INDIAN JOURNAL OF SCIENCE AND TECHNOLOGY

RESEARCH ARTICLE

Facial expression recognition for low resolution images using convolutional neural networks and denoising techniques

Pavan Nageswar Reddy Bodavarapu[1]*, **P V V S Srinivas**[2]

1 Student, CSE Department, Koneru Lakshmaiah Education Foundation, 534406, India. Tel.: +91-9182811847
2 Assistant Professor, CSE Department, Koneru Lakshmaiah Education Foundation, 502355, India

OPEN ACCESS

Received: 05.01.2021
Accepted: 21.03.2021
Published: 13.04.2021

Citation: Bodavarapu PNR, Srinivas PVVS (2021) Facial expression recognition for low resolution images using convolutional neural networks and denoising techniques. Indian Journal of Science and Technology 14(12): 971-983. https://doi.org/10.17485/IJST/v14i12.14

* **Corresponding author**.

Tel: +91-9182811847
170030144@kluniversity.in

Funding: None

Competing Interests: None

Published By Indian Society for Education and Environment (iSee)

ISSN
Print: 0974-6846
Electronic: 0974-5645

Abstract

Background/Objectives: There is only limited research work is going on in the field of facial expression recognition on low resolution images. Mostly, all the images in the real world will be in low resolution and might also contain noise, so this study is to design a novel convolutional neural network model (FERConvNet), which can perform better on low resolution images. **Methods**: We proposed a model and then compared with state-of-art models on FER2013 dataset. There is no publicly available dataset, which contains low resolution images for facial expression recognition (Anger, Sad, Disgust, Happy, Surprise, Neutral, Fear), so we created a Low Resolution Facial Expression (LRFE) dataset, which contains more than 6000 images of seven types of facial expressions. The existing FER2013 dataset and LRFE dataset were used. These datasets were divided in the ratio 80:20 for training and testing and validation purpose. A HDM is proposed, which is a combination of Gaussian Filter, Bilateral Filter and Non local means denoising Filter. This hybrid denoising method helps us to increase the performance of the convolutional neural network. The proposed model was then compared with VGG16 and VGG19 models. Findings: The experimental results show that the proposed FERConvNet_HDM approach is effective than VGG16 and VGG19 in facial expression recognition on both FER2013 and LRFE dataset. The proposed FERConvNet_HDM approach achieved 85% accuracy on Fer2013 dataset, outperforming the VGG16 and VGG19 models, whose accuracies are 60% and 53% on Fer2013 dataset respectively. The same FERConvNet_HDM approach when applied on LRFE dataset achieved 95% accuracy. After analyzing the results, our FERConvNet_HDM approach performs better than VGG16 and VGG19 on both Fer2013 and LRFE dataset. **Novelty/Applications**: HDM with convolutional neural networks, helps in increasing the performance of convolutional neural networks in Facial expression recognition.

Keywords: Facial expression recognition; facial emotion; convolutional neural network; deep learning; computer vision

1 Introduction

The raw data consists of noise like random variation of brightness or color information, removing noise from the images drastically improves the performance of the facial emotion recognition models. To eliminate noise from images there are many denoising techniques such as gaussian blur, bilateral filter, non-local means filtering. Gaussian Blur helps in blurring the edges and reducing the contrast, but it reduces the details [(1)]. A bilateral Filter decreases the noise by preserving the edges by replacing the intensity of pixels with a weighted average of intensity from surrounding pixels [(2)]. Gaussian Filter, Bilateral Filter and other traditional filtering techniques can remove image noise, but the image structure information is not retained enough. Non Local Means Filtering averages neighbors with similar neighborhoods, with much greater clarity and smaller extent loss of detail post filtering. The limitation of this technique is, efficiency is slightly lower when compared to traditional techniques. The computation complexity is quadratic in number of pixels in the image, so it is expensive to apply. To speed up the execution many techniques were designed, one such technique is the fast Fourier transform, it determines the similarity between two pixels by speeding up the algorithm by a factor of 50 and also maintains the quality of the result [(3)].

Conditional generative adversarial network is one of approach used to reduce the intra-class variations. The proposed approach consists of a generator G and discriminators (Di, Da and Dexp). For learning the generative and discriminative representations, three loss functions were designed. But there is one limitation in this approach is that the model is trained individually for each different datasets, a model which is trained on a particular dataset may result in poor accuracy on another dataset [(4)]. A method based on convolutional neural network and edge detection for facial emotion recognition is designed. For testing this they created a simulation experiment by combining the fer-2013 database with LFW dataset. The average recognition obtained by this method is 88.56% and the train speed on the training dataset is 1.5 times faster than the traditional method [(5)]. Hybrid transfer learning model, which is based on Convolution Restricted Boltzmann Machine (CRBM) model and a Convolutional Neural Network (CNN) model, since there are some content differences between the datasets during traditional transfer learning, which affects the classification performance of the model. In this model CRBM replaces the full connection layer in the CNN model. The added CRBM layer learns about the unique statistical characteristics of the target set. This helps in eliminating the content differences between the datasets [(6)].

Emotion-specific activation maps are constructed to set up infrared thermal facial image sequences as a different approach to finding out the correlation between emotional triggers and changes in facial temperature. Data that is stored in the International Affective Picture System are used to create emotional clips during the testing process. The results show the difficulty of selecting local regions when examining frame temperature had been resolved [(7)]. Simple multi-layer perceptron (MLP) classifier, which can find out the current classification result is reliable or not is created. If the present classification result is classified as unreliable, then that face image is used as a query to search for similar images. Then Facial Action units are used to discover the images with similar face expression. After finding such images another multi-layer perceptron is trained to recognize the final emotion category. This method obtained improved accuracies of 1.03% 1.02% and 1.06% for DenseNet, GoogLeNet and VGG-Face respectively [(8)]. An end-to-end Action Unit-oriented graph classification network is designed, the network extracts the action unit features with the use of 3D ConvNets. For discovering the dependency laying between action unit nodes for micro-expression categorization, graph convolutional network layers are applied. The results show that this approach performs better than convolutional neural networks based on micro-expression recognition [(9)].

Two convolutional neural networks, one for face identification and the second one for expression recognition are used for facial expression recognition. For face recognition two models are used, one is known for low reasoning speed, but very accurate and the other model is known for high reasoning speed, but less accurate [(10)]. Three classifiers (1) baseline classifier with one convolutional layer (2) CNN with five convolutional layers (3) deeper CNN for facial expression recognition. There are three stages in training the model (1) Raw image (2) Normalization (3) CNN train (4) CNN weights [(11)].

Group-based emotion recognition plays a vital role in real world applications, Multivariate Local Texture Pattern, Local energy based shape Histogram and gray-level co-occurrence matrix are used for feature extraction. The proposed model achieved 99.16% accuracy 99.33% recall 99% precision and 99.93% sensitivity. This method achieved 87.8% accuracy on low resolution images [(12)]. The issue with facial expression recognition field is small size dataset. This issue can be solved by joining convolutional neural networks with the augmented dataset. The augmented dataset helps to avoid over-fitting. Horizontal flip, shift, scaling, rotation are the data augmentation methods used to increase the size of dataset. This approach, when combined with the convolutional neural network achieved 99.5% accuracy on the ORL face dataset [(13)].

A 3 Dimensional convolutional neural network is designed for facial expression recognition on videos. The proposed method was carried out by using Tensorflow(deep learning framework) [(14)]. A new method based on hierarchical deep learning, which combines the result of softmax function by considering the error correlated with the second highest emotion recognition result. This model is compared with CK+ and JAFFE dataset. The results show up to 3% of accuracy improvement in CK+ dataset and 7% of accuracy improvement in JAFFE dataset [(15)]. Global average layer is used for avoiding over-fitting for better

facial expression recognition[16]. An automatic emotion recognition method is designed which uses body posture and facial expression information for facial expression recognition. This approach can improve the performance of an facial expression recognition system. A database was developed which contains spontaneous expressions of various children of different ages[17].

There is little research work going in the field of Facial Expression Recognition in low resolution Images. So, in this work we are proposing a novel convolutional neural network and a novel hybrid denoising method for facial expression recognition. The proposed neural network is a simple architecture and this proposed model is compared with state-of-art models on Fer2013 dataset. The batch size employed in this work is 64 and the model is trained for 100 epochs. In order to deal with over fitting, dropout and batch normalization are used.

For recognizing facial expressions from low resolution images, we created a low resolution facial expression (LRFE) dataset, which contains more than 6000 images of seven types of facial expression. FERConvNet, FERConvNet with hybrid denoising method (FERConvNet_HDM), VGG16, VGG19 are tested on this dataset and compared the results.

Our primary contributions in this research paper can be outlined as follows: (1) Novel convolutional neural network model is proposed for facial expression recognition (2) Novel hybrid denoising method is presented (3) We created a low resolution facial expression (LRFE) dataset for facial expression recognition in low resolution images.

Table 1. Outline of existing face databases and LRFE dataset

Category	JAFFE	MMI	FER2013	LRFE
Static images	219	740	35887	6100
Downloadable	Yes	Yes	Yes	No
No.of emotion expression	7	7	7	7
Gender	Female	Female/Male	Female/Male	Female/Male
Glasses	No	Yes	Yes	Yes

2 Proposed Methodology

2.1 Filter description

The main aim of our research is to compare the proposed convolutional neural network with state-of-art models. Filtering techniques like Gaussian, Bilateral, Non local Means are applied to the images to remove any unwanted noise from the images, because having any noise in images can decrease the performance of convolutional neural network. A hybrid denoising method is proposed by combining the Gaussian, Bilateral, Non local means denoising techniques. Gaussian filter is a 2D convolution filter, which blur the image, helping in removal of noise. The only limitation with this technique is, the loss of image details is high when compared to other techniques. Bilateral is a non-linear filtering technique used to remove noise from the image by preserving the edges. The limitation of this technique is that it introduces false edges in the image. Non local means filter, unlike taking the mean value of a group of pixels, Non local means takes a mean of all pixels and unlike other techniques which blur the image, Non local means can restore the texture of image.

Equations used for each filtering techniques are given below,

Gaussian Filtering,

$$img\,(a,b) = 1/2\pi\sigma^2 \ * \ e^{\frac{a^2+b^2}{2\sigma^2}} \tag{1}$$

Non Local Means Filtering,

$$img_{filtered}\,(x) = \frac{1}{N(x)} \int v\,(y)\,f(x,y)dy$$

Normalizing factor N(x) is given by,

$$N\,(x) = \int f(x,y)dy \tag{2}$$

2.2 Dataset description

The existing Fer2013 dataset contains 35887 images of facial expressions belonging to seven expressions (Happy, Disgust, Fear, Sad, Neutral, Angry, Surprise). This dataset contains 4593 angry images, 547 disgust images, 5121 fear images, 8989 happy images, 6077 sad images, 4002 surprise images, 6198 neutral images. All these images are grayscale and 48X48 sized. We created LRFE dataset by collecting images from various sources, nearly 6000 images are collected, which belong to seven categories of facial expression. Since the raw images collected are having different file extension formats (.JPG, .PNG, .GIF), we converted all these into .JPG format. Since convolutional neural networks require large samples of training images, we used three image appearance filters and four affine transform matrices. The three filters are average, Gaussian, Bilateral Filters. Therefore, the number of samples in LRFE dataset is 35000, which belong to seven facial expressions. All these images are then converted into grayscale and then resized to 48X48 pixels. We now divided this dataset into a training set and testing set in the ratio 80:20.

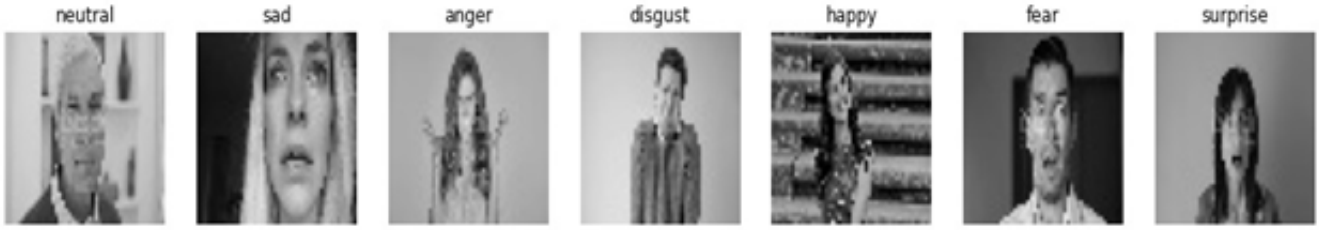

Fig 1. Sample images in LRFE dataset

2.3 Model description

A novel convolutional neural network is proposed for facial expression recognition and compared it with state-of-art models. Various Filtering techniques like Bilateral filter, Gaussian filter, Nonlocal means denoising are applied to all images to remove the noise from the images. A Hybrid denoising method is designed by combining the Gaussian, bilateral, non-local means denoising techniques. For dividing the data into train and test sets, we used 80:20 ratio. Then the model is trained on the train set and evaluated with the test set and the performance metrics are displayed.

Step 1: Firstly, all the images are selected, and Bilateral Filtering is applied and the resulted images are stored in bilateral image dataset.
Step 2: The same process is repeated with Gaussian Filter, Non-Local Means denoising techniques and the resulted images are placed in Gaussian and Non-Local Means image datasets, respectively.
Step 3: Then, n random images are selected from the image dataset with help of randomSelect function designed in this algorithm.
Step 4: Bilateral Filtering is applied on these randomly selected images, the resulting images are placed in Hybrid image dataset and these randomly selected images are taken out from the initial image dataset.
Step 5: The same process is repeated for Gaussian, Non-Local Means de-noising techniques and finally, a Hybrid image dataset is formed, which contains images that belong to various filtering techniques.
Step 6: Assign the labels to the Hybrid image dataset.
Step 7: Divide Bilateral, Gaussian, Non-Local Means denoising and Hybrid images datasets into training set and testing set in the ratio 80:20.
Step 8: Train the Novel proposed model with the train set, then the model is evaluated with test set.

2.4 Algorithm

Input : *ImageDataset containing seven different Facial Emotion images.*
Output : *HybridImageDataset after applying different denoising filtering techniques.*
$HybridAlgo\,(ImageDataset[\,]), HybridImageDataset[\,] = \{\phi\}$
Begin
  if $size\,(ImageDataset[\,] \neq \varnothing)$ *then*
    Begin
      for all the images in ImageDataset
      $x = size(ImageDataset[\,])$ //Resizes all images into size of 48X48
      $while(x > 0)$
        Beginloop
          $img_j[\,] = randomSelect(ImageDataset[],x)$
          $img_{fb}[\,] = BilateralFilter(img_j(\,])$
          $HybridImageDataset[\,] = HybridImageDataset[\,] \cup img_{fb}[\,]$
        $ImageDataset = ImageDataset - img_j[\,]$
        $x = x - size(img_j[\,])$
        $img_k[\,] = randomSelect(ImageDataset[],x)$
        $img_{fg}[\,] = GaussianFilter(img_k[])$
          $HybridImageDataset[\,] = HybridImageDataset[\,] \cup img_{fg}[\,]$
          $ImageDataset = ImageDataset - img_k[\,]$
        $x = x - size(img_k[\,])$
        $img_l[\,] = randomSelect(ImageDataset[],x)$
        $img_{fn}[\,] = NonLocalMeans(img_l[])$
          $HybridImageDataset[\,] = HybridImageDataset[\,] \cup img_{fn}[\,]$
          $ImageDataset = ImageDataset - img_l[\,]$
        $x = x - size(img_l[\,])$
        Endloop
      End
    Endif
  End

Input : *Img Dataset consists images of Seven different Facial Emotions after applying*
Filtering techniques. Gaussian, Bilateral, Non Local Means, Hybrid Denoised Images are given for Evaluation
Output : *Classification of Image base on Emotion*
  $fernet\,(InputImgDataset)$
  Begin
    $x = size(InputImgDataset)$
    $for\ j\ in\ 1\ to\ x$
    {
    $ImageLabelSet \leftarrow Label(InputImgDataset\,(img_j])$
    $TrainSt, TestSt \leftarrow Split\,(ImageLabelSet, 80, 20)$
    $FERNET\ Model \leftarrow FERNET\ Model\,(TrainSt)$
    $Evaluate \leftarrow FERNET\ Model(TestSt)$
    }
    return Emotion
  end

2.5 FERConvNet Architecture

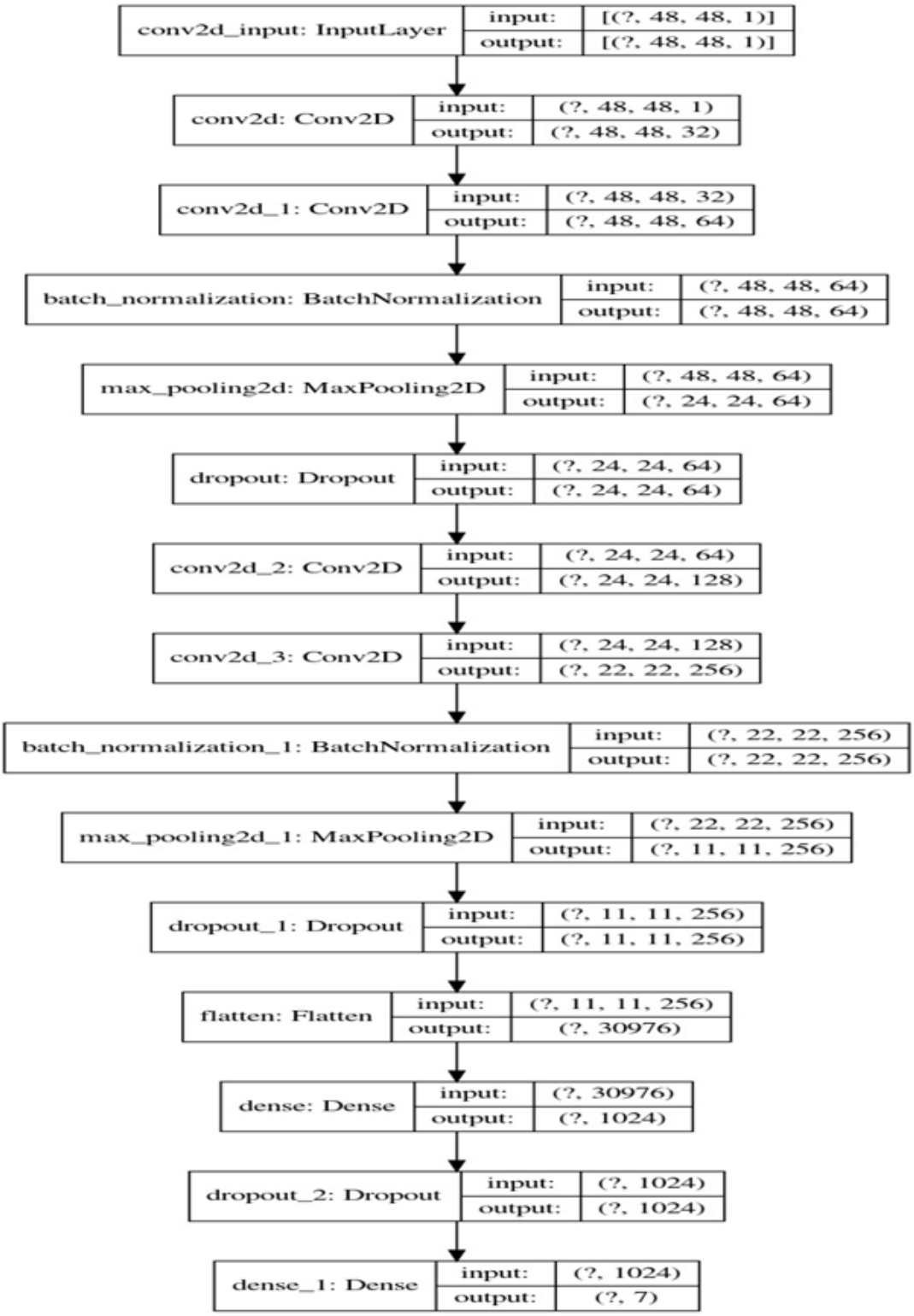

Fig 2. Proposed model (FERConvNet) Architecture

3 Results and Discussion

3.1 Dataset and Execution

The Execution of the model is done on kaggle platform, which provides a single NVIDIA Tesla P100, TPU V3-8, 9 hours execution time, 20 Gigabytes of disk space. The GPU specifications are 2 CPU cores and 13 Gigabytes of RAM. For implementing this model we used LRFE dataset and Fer2013 dataset. The Fer2013 contains 35887 images, which are divided into train and test sets in the ratio 80:20. The train set and test set of Fer2013 dataset contain 28709 and 7178 images respectively. LRFE dataset contains 6000 images of facial expression belonging to seven emotions (Happy, Sad, Surprise, Neutral, Fear, Disgust, Angry), which are collected from various sources.

3.2 Performance on Fer2013 dataset

Fer2013 dataset contains 35887 images, each image labeled as one of the seven emotions. All the images are in grayscale format and 48X48 pixels. Both posed and unposed images are present in Fer2013 dataset. This dataset contains 4953 angry images, 547 disgust images, 5121 fear images, 8989 happy images, 6077 sad images, 4002 surprise images, 6198 neutral images.

Table 2. Outline of number of samples in each expression of Fer2013 dataset

Dataset	Name & No. of images in each emotion						
	Happy	Sad	Angry	Disgust	Sad	Surprise	Neutral
Fer2013	8989	6077	4953	547	6077	4002	6198

Fig 3. Sample images of each expression in Fer2013 dataset

We now present the results on Fer2013 dataset, where the Fer2013 dataset is divided in the ratio 80:20 for training and testing, validation. The batch size used is 64 and trained for 100 epochs.

Table 3. Accuracy and Loss Comparisons of Different Models on Fer2013 dataset

S.no	Model Name	Dataset	Train Accuracy	Test Accuracy	Train loss	Test loss
1	VGG16	Fer2013	0.63	0.60	1.01	1.10
2	VGG19	Fer2013	0.54	0.53	1.22	1.20
3	FERConvNet	Fer2013	0.79	0.65	0.69	1.07
4	EfficientNetB7	Fer2013	0.63	0.60	1.10	1.09

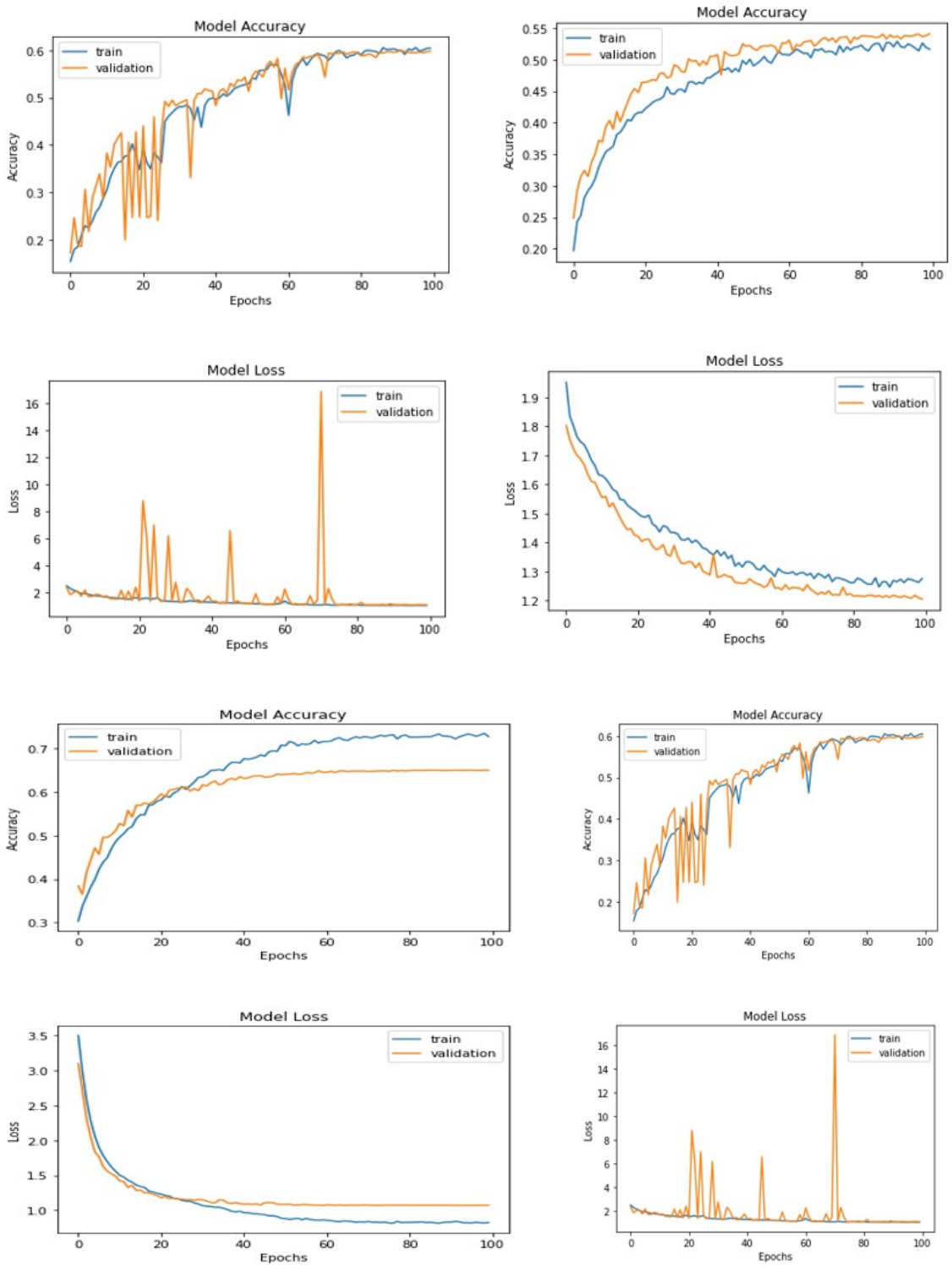

Fig 4. a. Accuracy & Loss Comparison of VGG 16 Model on FER 2013, b. Accuracy & Loss Comparison of VGG 19 Model on FER 2013, c. Accuracy & Loss Comparison of Proposed FERConvNet Model on FER 2013, d. Accuracy & Loss Comparison of EfficientNetB7 model on Fer2013

Table 4. Accuracy and Loss Comparison of Proposed FERConvNet Model on FER2013 datasetafter Applying Various Denoising Techniques

S.no	Model Name	Dataset	Train Accuracy	Test Accuracy	Train loss	Test loss
1	FERConvNet_Gaussian	Fer2013	0.65	0.55	1.05	1.33
2	FERConvNet_Bilateral	Fer2013	0.80	0.65	0.69	1.09
3	FERConvNet_Nonlocal Means	Fer2013	0.79	0.65	0.71	1.09
4	FERConvNet_HDM	Fer2013	0.87	0.85	0.47	0.56

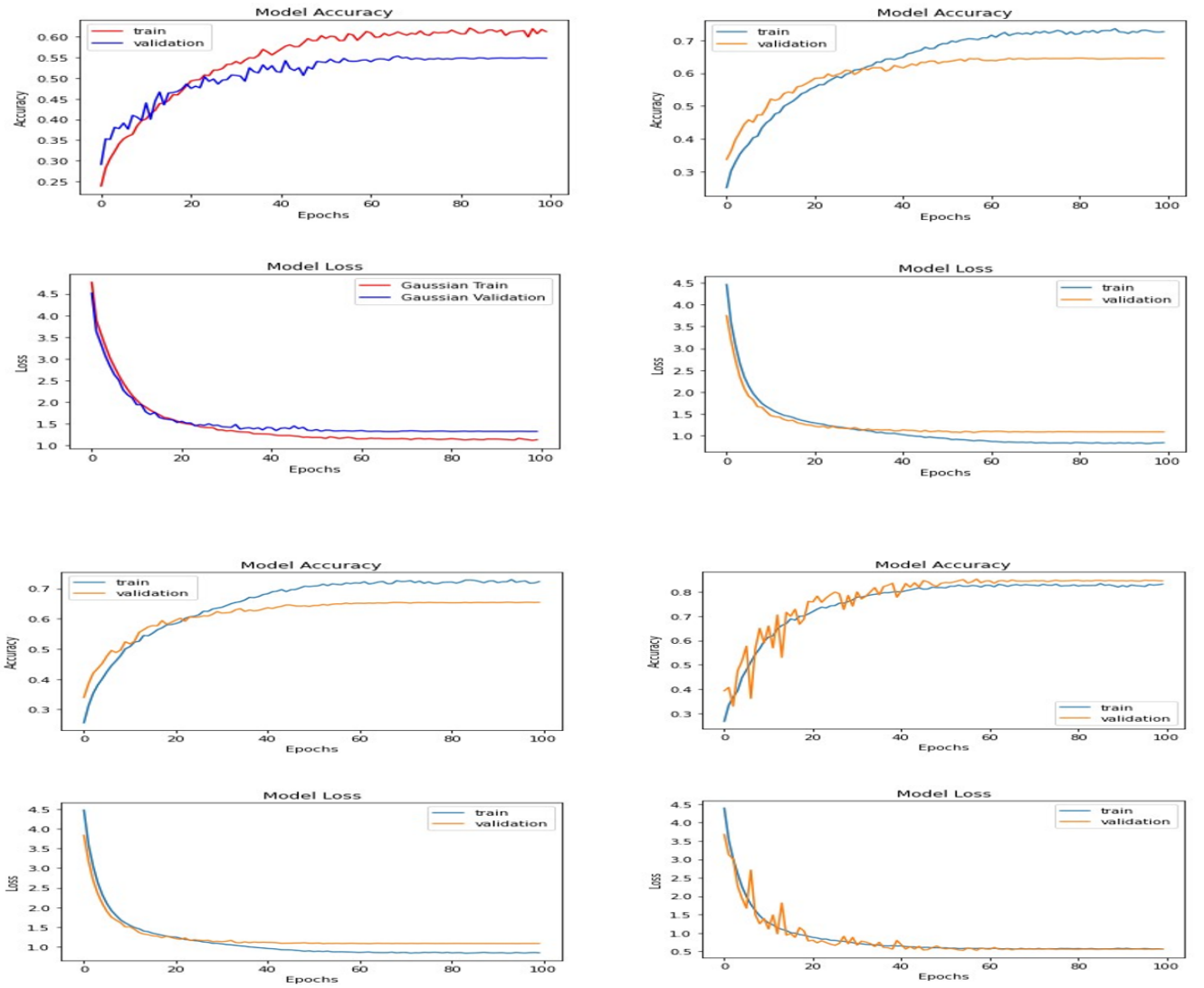

Fig 5. a. Accuracy & Loss Comparison of ProposedFERConvNet Model on FER 2013 with Gaussian Denoising, b. Accuracy & Loss Comparison of Proposed FERConvNet Model on FER 2013 with Bilateral Denoising, c. Accuracy & Loss Comparison of Proposed FERConvNet Model on FER 2013 with Nonlocal Means Denoising, d. Accuracy & Loss Comparison of Proposed FERConvNet Model on FER 2013 with Hybrid Denoising

Tables 3 and 4 indicate the accuracy and loss of various deep learning models on the Fer2013 dataset. The proposed model (FERConvNet) achieved 79% accuracy on training set and 65% accuracy on testing set. After analyzing the results, our proposed model (FERConvNet) performs better than VGG16 and VGG19, which are state-of-art models in deep learning. The results show that FERConvNet has better accuracy on both training and testing set, when compared to VGG16 and VGG19 models.

Although FERConvNet has lesser layers than VGG16 and VGG19, it performed better than the VGG16 and VGG19. The test accuracy of VGG16 on Fer2013 dataset is 60% and VGG19 on Fer2013 dataset is 53%. The latest deep learning model efficientNetB7 is also implemented on FER2013 dataset, the training accuracy is 63% and test accuracy is 60%. This model seems like, it is not effective in extracting features from the facial images, which resulted in poor train accuracy. Here also clearly the FERConvNet performs better than the efficientNetB7 in terms of train and test accuracies.

We then applied filtering techniques like Gaussian, Bilateral and Non local Means on the Fer2013 dataset. The results show that proposed model (FERConvNet) with Guassian Filter, Bilateral Filter, Non local Means Filter obtained 55% 65% 65% accuracies respectively on Fer2013 dataset. Similarly when the proposed novel hybrid denoising method, which is a combination of Gaussian, Bilateral, Non local means Filters, applied on Fer2013 dataset, the proposed model with hybrid denoising method (FERconvNet_HDM) achieved 85% accuracy on the test set. The FERConvNet_HDM, when compared to traditional filtering techniques performs better for facial expression recognition. The Tables 3 and 4 clearly show that FERConvNet performs better than the VGG16 and VGG19 on Fer2013. When the hybrid denoising method is applied to FERConvNet, the test accuracy increased from 65% to 85%. This clearly shows that our FERConvNet_HDM model outperformed the state-of-art VGG variants.

3.3 Performance on LRFE dataset

Low resolution facial expression (LRFE) dataset is created from various sources, the primary intention to create this dataset is, there are no existing datasets that contain images of low resolution for facial expression recognition. All the existing work in facial expression recognition is done on recognizing emotions on well posed conditions, but not in wild or real world conditions. So we created this LRFE dataset, where the images are taken in real world conditions. This dataset contains nearly 6000 images belonging to seven emotions (Happy, Sad, Surprise, Angry, Neutral, Disgust, Fear). Since all the images are collected from various resources, they are of different file extension formats (.JPG, .PNG, .GIF), we converted all the images to .JPG format. We used three image appearance filters and four affine transform matrices to increase the number of samples since convolutional neural networks require a large number of samples for training purposes.

The three image appearance filters used are average, bilateral, Gaussian filters. Therefore, the number of samples in LRFE dataset are now 35000. Now all the images are converted to grayscale and resized to 48X48 pixels. Then LRFE dataset is divided into training and testing set in the ratio 80:20 (80% training and 20% testing).

Table 5. Outline of number of samples in each expression of LRFE dataset

Dataset	Name & No. of images in each emotion						
	Happy	Sad	Angry	Disgust	Fear	Surprise	Neutral
LRFE	5162	5148	5218	5155	5002	5194	5274

Fig 6. Sample images of each expression from LRFE dataset

We now present the results on LRFE dataset, where the LRFE dataset is divided in the ratio 80:20 for training and testing, validation. The batch size used is 64 and trained for 100 epochs.

Table 6. curacy and loss of different models on LRFE dataset

S.no	Model Name	Dataset	Train Accuracy	Test Accuracy	Train Loss	Test Loss
1	VGG16	LRFE	0.87	0.69	0.40	1.16
2	VGG19	LRFE	0.84	0.66	0.47	0.96
3	FERConvNet	LRFE	0.95	0.71	0.16	1.19
4	EfficientNetB7	LRFE	0.79	0.65	0.71	1.09

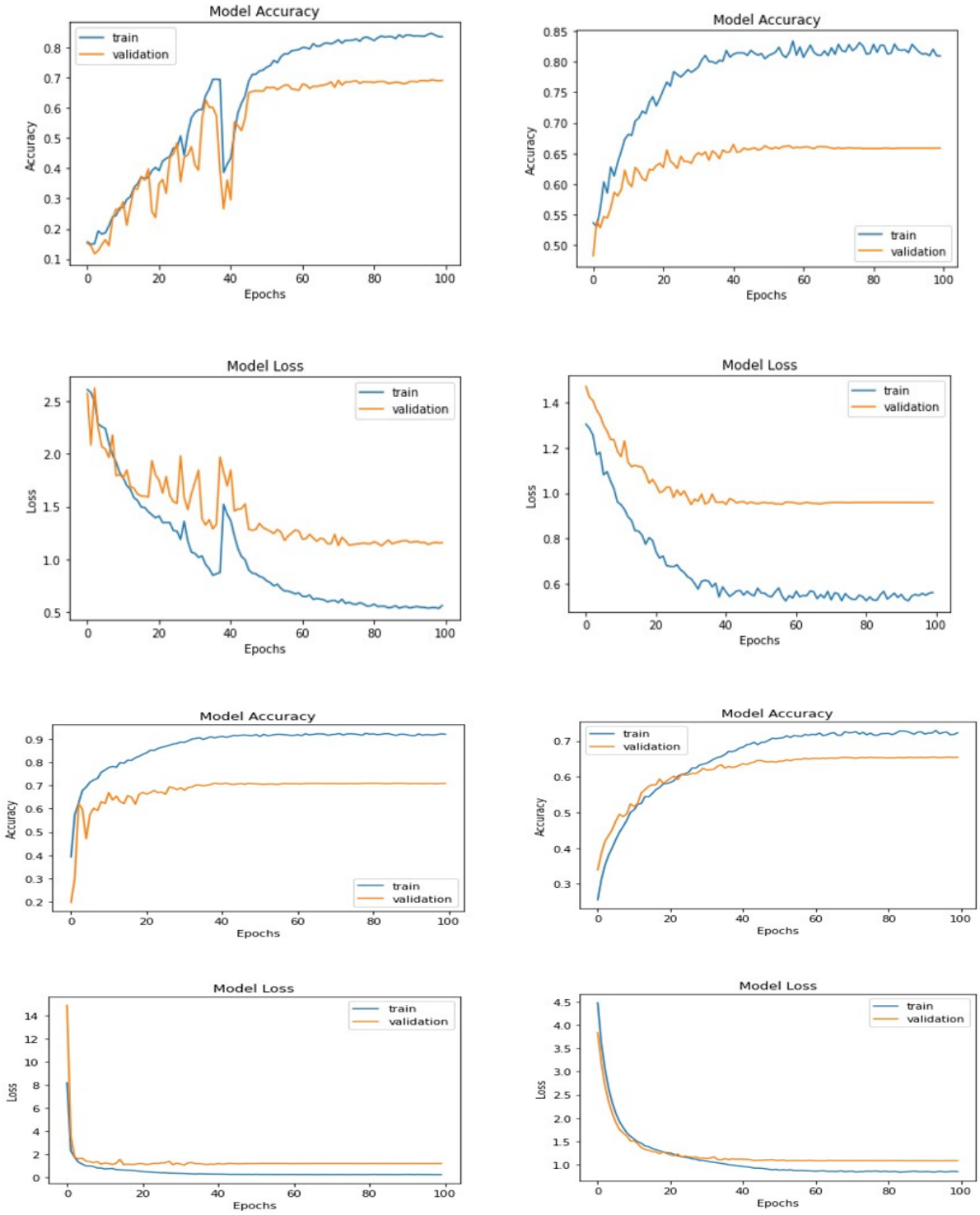

Fig 7. a. Accuracy & Loss Comparison of VGG 16 Model on LRFE, b. Accuracy & Loss Comparison of VGG 19 Model on LRFE, c. Accuracy & Loss Comparison of Proposed FERConvNet Model on LRFE, d. Accuracy & Loss comparison Of EfficientNetB7 model on LRFE

Table 7. Accuracy and Loss Comparison of Proposed FERConvNet Model on FER2013 dataset after Applying Various Denoising Techniques

S.no	Model Name	Dataset	Train Accuracy	Test Accuracy	Train Loss	Test Loss
1	FERConvNet_Gaussian	LRFE	0.98	0.58	0.30	3.00
2	FERConvNet_Bilateral	LRFE	0.98	0.63	0.43	2.52
3	FERConvNet_Nonlocal Means	LRFE	0.93	0.61	0.79	2.32
4	FERConvNet_HDM	LRFE	0.98	0.95	0.07	0.33

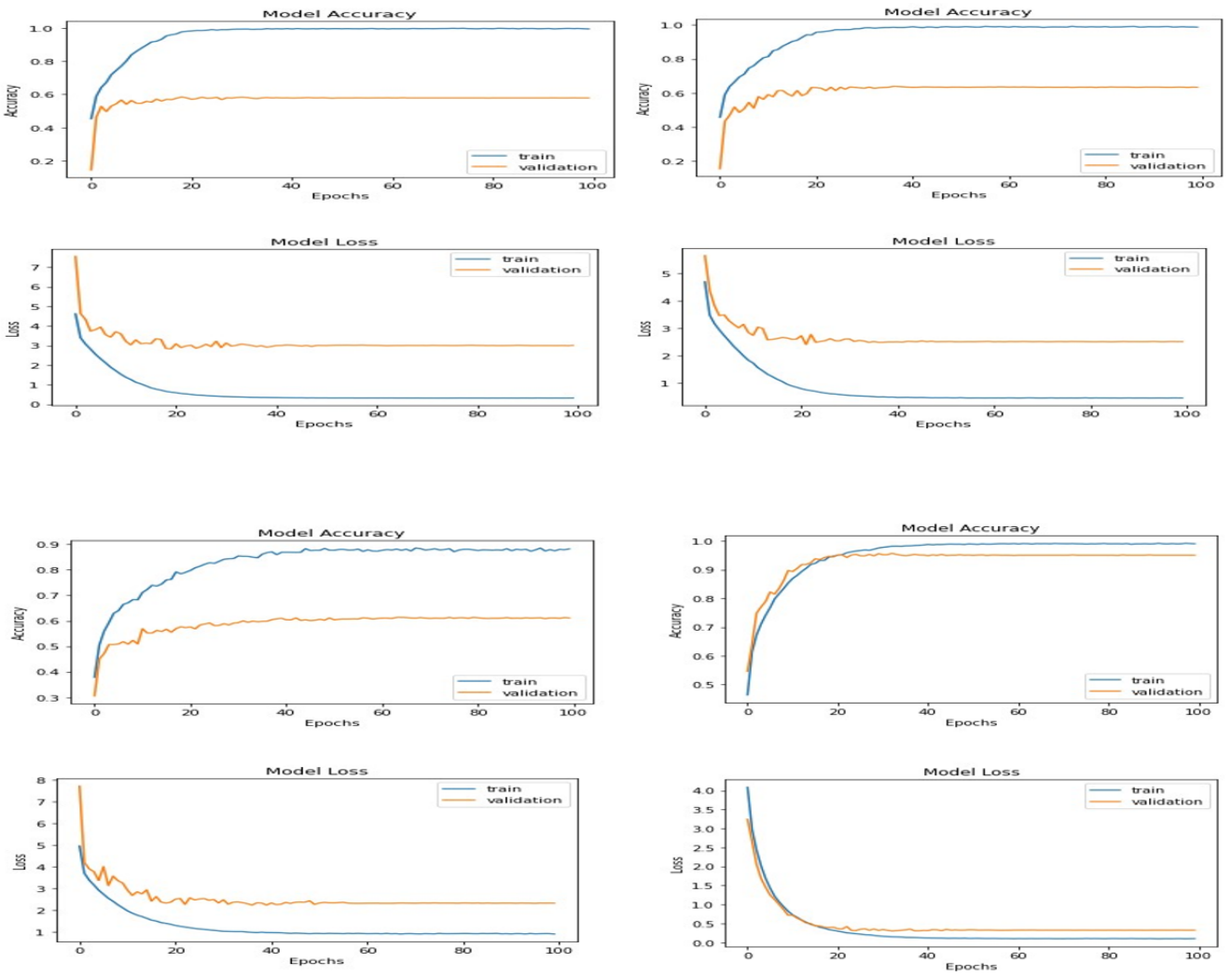

Fig 8. a. Accuracy & Loss Comparison of Proposed FERConvNet Model on LRFE with Gaussian Denoising, b. Accuracy & Loss Comparison of Proposed FERConvNet Model on LRFE with Bilateral Denoising, c. Accuracy & Loss Comparison of Proposed FERConvNet Model on LRFE with Nonlocal Means Denoising, d. Accuracy & Loss Comparison of Proposed FERConvNet Model on LRFE with Hybrid Denoising

Tables 6 and 7 show the results of different convolutional neural network models on Low resolution facial expression (LRFE) dataset. State-of-art models like VGG16, VGG19 and EfficientNetB7 obtained 69% 66% 65% accuracies on LRFE dataset. The average time taken per each epoch for VGG variants on LRFE dataset is 5sec, the average time taken per epoch for EfficientNetB7 on LRFE dataset is 7sec and the average time taken by FERConvNet on LRFE dataset is 2sec, since the FERConvNet model has less layers than VGG variants and EfficientNetB7. The proposed convolutional neural network model (FERConvNet) achieved 71% accuracy on this dataset, which is having better performance than VGG variants (VGG16 and VGG19) and EfficientNetB7.

The train loss of the FERConvNet is only 0.16, which is better than VGG16, VGG19 and EfficientNetB7 train loss. But the FERConvNet model accuracy is 71%, which is slightly better than both VGG16 and VGG19. We tried different techniques to increase the accuracy of FERConvNet on LRFE dataset, Tables 6 and 7 shows the performance of different techniques with FERConvNet. FERConvNet with Gaussian technique(FERConvNet_Gaussian) obtained only 58% accuracy, since loss of details in image is high, when Gaussian technique is used. So we applied hybrid denoising method(HDM) to FERConvNet, which is FERConvNet_HDM. This approach achieved 95% accuracy, outperforming the VGG16, VGG19 and EfficientNetB7 state-of-models. The train and test loss of FERConvNet_HDM are 0.07 and 0.33 respectively, these results show that this approach is overcoming the overfitting problem in convolutional neural networks for facial expression recognition.

4 Conclusion

In this study, a novel convolutional neural network (FERConvNet) and a new hybrid denoising method, which is a combination of Gaussian, Bilateral and Non local means filters, are presented. Since there are no existing datasets for low resolution images for facial expression recognition, we created a low resolution facial expression (LRFE) dataset. This dataset contains nearly 6000 images of seven different emotions (Happy, Sad, Surprise, Fear, Angry, Neutral, Disgust). Since convolutional neural networks require large number of samples for training, we used three image appearance filters and four affine transform matrices to increase the number of samples. After applying these techniques, the number of samples in LRFE dataset increased to 35000. The proposed FERConvNet_HDM approach achieved 85% accuracy on Fer2013 dataset, outperforming the VGG16, VGG19 and EfficientNetB7 models, whose accuracies are 60% 53% 60% on Fer2013 dataset respectively. The same FERConvNet_HDM approach when applied on LRFE dataset achieved 95% accuracy. After analyzing the results, our FERConvNet_HDM approach performs better than VGG16, VGG19 and EfficientNetB7 on both Fer2013 and LRFE dataset. Our approach is computationally simple and robust in terms of low resolution images, which are close to real world conditions, making our proposed model as promising for real world applications.

References

1) Kopparapu S, Kumar M, Satish. Identifying Optimal Gaussian Filter for Gaussian Noise Removal. In: and others, editor. 2011 Third National Conference on Computer Vision, Pattern Recognition, Image Processing and Graphics. 2011. Available from: 10.1109/ncvpripg.2011.34.
2) Tomasi C, Manduchi R. Bilateral filtering for gray and color images. In: and others, editor. Sixth International Conference on Computer Vision. IEEE. . Available from: 10.1109/iccv.1998.710815.
3) Deledalle CA, Duval V, Salmon J. Non-local Methods with Shape-Adaptive Patches (NLM-SAP). *Journal of Mathematical Imaging and Vision*. 2012;43(2):103–120. Available from: https://dx.doi.org/10.1007/s10851-011-0294-y.
4) Deng J, Pang G, Zhang Z, Pang Z, Yang H, Yang G. cGAN Based Facial Expression Recognition for Human-Robot Interaction. *IEEE Access*. 2019;7:9848–9859. Available from: https://dx.doi.org/10.1109/access.2019.2891668.
5) Zhang H, Jolfaei A, Alazab M. A Face Emotion Recognition Method Using Convolutional Neural Network and Image Edge Computing. *IEEE Access*. 2019;7:159081–159089. Available from: https://dx.doi.org/10.1109/access.2019.2949741.
6) Wang Y, Li Y, Song Y, Rong X. The Application of a Hybrid Transfer Algorithm Based on a Convolutional Neural Network Model and an Improved Convolution Restricted Boltzmann Machine Model in Facial Expression Recognition. *IEEE Access*. 2019;7:184599–184610. Available from: https://dx.doi.org/10.1109/access.2019.2961161.
7) Jian BL, Chen CL, Huang MW, Yau HT. Emotion-Specific Facial Activation Maps Based on Infrared Thermal Image Sequences. *IEEE Access*. 2019;7:48046–48052. Available from: https://dx.doi.org/10.1109/access.2019.2908819.
8) Pham TTD, Kim S, Lu Y, Jung SW, Won CS. Facial Action Units-Based Image Retrieval for Facial Expression Recognition. *IEEE Access*. 2019;7:5200–5207. Available from: https://dx.doi.org/10.1109/access.2018.2889852.
9) Lo L, Xie HX, Shuai HH, Cheng WH. MER-GCN: Micro-Expression Recognition Based on Relation Modeling with Graph Convolutional Networks. *2020 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR)*. 2020. doi:10.1109/mipr49039.2020.00023.
10) Melinte DO, Vladareanu L. Facial Expressions Recognition for Human–Robot Interaction Using Deep Convolutional Neural Networks with Rectified Adam Optimizer. *Sensors*. 2020;20(8):2393–2393. Available from: https://dx.doi.org/10.3390/s20082393.
11) Ch S, Kiruthika K, Priya B, Jayalakshmi R, J. Detection ofHuman Facial Expression using CNN Model. *International Journal of Engineering and Advanced Technology*. 2020;9(5):2249–8958. doi:10.35940/ijeat.E9615.069520.
12) Sreenivas V, Namdeo V, Kumar EV. Group based emotion recognition from video sequence with hybrid optimization based recurrent fuzzy neural network. *Journal of Big Data*. 2020;7(1):1–21. Available from: https://dx.doi.org/10.1186/s40537-020-00326-5.
13) Lu P, Song B, Xu L. Human face recognition based on convolutional neural network and augmented dataset. *Systems Science & Control Engineering*. 2020. Available from: https://dx.doi.org/10.1080/21642583.2020.1836526.
14) Deng L, Wang Q, Yuan D. Dynamic Facial Expression Recognition Based on Deep Learning. 2019. Available from: 10.1109/iccse.2019.8845493.
15) Kim JH, Kim BG, Roy PP, Jeong DM. Efficient Facial Expression Recognition Algorithm Based on Hierarchical Deep Neural Network Structure. *IEEE Access*. 2019;7:41273–41285. Available from: https://dx.doi.org/10.1109/access.2019.2907327.
16) Yue Z, Yanyan F, Shangyou Z, Bing P. Facial Expression Recognition Based on Convolutional Neural Network. In: and others, editor. 2019 IEEE 10th International Conference on Software Engineering and Service Science (ICSESS). 2019. Available from: 10.1109/icsess47205.2019.9040730.
17) Filntisis PP, Efthymiou N, Koutras P, Potamianos G, Maragos P. Gerasimos Potamianos, and Petros Maragos. "Fusing Body Posture With Facial Expressions for Joint Recognition of Affect in Child-Robot Interaction. *IEEE Robotics and Automation Letters*. 2019;4(4):4011–4018. doi:10.1109/LRA.2019.2930434.