Kansizoglou, I., Misirlis, E., Tsintotas, K.A., & Gasteratos, A. (2022). Continuous Emotion Recognition for Long-Term Behavior Modeling

through Recurrent Neural Networks. *Technologies*.10;59

[11]Xie, Y., Tian, W., Zhang, H. *et al.* (2023). Facial expression recognition through multi-level features extraction and fusion. *Soft Comput* 27, 11243–11258. https://doi.org/10.1007/s00500-023-08531-z

[12]S. Siriwardhana, T. Kaluarachchi, M. Billinghurst and S. Nanayakkara, (2020). Multimodal Emotion Recognition With Transformer-Based Self Supervised Feature Fusion, in *IEEE Access*, 8;176274-176285, doi: 10.1109/ACCESS.2020.3026823.

[13]I. P. Adegun and H. B. J. S. A. Vadapalli, (2020). Facial micro-expression recognition: A machine learning approach. 8;e00465,. https://doi.org/10.1016/j.sciaf.2020.e00465

[14]M. Aly, A. Ghallab, and I. S. Fathi, (2023). Enhancing Facial Expression Recognition System in Online Learning Context Using Efficient Deep Learning Model, *IEEE Access*, 11;121419-121433.

[15]M. Arul Vinayakam Rajasimman, R. K. Manoharan, N. Subramani, M. Aridoss, and M. G. Galety, (2022). Robust facial expression recognition using an evolutionary algorithm with a deep learning model," *Applied Sciences*, 13;468.

[16]Bodavarapu, P.N., & Srinivas, P.V. (2021). Facial expression recognition for low resolution images using convolutional neural networks and denoising techniques. *Indian Journal of Science and Technology*. 14;971-983.

[17]K. Xie, G. Cai, G. Kaddoum, and J. J. a. p. a. He, (2023). Performance Analysis and Resource Allocation of STAR-RIS Aided Wireless-Powered NOMA System,. https://doi.org/10.48550/arXiv.2301.08865

[18]L. Zahara, P. Musa, E. P. Wibowo, I. Karim, and S. B. Musa, (2020). The facial emotion recognition (FER-2013) dataset for prediction system of micro-expressions face using the convolutional neural network (CNN) algorithm based Raspberry Pi, in *2020 Fifth international conference on informatics and computing (ICIC)*, 2020, pp. 1-9.

[19]Saurav, S., Saini, R. & Singh, S. (2021). EmNet: a deep integrated convolutional neural network for facial emotion recognition in the wild. *Appl Intell* 51, 5543–5570. https://doi.org/10.1007/s10489-020-02125-0