Transforming Facial Expression Prediction: Amplifying Accuracy with ResNet50 Features and Innovated XG-Boost Algorithm

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

[*1]Research scholar, Institute of Computer Science and Information Science, Srinivas University, Mangalore.
[2]Professor & HOD, Cyber security and Cyber Forensics, Institute of Engineering and Technology, Srinivas University, Mukka, Mangaluru, Karnataka, India.
[*1]Corresponding Author: rajimuraleedhar@gmail.com. Orcid ID: 0000-0001-9613-4967
[2]Co-author mail Id: krishnaprasadkcci@srinivasuniversity.edu.in. Orcid ID: 0000-0001-5282-9038

Abstract
Facial expression is one of the most natural and a non-verbal way in expressing human emotions and interactions. Assessing the well-being via communication are the first signs that transmits the emotional state. This domain attracts more of research and are interested by the modality in the specificity of the domain. With a suspect able increase in the AI in domains, gaining a reasonable implementation of DL with its advanced ability in case of Facial Expression Detection known as FER. Existing studies adapting ML approaches, in the FER, have deployed in adequate laybacks such as high computational time, inability to deal with large datasets, and fail in bringing timely accurate ranges of predictions. In consideration to these aspects, the proposed study admits the system design, in aim of FER comprising feature extraction and the classification of the emotions using the DL models. These are done using the proposed approach uses Deep Multi-level feature extraction using ResNet50, which is more appropriate in optimal and exact feature selection mechanism. Followed by Weight-normalized XG-Boost classifier for the process of classifying various emotional expressions. This is adapted in aim, of maintaining the gradient descent step and admitting using larger dataset for learning. The input images are collected from the FER13, dataset consisting 28,709 sample image data and the test data consists of about 3,589 image data. These are initially pre-processed for better accuracy rates during feature extraction and classifications. The complete model in effective FER is evaluated using the performance metrics comprising Accuracy, Recall. F1-score and the precision rates. This analysis of the performance will aid in affirming the overall efficacy of the proposed system.

Keywords Facial Expression Recognition, Weight Normalized XG-Boost Algorithm, Deep multi- Level Feature Extraction, ResNet50.

(Received December 12th, 2023 / Accepted June 28th, 2024)

I. Introduction

The facial expression from individual's emotions are the key features in establishing the human interaction [1]. Deceptively, both humans and animals are able to produce more specific form of muscle movements and interactions which expresses a certain inner mental state. Amid these features, particularly, the human face expression contribute more of interesting and observable form of expressions comprising happiness, fear, sad, neutral state, excitement and disgust. These are the expressions which are commonly used in practice than the other uncommon facial features. Though capturing of the facial images are associated with less palpable and are highly invasive than the other emotional form of facial features. Recognizing and classifying these forms of facial emotions are contributed with various applications in fields of computer research areas, which constitutes computer interaction, animation, security and in faith criminal resolved issues [2].

Recently, Automatic Facial Expression Recognition (FER), which is particular and aims in analyzing and in understanding the inner emotions of an individual via human facial behavior [3]. This domain have now become a hot research field, especially in the domains of Computer vision, Artificial Intelligence (AI) and also in the Pattern recognition fields. This is all because of FER having a wide and potential scale of solicitations, such as Human Emotion Perception recognized as HEP, Social Robotics, human computer interaction and in healthcare sectors [4]. Researchers in this domain are enhancing the case of interest to interpret, code the facial expression delivers and in extracting the features for making a better prediction by digital systems. The emotional state delivered by humans can be obtained by both the verbal and using the non-verbal data which are captured using various sensors, in detecting the facial changes [5]. With a remarkable and a notable success made using the DL techniques in case of FER, various architecture in these province are exploited in achieving better rates of performance. The process of feature extraction from one face to the other is a difficult and are of sensitive task in making a better part of classification. Recently, DL have been efficacious and an proficient form of approach, for allowing an automated form of feature extraction, and in the process of classification using Convolutional Neural Network (CNN), Recurrent Neural Network (RNN), where, these techniques are prompt for using them in the recognition of human emotions. But these methods are only suitable in making the classification process using the 1D and the 2D images, and devoid in acting with the depth of the data. Several other case of approaches have been carried out by the researchers in initiating to use these techniques in recognizing various human emotions using NN architectures, which can be accomplished in establishing satisfactory results in these fields [6].

One of the success determining factors of DL is, the training the NN with several FER databases, and each one of them differing with various levels in size of images, videos and with varying levels of illumination, population and in face pose. Despite the notable grades of success in the traditional FER, which are extracted from difficult cases of handcrafted features, where over the past few decades, researchers have directed along the DL approaches, only due to its high scalability in exposing automatic recognition capacity [7]. Recently, Machine Learning (ML) techniques have expressed its revival mostly in case of progress in DL with advanced ranges of optimization techniques. In this context, the proposed study have expressed the FER using some of the DL techniques, in obtaining better rates of prediction. DL are used in various and in numerous fields of research which relates to image recognition, from the ranges of urban dynamic till the general tracking of the human and animal expression [8].

These are some of the common and crucial cases of attributes used in the fields of neuroscience, digital pathology and in cases of behavioral biology. These DL networks are successfully used in present, under many applications, involving the acquisition of the data about the features, of a particular image that are only perceived by humans and animals. As in case of advancement, these FER are becoming more common in terms of collecting the acoustic data which are described at a petabyte scale levels, which in turn can be used in making complete analysis being reasonable and feasible [9]. As a result, the present research now tend to grow in case of automated detection and in the classification systems recognized as DCS, being more widespread and tend to continue in use. From a perspective of ML in the fields of DCS, which is more hierarchical in model. This consisting both the binary classifier and in recognition whether the signal of interest. This is present in the acoustic recording combined with the multiclass classifier used in determining the source of signal. Moreover, adapting the DL techniques which is more potential in building the intelligent systems in the accurate recognition of the emotions are closer in reality [10]

Whereas, the annotation of the large-scales of facial expressions are difficult in case of uncertainties which are due to the ambiguous facial expressions, low-quality facial expressions and images and the subjectiveness of the particular annotators. These uncertainties will enhance from high-quality to low-quality and even in case of micro-expressions. These case of uncertainties will result in inconsistent and incorrect labels, which results in suspending the progress of large-scale FER. Moreover, training the DL with uncertainties of FER, may result in several issues such as Overfitting, mislabeled data, and making the model unsuitable for learning. This results in the model disconvergence, leading to early stage of optimization [11]. To overcome these issues, the multi-level feature extraction is carried out in the proposed study using the ResNet 50, which is used in extracting the rich and the deep levels of the feature present in the image data. These data are then split into train and test data, where the train data are used in the classification of the emotion delivered via facial expression. These data are classified using the XG-boost model which is based on the process of weight normalization, which is advantageous in making a stable gradient descent step and can enable using higher learning rates and aids in faster convergence for the provided learning rate. The complete ability of the model is validated using the test data, via the performance metrics comprising the accuracy, Recall rates, F1-score and the precision rates.

1.1 Aim and Objective

The main aim of the study lies in accurate detection of FER, using different DL approaches achieving the aim by adhering to the objectives projected below.

- To implement Deep Multi-level Feature Extraction using ResNet 50 for accurate and high level feature extraction mechanism from the image dataset.
- To use Weight normalized XG-Boost algorithm for 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

Section II deliberates the review of conventional works with the problems identified. Following this, section III expounds the projected procedures with apposite flow, algorithms and their mathematical derivations. Subsequently, section IV presents the results attained after simulating the proposed work. The overall study is concluded in section V with future suggestions

II. Review of Existing Works

Some of the existing approaches carried out in case of FER using the DL methods and approaches are discussed in the respective section.

As FER plays a major role in cognitive psychology research, measuring these emotions remains a challenging task. These are data transit from the lab-controlled environments, although DL approaches are employed, over fitting due to inadequate availability of the training data and complications which are associated with the expression and identity bias [12]. Thus accuracy of the expression recognition is required in extracting the relevant facial expression used in the accurate prediction of the emotion recognition. Due to few facial features and their intensity for the DL models, the improvement in the performance of the FER still remains challenging. Thus an accurate prediction of the facial emotion model has to be validated [13]. The dynamic conditions of the FER are the vital characteristic hindering these models affecting the general affect recognition. This can be reduced using the enhanced learning facial features in the last layer of the network and by reducing the levels of trainable parameters. But these process require more of data processing capabilities in accurate rates of FER [14]. In order to suppress the difficulty, by annotating the large-scale dataset can be met using self-Attention Mechanism, with the careful labelling technique, but the RAF_DB, Synthetic forms of FER data suppressed the approach [15].

Some of the same DL approaches are used in case of video sequences and in online learning platforms. But the DL model using the DNN with the transfer learning approach are made as a hybrid model which involves more of computational cost and requires more of time in predicting the accurate FER [16]. Moreover, a FER based on DL approaches can be used as a passive factor for a two-factor authentication process [17]. Similarly, ensemble model are used in the real time FER that are used in the extraction of the facial regions in real time and are used in the compression of the data using multiple features. This results in insufficient data and in the expression which are related to the intra-class differences [18]. At certain points, conditional GAN are used in the data augmentation process, but employing only three layers of the sub-network failed in making a contemporary DNN which are trained on larger dataset [19]. Similarly, a FER should be examined on the basis of detect the FER, either from the image or in a video content. Human emotion interaction and emotional expression play a viral role in enhancing the mode of interaction with one another, when they are applied to AI recognitions especially in the healthcare and in medical fields [20].

In view of establishing these tasks, efficient feature extraction and data pre-processing techniques are used in retrieval of suitable data from the FER can result in accurate and in more of subsequent FER [21]. As a part of DL theory, not only works on the basis of the depth of the model but also emphasizes the significance of the FL of the network model [22]. In view of these characteristics the existing study used ABASNet algorithm, with H-softmax method for optimizing the threshold levels in the NN, which can be used either in embedded services and in other classification tasks achieving higher applied values on practical significance with NN [23]. Real-time emotion recognition in children with autism which are detected using the CNN and the LSTM classifier by the collection of the EEG signals, and are cross-validated using five-cross fold validation for both the LSTM and in the CNN classifiers [24]. But these methods suffered with certain limitations in the FER such as light intensity, position of face, and backgrounds. These EEG signals are pretentious with noise and other artifacts. Procuring this the multi-model systems used in the FER, are increasing at exponential levels, however combining several face expressions with the modalities such as signals of speech, and bio signals are not effective in detecting the precise emotions [25].

Many such techniques have been suggested in enhancing the accuracy of FER from facial cues. But these models require enormous amounts of computational power to train and to process the emotional recognition mechanism at faster rates [26]. Whereas, with the increase in security issues universally, micro-expression which are characterized in a short-duration with low intensity have resulted in lower range of performance [27]. Considering this, ML have been used as a primitive approach in various classification problems but only few were carried out in this era, comprising SVM but are comprised with complexity and are slower in rates of learning [28]. Hybrid forms of DCNN-SVM are used in the video FER, for analyzing the sleepawake pattern [29]. Additionally, face and facial parts using Viola-John algorithm with the KNN classifier, are used in the face tracking in real time scenarios. These methods are limited to detect more of features at one single time [30]. Some of the existing approaches using the DCNN models such as DenseNet-161 [31], ResNet-34, ResNet -18, VGG19 and VGG-16 along with JAFFE and KDEF which are used on the enhanced FER by adapting 10 fold cross validation approach [32].Additionally, the emotion like happy,sad,anger and so on are identified by the models like RF,SVM and XGBoost has been showed a better accuracy [33]. Moreover, the XGBoost has been utilized to select the features from FER2013, CK+ and attain 66.1%,98.32% accuracy [34].

2.1 Literature Gap

Some of the existing approaches have endured with limitations which led to further implications than the existing studies.

- CNN considers only few layers though the model is deeper which results in less accuracy rates of FER [33].
- Backward propagation could not enhance the classification process of the image. Additionally this drawback reduces the accuracy rate of the image prediction and classification rates [34].

IV. Proposed methodology

The proposed method uses the FER13 dataset, for Face Expression Recognition. This is carried out in the perception of analyzing and in detecting the various emotions expressed via facial expressions. With an intention in resolving the existing drawbacks endured with less accuracy rates, redundant data levels and low convergence rates, the aim of the proposed technique lies in case of predicting the appropriate expression delivered using the facial expression using the DL techniques by adapting the Deep multi-level feature extraction technique using ResNet50 and the classification of the emotions delivered using Weight normalized XG-Boost algorithm. Initially, the images from the dataset are obtained using proper image pre-processing techniques which are suitable for the further feature extracting approaches. These pre-processing techniques enhance the performance of the face expression recognition levels. This image pre-processing techniques in the proposed methodology indulges the process of Image resizing which can bring up various advantages to the input images such as enhancing the image clarity, scaling levels, adjusting the contrast levels and other associated enhancement process by reducing the noise levels in enhancing the expression frames. The complete overview via flow representation is provided in figure 1.

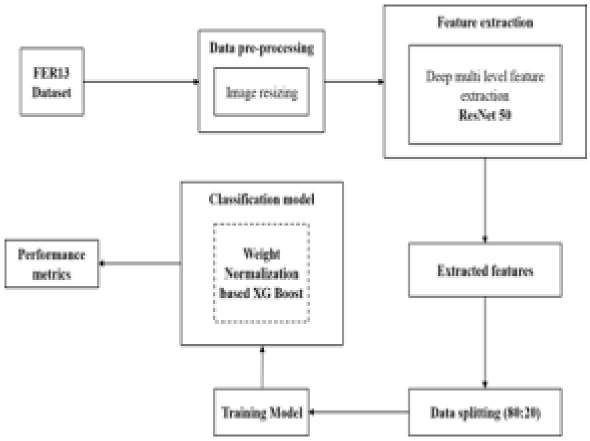

Figure 1 Overall view of the proposed study

This is followed by the feature extraction technique, where the proposed study uses the multi-level feature extraction using the deep multi-level Feature extraction which are opted for a more relevant and optimal feature extraction mechanism by reducing the other redundant data additionally bringing advantages such as by preventing over fitting issues and in bringing optimal outcomes. .Moreover, the algorithms such as RF, SVM and ANN gives the good results. Whereas, combination of XGBoost provides scalability, in-build regularization methods, which aids in preventing over-fitting and flexibility makes it effective. It allows for the modification of the loss function and supports parallel and distributed calculating that makes it scalable and suitable for the computational efficiency. Hence, in the proposed study, Weight normalized XG-Boost mechanism are used under specific activity for the efficient classification techniques, used in aspects of bringing advantages in stabilizing the learning phase of the model making accurate classification levels. These are additionally endured with the advantages of normalizing the individual features. This procedure is followed by efficient train test data split which is split in the range of 80:20 respectively. These train data are taken for the process of classification which are done using the weight normalized XG-Boost model. These are then analyzed using the performance metrics, comprising the accuracy levels, precision rates, F1-score and the recall rates.

3.1 Deep multi-level Feature extraction using ResNet 50

ResNet 50, is one of an optimal model used in the accurate extraction of the deep feature. These are associated with the input data using the DL approach. This model is applied along the layer which is upgraded in the architecture in order to keep the optimal computational efficiency. ResNet 50 is used in the process of feature extraction, vitally used in case of avoiding the over fitting problems, decrease the loss values and also in reducing the problems of fluctuation. The ResNet50, adapts the process of adaptive learning methods than a fixed learning rate method in aim of adjusting the weight of the layers. This ResNet is a type of DL network which mimics the biological network and the pyramidal cells in the human cortex, this is iteratively used in fields of the computer vision. The ResNet contains 48 coiled layers, 1 maxpool, intermediate pool layer and ReLU activation method. These are used in case of feature extraction as they overcome over fitting issues where the front-feed NN contains a hidden layer, which consists several specific neurons which mainly works in the aspect of reducing the errors during the training of data using $y1 = h(x1) + f(xi, wi)$ and $Xi + 1 = f(y1)$. Where $x1$ is the input feature, Wi is the set of weights f is the residual function and h is the set function used as an identity in mapping the features which are done using, $xL - xi + \sum_{i-1}^{L-1} F(xi, Wi)$. Whereas in the proposed study, the multi-level feature extraction is used in view of extracting the multiple features from each of the layers which assisted in enhanced feature extraction mechanism. During CNN training, for a specific feature extraction process, the kernels are randomly generated and each of the kernel produces a different feature. The complete pseudo code of the proposed ResNet 50 is provided in Pseudo code I.

3.2 Classification using Weight normalised XG-Boost algorithm

XG-boost is considered to be an optimal and a well performing algorithm in various cases, where the notion of the stems to the construction of the additive models. Here the K being in values of 1,2…M where M represents the number of base learners. The S_k calculated using $s_k = \frac{\partial L(y,f)}{\partial f}$ and the V_k is calculated using $v_k = \frac{\partial^2 L(y,f)}{\partial f^2}$

Finally, the binary splits with a maximum gain of approach is calculated using equation (1)

$$A = \frac{1}{2}\left[\frac{S_{Left}^2}{V_{Left}} + \frac{S_{Right}^2}{V_{Right}} - \frac{S^2}{V}\right] \quad (1)$$

Whereas the weight of the leaf is determined using $w^* = -\frac{S}{V}$ and the base learner is determined using equation 2

$$\hat{b}(x) = \sum_{j=1}^{T} wI \quad (2)$$

Further addition of trees are done using equation (3)

$$f_k(x) = f_{k-1}(x) + \hat{b}(x). \quad (3)$$

The complete pseudo code of the Weight normalised XG-Boost algorithm is given in Pseudo code II.

<u>Pseudo code-II- Weight normalised XG-Boost algorithm</u>
Step 1: Initialize the ensemble of weak classifiers. **Step 2**: For each boosting iteration: a. Calculate the gradients for each training instance: For each training instance, compute the gradient of the loss function with respect to the predicted value. b. Update the weights of the training instances: For each training instance, compute the optimal step size (learning rate) for updating the weight. Update the weight of each training instance based on the learning rate and the gradient. c. Fit a new weak classifier to the updated weights of the training instances. d. Add the new weak classifier to the ensemble. **Step 3**: Repeat the boosting iterations until a stopping criterion is met. **Step 4**: Output the final ensemble of weak classifiers obtained from the weight updating process in XGBoost.

Pseudo code-I- Deep multi-level feature extraction using ResNet 50
Step 1: Initialize the input image and network weights with the desired input shape. **Step 2**: Pass the input image through the first convolutional layer to extract features. **Step 3**: Use a series of residual blocks to process the output of the convolutional layer. **Step 4**: Each residual block applies convolutional layers and adds the input to the output. **Step 5**: The final output of the residual blocks is the output of the last residual block. **Step 6**: Add a fully connected layer to the final output for classification. **Step 7**: Define the feature extraction using global average pooling and weights multiplication. **Step 8**: Optimize the model using techniques like batch normalization and dropout.

III. 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.

3.1 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 centers 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. The training data consists of 28, 709 sample image data and the test data consists of about 3, 589 image data.

3.2 Exploratory Data Analysis

The EDA is used in case of legalizing the data by adapting various methodologies for visualization. EDA is utilized in defining the patterns to legalize the expectations by implementing a graphical representations and statistical summaries. Moreover, EDA offers data which helps in understanding the dataset.

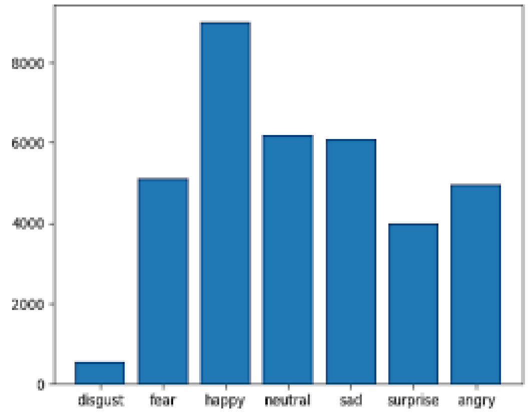

Figure 2 image counts of the dataset

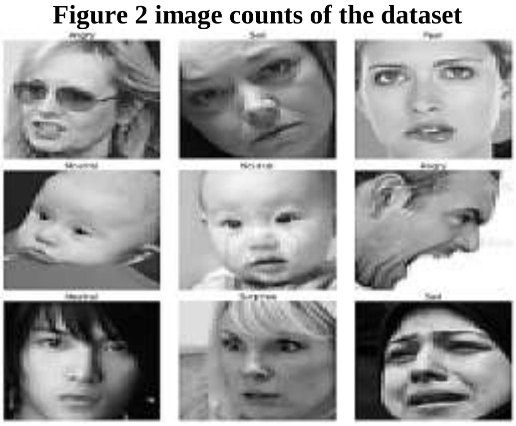

Figure 3 Various Facial Expression (FE) used

Figure 2 represents the number of varying image counts used in the proposed study from the FER13 dataset. These are under various expressions such as fear, happy, neutral, sad and few. Whereas the Figure 3, 4 and 5 are the various emotions expressed in face collected using the FER13 dataset used in the analysis of various Facial Expressions using the proposed DL methods.

Figure 4 Various FE used in proposed study

Figure 5 Various FE used in proposed study

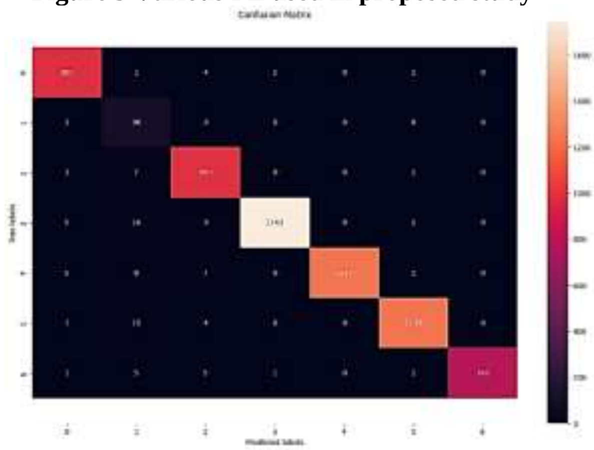

Figure 5 Confusion matrix of the dataset used in study

Figure 5 in the respective section explains the correlation map or the Confusion Matrix (CM) which can be used in correlation of the values present in the dataset. This CM is also used in screening the possible pairs of the values which can be created from the table. CM is considered to be the powerful tool in identifying and in envisaging the patterns in the given data.

3.3 Comparative Analysis

FER13 dataset are adapted and are used in analyzing various Facial Expression and are analyzed 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. 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 Comparative analysis among proposed and existing methodology in FER in terms of accuracy rates

MODEL	Accuracy
DCNN Model1	72
DCNN Model2	72.02
EmNet (average fusion)	74.11
EmNet (weighted maximum fusion)	74.06
Proposed	99.19

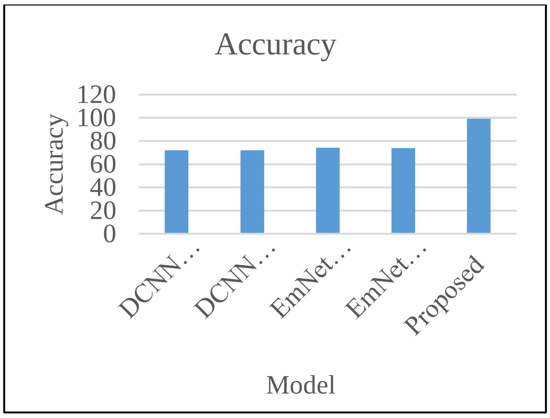

Figure 6 Comparative analysis among proposed and existing methodology for FER in terms of accuracy rates

The table 1 and figure 6 represented in the section deliberates that the proposed model have outperformed the existing approaches carried out in terms of FER, using the DL methods such as DCNN (model 1 and 2), EmNET, with average and weighted Max fusion techniques used in FER. These models have achieved accuracy rates of 72, 72.02, 74.1, 74.06 and the proposed model have achieved the accuracy rate more than in a range of 25.08 resulting in 99.19% in terms of the accuracy rate of the proposed model.

Table 2 Comparative analysis among proposed and existing methodology in FER.

MODEL	Accuracy
Existing model	65.97
Proposed	**99.19**

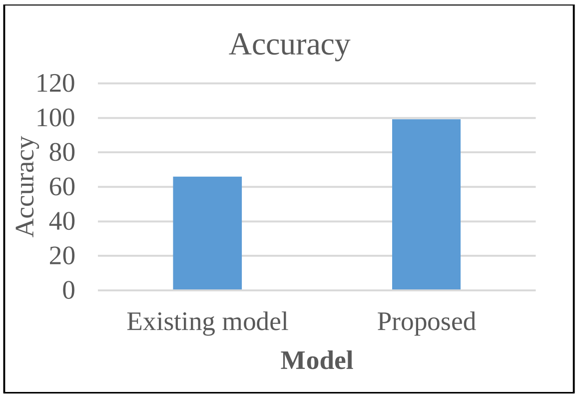

Figure 7 Comparative analysis among proposed and existing methodology for FER in terms of accuracy rates

Figure 7 and the table 2 clearly depicts that the proposed model have outperformed the existing model in terms of the accuracy rates where the existing model used the CNN as a primitive approach in case of FER, based on Raspberry Pi, and have achieved an accuracy rate of 65.97 whereas the proposed model have used efficient feature extraction and classification techniques and have achieved an accuracy rate of 99.19.

Table 3 Comparative analysis among proposed and existing methodology in FER

MODEL	Train Accuracy	Test Accuracy
FERConvNet_Gaussian	0.98	0.58
FERConvNet_HDM	0.98	0.95
FERConvNet_Nonlocal Means	0.93	0.61
FERConvNet_Bilateral	0.98	0.63
Proposed	0.99	0.99

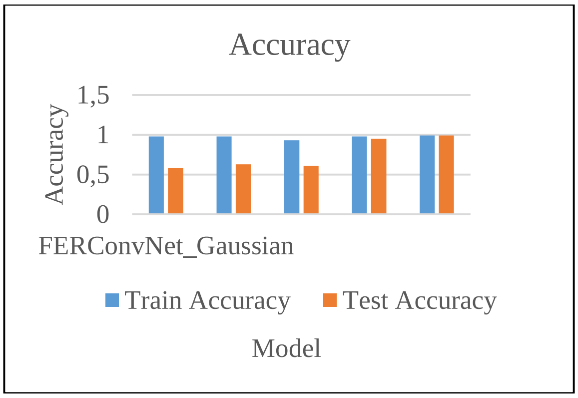

Figure 7 Comparative analysis among proposed and existing methodology for FER in terms of train and test accuracy rates

Figure 7 and table 3 represents the evaluation of both the train and the test accuracy rates which are compared with the existing models comprising FER ConvNet_Gaussian model, FERConvNet_Bilateral, FER ConvNet_Nonlocal means and

FERConvNet_HDM. The existing model have achieved the train and the test accuracy rates at a scale of 0.98, 0.58. Followed by 0.98 and 0.63, 0.93 and 0.61 and finally 0.98 and 0.95 in terms of existing models. Whereas the proposed model have achieved the train and the test accuracy rates in range of 0.99 and 0.99 on both the cases.

Table 4 Comparative analysis among proposed and existing methodology in FER

MODEL	Accuracy	Precision	Recall	f1_score
Adam	77.17	66.6236	66.8845	66.6779
SGD	76.17	63.0118	61.0729	61.0932
Proposed	99.19	98	98	98

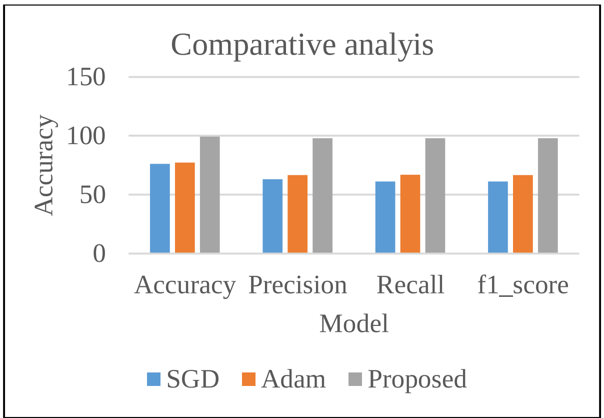

Figure 8 Comparative analysis among proposed and existing methodology for FER in terms of accuracy rates

Table 4 and Figure 8 in the section represents the outperformance of the proposed model compared with the existing techniques comprising SGD known as Stochastic Gradient Descent using the weighted maximum fusion scheme by adapting the FER13 dataset , Adam Optimization Algorithm (AOA). The existing models achieved in both SGD and in the AOA in terms of Accuracy, Precision, Recall and on the basis of F1-score rates are 76.17, 63.08, 61.07 and 61.09 whereas the Adam have achieved them in rates of 77.17, 66.6, 66.8 and 66.6. But the proposed model have outperformed the existing approaches by achieving the accuracy, precision, recall and the F1-score in the rates of 99.1, 98, 98 and 98. Thus, the proposed model using the deep multi-level feature extraction mechanism which are more obvious in enhancing the model accuracy, enhances the speed of leaning mechanism, reduces all of the redundant data and are more peculiar in producing accurate levels of the data for further processing techniques. Whereas on the other hand, the weight normalized methods adapted for the classification technique, which are more advantageous in making high class of classification with better accuracy terms in a shorter period of time and with faster convergence rates. Weight normalization process also results in the enhanced speed in making the values learn and in making a stabilized learning process, Adapting these techniques in the proposed study have resulted in enhancing the attributes of the performance metrics such as accuracy, precision rates, recall and the F1-score ranges.

V. Conclusion

The research aimed to recognize facial emotions with deep feature extraction and machine learning classification. For performing this, ResNet-50 was used for extracting significant features, while, WN-XG-Boost algorithm was employed for classification. A confusion matrix was also attained through the execution phase, wherein, the misclassification rates of the proposed system was less than correct classification rate. This reveals the effectual performance of the proposed system. However, to expose the better performance of the proposed system than conventional works, Comparisons were undertaken with four traditional systems with regard to the four metrics (accuracy, recall, F1-score and precision). From the comparative results, it were unveiled that, the proposed method explored outstanding performance than conventional works with 99.19% accuracy, 98% recall, 98% f1-score and 98% precision. Future of the current work hinges on incorporating numerous features and advanced DL models.

REFERENCES

[1] B. Catanzariti, "Seeing affect: knowledge infrastructures in facial expression recognition systems," 2023.

[2] G. Bargshady, X. Zhou, R. C. Deo, J. Soar, F. Whittaker, and H. Wang, "Enhanced deep learning algorithm development to detect pain intensity from facial expression images," *Expert Systems with Applications*, vol. 149, p. 113305, 2020.

[3] J. Chen, Y. Lv, R. Xu, and C. Xu, "Automatic social signal analysis: Facial expression recognition using difference convolution neural network," *Journal of Parallel and Distributed Computing*, vol. 131, pp. 97-102, 2019.

[4] S. Zhang, X. Pan, Y. Cui, X. Zhao, and L. Liu, "Learning affective video features for facial expression recognition via hybrid deep learning," *IEEE Access*, vol. 7, pp. 32297-32304, 2019.

[5] A. Khattak, M. Z. Asghar, M. Ali, and U. Batool, "An efficient deep learning

technique for facial emotion recognition," *Multimedia Tools and Applications,* pp. 1-35, 2022.
[6] M. Aamir, Z. Rahman, W. A. Abro, U. A. Bhatti, Z. A. Dayo, and M. Ishfaq, "Brain tumor classification utilizing deep features derived from high-quality regions in MRI images," *Biomedical Signal Processing and Control,* vol. 85, p. 104988, 2023.
[7] N. Andresen, M. Wöllhaf, K. Hohlbaum, L. Lewejohann, O. Hellwich, C. Thöne-Reineke, *et al.*, "Towards a fully automated surveillance of well-being status in laboratory mice using deep learning: Starting with facial expression analysis," *PLoS One,* vol. 15, p. e0228059, 2020.
[8] M. Flynn, D. Effraimidis, A. Angelopoulou, E. Kapetanios, D. Williams, J. Hemanth, *et al.*, "Assessing the effectiveness of automated emotion recognition in adults and children for clinical investigation," *Frontiers in human neuroscience,* vol. 14, p. 70, 2020.
[9] C. Pabba and P. Kumar, "An intelligent system for monitoring students' engagement in large classroom teaching through facial expression recognition," *Expert Systems,* vol. 39, p. e12839, 2022.
[10] N. Andresen, M. Wöllhaf, K. Hohlbaum, L. Lewejohann, O. Hellwich, C. Thöne-Reineke, *et al.*, "Towards a fully automated surveillance of well-being status in laboratory mice using deep learning," *BioRxiv,* p. 582817, 2019.
[11] Y. Xie, W. Tian, H. Zhang, and T. Ma, "Facial expression recognition through multi-level features extraction and fusion," *Soft Computing,* pp. 1-16, 2023.
[12] M. Karnati, A. Seal, D. Bhattacharjee, A. Yazidi, and O. Krejcar, "Understanding deep learning techniques for recognition of human emotions using facial expressions: a comprehensive survey," *IEEE Transactions on Instrumentation and Measurement,* 2023.
[13] T. Shahzad, K. Iqbal, M. A. Khan, and N. Iqbal, "Role of zoning in facial expression using deep learning," *IEEE Access,* vol. 11, pp. 16493-16508, 2023.
[14] P. Barros, N. Churamani, and A. Sciutti, "The facechannel: a fast and furious deep neural network for facial expression recognition," *SN Computer Science,* vol. 1, pp. 1-10, 2020.
[15] Z. Zhao, Q. Liu, and S. Wang, "Learning deep global multi-scale and local attention features for facial expression recognition in the wild," *IEEE Transactions on Image Processing,* vol. 30, pp. 6544-6556, 2021.
[16] R. K. Mishra, S. Urolagin, J. A. A. Jothi, and P. Gaur, "Deep hybrid learning for facial expression binary classifications and predictions," *Image and Vision Computing,* vol. 128, p. 104573, 2022.
[17] C. Benegui and R. T. Ionescu, "Convolutional neural networks for user identification based on motion sensors represented as images," *IEEE Access,* vol. 8, pp. 61255-61266, 2020.
[18] L. Wang, Z. He, B. Meng, K. Liu, Q. Dou, and X. Yang, "Two-pathway attention network for real-time facial expression recognition," *Journal of Real-Time Image Processing,* vol. 18, pp. 1173-1182, 2021.
[19] A. H. Farzaneh and X. Qi, "Facial Expression Recognition in the Wild via Deep Attentive Center Loss."
[20] M. Egger, M. Ley, and S. Hanke, "Emotion recognition from physiological signal analysis: A review," *Electronic Notes in Theoretical Computer Science,* vol. 343, pp. 35-55, 2019.
[21] A. T. Wasi, K. Šerbetar, R. Islam, T. H. Rafi, and D.-K. Chae, "ARBEx: Attentive feature extraction with reliability balancing for robust facial expression learning," *arXiv preprint arXiv:2305.01486,* 2023.
[22] R. Banerjee, S. De, and S. Dey, "A survey on various Deep Learning algorithms for an efficient facial expression recognition system," *International Journal of Image and Graphics,* vol. 23, p. 2240005, 2023.
[23] F. Wang, F. Xie, S. Shen, L. Huang, R. Sun, and J. Le Yang, "A novel multiface recognition method with short training time and lightweight based on abasnet and h-softmax," *IEEE Access,* vol. 8, pp. 175370-175384, 2020.
[24] A. Hassouneh, A. Mutawa, and M. Murugappan, "Development of a real-time emotion recognition system using facial expressions and EEG based on machine learning and deep neural network methods," *Informatics in Medicine Unlocked,* vol. 20, p. 100372, 2020.
[25] J. Zhang, Z. Yin, P. Chen, and S. Nichele, "Emotion recognition using multi-modal data and machine learning techniques: A tutorial and review," *Information Fusion,* vol. 59, pp. 103-126, 2020.
[26] A. Chowanda, "Separable convolutional neural networks for facial expressions recognition," *Journal of Big Data,* vol. 8, p. 132, 2021.
[27] I. P. Adegun and H. B. Vadapalli, "Facial micro-expression recognition: A machine learning approach," *Scientific African,* vol. 8, p. e00465, 2020.

[28] H. I. Dino and M. B. Abdulrazzaq, "Facial Expression Classification Based on SVM, KNN and MLP Classifiers."

[29] M. Awais, X. Long, B. Yin, S. F. Abbasi, S. Akbarzadeh, C. Lu, *et al.*, "A hybrid DCNN-SVM model for classifying neonatal sleep and wake states based on facial expressions in video," *IEEE Journal of Biomedical and Health Informatics,* vol. 25, pp. 1441-1449, 2021.

[30] K. S. Yadav and J. Singha, "Facial expression recognition using modified Viola-John's algorithm and KNN classifier," *Multimedia Tools and Applications,* vol. 79, pp. 13089-13107, 2020.

[31] J. M. Song, W. Kim, and K. R. Park, "Finger-vein recognition based on deep DenseNet using composite image," *Ieee Access,* vol. 7, pp. 66845-66863, 2019.

[32] M. Akhand, S. Roy, N. Siddique, M. A. S. Kamal, and T. Shimamura, "Facial emotion recognition using transfer learning in the deep CNN," *Electronics,* vol. 10, p. 1036, 2021.

[33] A. Khan, A. Sohail, U. Zahoora, and A. S. Qureshi, "A survey of the recent architectures of deep convolutional neural networks," *Artificial intelligence review,* vol. 53, pp. 5455-5516, 2020.

[34] S. Liu, "Image Classification of Static Facial Expressions in the Wild Based on Bidirectional Neural Networks (Based on Pytorch)."