Deep Leaning-Based Facial Expression Recognition in FER2013 Database: An in-Vehicle Application

Goutam Kumar Sahoo*,1, Jayakrishna Ponduru*,2, Santos Kumar Das*,3 and Poonam Singh*,4
*Department of ECE, National Institute of Technology Rourkela, Odisha, India
Email: 1goutamkrsahoo@gmail.com, 2pondurujayakrish@gmail.com, 3dassk@nitrkl.ac.in, 4psingh@nitrkl.ac.in

2022 IEEE 19th India Council International Conference (INDICON) | 978-1-6654-7350-7/22/$31.00 ©2022 IEEE | DOI: 10.1109/INDICON56171.2022.10040121

***Abstract*—This work presents a deep learning-based approach for the evaluation of facial expression recognition (FER) performance. The main objective is to develop a deep convolutional neural network (CNN) to perform FER using the publicly available benchmark dataset, the FER2013 dataset. The FER2013 dataset includes hand-based facial occlusion, incorrectly cropped or partial images, images with glasses, low-resolution images, etc., which are close to real driving complex scenarios. Two custom CNN models and a pre-trained VGG16 model are evaluated for the FER task. The Deep CNN model with 10-layer architecture shows the best performance accuracy of 68.34%. This deep CNN model can be used to monitor driver behavior from front face images captured via dashboard camera and alert the driver to improve their driving style for a safer drive.**

***Index Terms*—Facial Expression Recognition (FER), Driver Behavior, Deep Learning, FER2013 Dataset, Driving Safety.**

I. Introduction

A person's feelings are expressed physically through their facial expressions. Facial expression is one of the most natural and universal cues to express human emotional state and intentions. Literature studies have shown automated facial expression analysis to be of practical importance in health care, driver fatigue monitoring, and many other human-computer interaction (HCI) systems [1]. The field of computer vision is one of the areas where deep neural network (DNN) systems have appeared with exceptional performance to solve various problems. The use of DNN to recognize facial expressions with a given image dataset is often defined as emotion recognition. Potential applications of facial emotion recognition could be used as a coordination tool to assess people's temperament or to identify responses to diverse stimuli. Ekman and Frisson defined six basic emotions, indicating that humans understand emotions such as anger, disgust, fear, joy, sadness, and surprise [2]. Contempt or neutrality was later added as one of the core emotions [3]. There are seven categories of facial expressions available in most datasets, which are happy, sad, angry, disgust, neutral, fear and surprise. A work by Nandyala *et al.* [4] presented a driver monitoring system (DMS), which used machine learning and image processing algorithms to identify hidden emotions such as distraction and drowsiness. Alerts of different levels such as tired and completely sleepy were generated to avoid a fatal accident. Next, it used a convolutional neural network (CNN) model to analyze visual images to classify emotions into six basic emotion categories. Standard and Tata Alexi proprietary databases were used for model training and real-time testing of emotion recognition on board the Raspberry Pi.

Work by Suchitra *et al.* [5] presented CNN-based facial expression recognition using local octal pattern (LOP) feature descriptors. Recognition and classification of facial emotions was accomplished using a support vector machine (SVM) technique. Tasks performance metrics such as precision, recall and F-score were used to evaluate the proposed task. The model showed a high recall rate of 96.09% as compared to other cutting edge technologies. The work by Yan *et al.* [6] addressed the problem of insufficient training data and data redundancy to achieve better performance on this facial expression recognition task. It also discussed eliminating the results of expression-independent modifications (e.g. head posture, lighting conditions). Generative Adversarial Networks (GANs) were used to augment the dataset, filtering out irrelevant factors via cascading networks, and over-fitting problems brought on by insufficient training data and pointless changes in expression. Network integration was attempted to resolve the over-fitting problems. Various models such as VGG13, VGG16, VGG19, and ResNet-50+ deep convolutional GANs were evaluated using the FER-2013 dataset. According to [7], anger and aggression are two emotions that have a significant impact on how people drive and raise the possibility of accidents. Similarly, fatigue and stress are other causes of dangerous driving. Nervousness, sadness, and other strong emotions can also affect driving. Other factors that contribute to unsafe driving include exhaustion and stress. Driving can also be impacted by nervousness, sadness, and other powerful emotions. Recognizing a driver's feelings and making them aware of them are the first steps in managing their emotions. Additionally, in order to drive safely, a person's mental condition must be supported by skills including good traffic judgement, awareness, appropriate decision-making, and communication with other drivers. Facial expressions play a very important role in recognizing the current emotions of the driver.

This task uses a publicly available benchmark facial emotion recognition (FER) image dataset, "FER2013", for performance evaluation [8]. It is the most challenging and unbalanced dataset collected from the Internet, commonly used for FER purposes. The study offers a framework for recognising facial expressions that successfully applies characteristics taken from fully connected layers of a pre-trained VGG16 model. Also, the performance of the model is evaluated using 6-layer

CNN and 10-layer CNN models designed from scratch. The following are the paper's main contributions.

- Use of a 6-layer and 10-layer CNN framework for FER drivers
- The use of data pre-processing techniques to prepare the required inputs for deep learning networks.
- A "pre-trained VGG16" model is used to implement a framework for FER.
- FER2013 benchmark picture dataset is used performance evaluation.
- Result discussion of the proposed method with state-of-the-art (SOTA) methods.

II. Problem Identification and Methodology

A. Problem Definition

Generally, most of the road accidents are due to personal faults, negligence or drowsiness. Deep learning (DL) techniques have been used effectively to differentiate driving styles and recognize risky behavioral activities. From this point of view, identifying the facial expressions of drivers in such a way that we can alert the driver when he/she departs from normal driving activity. This task is designed to evaluate FER performance on two CNN models designed from scratch and two transfer learning-based VGG models. Basically in this, we aim to measure how accurately the model is predicting the labeled emotions in order to alert the driver to a safe drive.

B. Proposed Methodology

1) Using CNN-based Approach: Fig. 1 represents the proposed system architecture that performs FER tasks using a CNN model. The system would be fed images of the driver's front face captured by the vehicle's webcam. The suggested system's input requirements are then met by applying image pre-processing techniques such image scaling, data augmentation, etc. In order to classify face expressions, the algorithm then extracts depth information from the images.

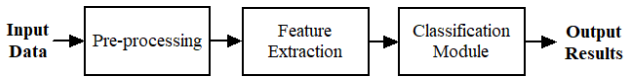

Fig. 1. FER System Overview Using CNN Model.

In classification tasks including object detection, picture recognition, and computer vision, convolutional neural networks are frequently utilised. It is widely used for image/video related tasks, but can also be used for text and voice data. The architecture of convolutional neural network is also similar to that of regular neural network. The main operation here is a convolution which means that the input image matrix is convolved with another matrix called image filter or kernel. So, here it says that the spatial structure of an image is being preserved in the reduced size of the input. This filter activates certain features from images, such as edges. The convolution layer, the pooling layer, and the fully connected layer are the three primary layers of a CNN [9]. The study presents a facial expression recognition framework to evaluate two models designed from scratch, one with a 6-layer CNN and the other with a 10-layer CNN. A step-wise approach to the FER method using CNN is presented in Algorithm 1.

Algorithm 1 : Recognizing facial expressions using CNN

1: Load the FER2013 dataset's CSV file.
2: Make use of the pre-processing methods.
3: Split the dataset into training and testing samples.
4: Build the CNN model using convolutional layers, maxPooling layer, dropout layer, fully connected layers and softmax classification layer.
5: Feature extraction from the dense layer of the model.
6: Test data to the model for emotion classification.
7: Calculate the performance measurement parameters.

Layer (type)	Output Shape	Param #
conv2d (Conv2D)	(None, 48, 48, 64)	640
conv2d_1 (Conv2D)	(None, 48, 48, 64)	36928
batch_normalization (BatchNo	(None, 48, 48, 64)	256
max_pooling2d (MaxPooling2D)	(None, 23, 23, 64)	0
dropout (Dropout)	(None, 23, 23, 64)	0
conv2d_2 (Conv2D)	(None, 23, 23, 64)	36928
conv2d_3 (Conv2D)	(None, 23, 23, 64)	36928
batch_normalization_1 (Batch	(None, 23, 23, 64)	256
max_pooling2d_1 (MaxPooling2	(None, 11, 11, 64)	0
dropout_1 (Dropout)	(None, 11, 11, 64)	0
conv2d_4 (Conv2D)	(None, 11, 11, 128)	73856
conv2d_5 (Conv2D)	(None, 11, 11, 128)	147584
batch_normalization_2 (Batch	(None, 11, 11, 128)	512
max_pooling2d_2 (MaxPooling2	(None, 5, 5, 128)	0
dropout_2 (Dropout)	(None, 5, 5, 128)	0
flatten (Flatten)	(None, 3200)	0
dense (Dense)	(None, 2048)	6555648
dropout_3 (Dropout)	(None, 2048)	0
dense_1 (Dense)	(None, 7)	14343

Total params: 6,903,879
Trainable params: 6,903,367
Non-trainable params: 512

Fig. 2. Different layer details of the proposed 6-layer CNN model.

Fig. 2 represents the model configuration of a 6-layer CNN model. Each input data of size 48×48 is used from the FER2013 dataset and 90% of the data is allocated for training, while 10% is allocated for testing., and 10% of the data from 90% of the train data is allocated for validation. Data augmentation is used here, along with rotating the images by 20 degrees and moving them horizontally and vertically. The categorical cross-entropy loss function and the Adam optimizer are used to train the model over 300 iterations with a mini-batch size of 32. A 10-layer deep model is also designed

as in Fig. 3 and evaluated with the same network parameters to verify the performance of the deep CNN model. The performance parameters of both the CNN models are evaluated for comparison purposes and it shows an improvement in performance with the increase in the number of layers.

Layer (type)	Output Shape	Param #
conv2d (Conv2D)	(None, 48, 48, 64)	640
conv2d_1 (Conv2D)	(None, 48, 48, 64)	36928
batch_normalization (BatchNo	(None, 48, 48, 64)	256
max_pooling2d (MaxPooling2D)	(None, 23, 23, 64)	0
dropout (Dropout)	(None, 23, 23, 64)	0
conv2d_2 (Conv2D)	(None, 23, 23, 64)	36928
conv2d_3 (Conv2D)	(None, 23, 23, 64)	36928
batch_normalization_1 (Batch	(None, 23, 23, 64)	256
max_pooling2d_1 (MaxPooling2	(None, 11, 11, 64)	0
dropout_1 (Dropout)	(None, 11, 11, 64)	0
conv2d_4 (Conv2D)	(None, 11, 11, 128)	73856
conv2d_5 (Conv2D)	(None, 11, 11, 128)	147584
batch_normalization_2 (Batch	(None, 11, 11, 128)	512
max_pooling2d_2 (MaxPooling2	(None, 5, 5, 128)	0
dropout_2 (Dropout)	(None, 5, 5, 128)	0
conv2d_6 (Conv2D)	(None, 5, 5, 128)	147584
conv2d_7 (Conv2D)	(None, 5, 5, 128)	147584
batch_normalization_3 (Batch	(None, 5, 5, 128)	512
max_pooling2d_3 (MaxPooling2	(None, 2, 2, 128)	0
dropout_3 (Dropout)	(None, 2, 2, 128)	0
conv2d_8 (Conv2D)	(None, 2, 2, 128)	147584
conv2d_9 (Conv2D)	(None, 2, 2, 128)	147584
batch_normalization_4 (Batch	(None, 2, 2, 128)	512
max_pooling2d_4 (MaxPooling2	(None, 1, 1, 128)	0
dropout_4 (Dropout)	(None, 1, 1, 128)	0
flatten (Flatten)	(None, 128)	0
dense (Dense)	(None, 2048)	264192
dropout_5 (Dropout)	(None, 2048)	0
dense_1 (Dense)	(None, 7)	14343

Total params: 1,203,783
Trainable params: 1,202,759
Non-trainable params: 1,024

Fig. 3. Different layer details of the proposed 10-layer CNN model.

2) Transfer Learning-based Approach: A FER system based on the transfer learning technique is shown in Fig. 4 [10]. The transfer learning-based model uses a pre-trained VGG network to extract deep features from images for facial emotion classification. The transfer learning approach, in general, reduces computational complexity and can be used to embedded systems for in-vehicle applications that demand real-time processing. This FER task using is evaluated using the publicly available benchmark dataset the FER2013. The FER-2013 dataset we utilise in this FER experiment is different from the VGG architecture, which was pre-trained on image size of ImageNet dataset. As a result, we resize each image in the FER-2013 dataset to match ImageNet's dimensions. A simple CNN architecture utilised in ImageNet contests is called VGG16 (Visual Geometry Group 16). A step-wise approach to the FER method using transfer learning-based approach is presented in Algorithm 2.

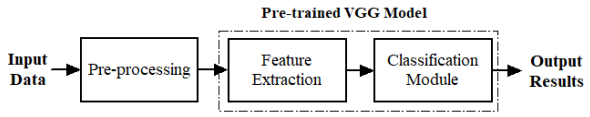

Fig. 4. Overview of a FER system based on a pre-trained VGG model.

Algorithm 2 : FER Using Pre-trained VGG16 Model

1: Load the FER2013 dataset's CSV file.
2: Make use of the pre-processing methods.
3: Create training and test samples from the dataset.
4: Utilize pre-trained VGG16 network parameters for the model.
5: Extraction of features from the model's dense layer.
6: Test image data for the emotion categorization model.
7: Determine the parameters for performance measurement.

Fig. 5 represents the network configuration of the transfer learning VGG16 model. FER2013 dataset is 48×48 which is not suitable to be used in transfer learning VGG model, so the data is converted to 224×224 as required for transfer learning VGG model. 80-20% of the data-split is used, where 80% of the data is allocated for training, and 20% for data validation. Data augmentation is used here, by rotating the images 20 degrees and rotating them horizontally and vertically. The model is trained with a categorical cross-entropy loss function and the Adam optimizer for 100 epochs with a mini-batch size of 32.

III. Results and Discussion

The Google Colab platform was used for all experiments. Deep learning models are constructed to analyze FER tasks using Python, Numpy, pandas, Keras, and TensorFlow libraries.

A. Database Used

The benchmark FER database 'FER2013' is well-known and is publicly available [8]. A total of 35887 images were collected from the Internet using Google Image Search for each particular emotion. However, these databases are mainly captured under certain lighting conditions in an indoor environment. The dataset includes hand-based facial occlusion, incorrectly cropped or partial images, images with glasses, low-resolution images, etc., which are close to real complex scenarios. Also, the number of images in each emotion category is not the same, making the dataset unbalanced [11], [12]. Fig. 6 shows the training data imbalance across multiple

Layer (type)	Output Shape	Param #
input_1 (InputLayer)	[(None, 224, 224, 3)]	0
block1_conv1 (Conv2D)	(None, 224, 224, 64)	1792
block1_conv2 (Conv2D)	(None, 224, 224, 64)	36928
block1_pool (MaxPooling2D)	(None, 112, 112, 64)	0
block2_conv1 (Conv2D)	(None, 112, 112, 128)	73856
block2_conv2 (Conv2D)	(None, 112, 112, 128)	147584
block2_pool (MaxPooling2D)	(None, 56, 56, 128)	0
block3_conv1 (Conv2D)	(None, 56, 56, 256)	295168
block3_conv2 (Conv2D)	(None, 56, 56, 256)	590080
block3_conv3 (Conv2D)	(None, 56, 56, 256)	590080
block3_pool (MaxPooling2D)	(None, 28, 28, 256)	0
block4_conv1 (Conv2D)	(None, 28, 28, 512)	1180160
block4_conv2 (Conv2D)	(None, 28, 28, 512)	2359808
block4_conv3 (Conv2D)	(None, 28, 28, 512)	2359808
block4_pool (MaxPooling2D)	(None, 14, 14, 512)	0
block5_conv1 (Conv2D)	(None, 14, 14, 512)	2359808
block5_conv2 (Conv2D)	(None, 14, 14, 512)	2359808
block5_conv3 (Conv2D)	(None, 14, 14, 512)	2359808
block5_pool (MaxPooling2D)	(None, 7, 7, 512)	0
flatten (Flatten)	(None, 25088)	0
dense (Dense)	(None, 7)	175623

Total params: 14,890,311
Trainable params: 175,623
Non-trainable params: 14,714,688

Fig. 5. Different layer details of the transfer learning VGG16 model.

observations and data pose variation can also be observed from the images of different classes. The "happy" class has the highest 6292 and "disgust'' class has the lowest 383 observations, with model training data splitting 90%.

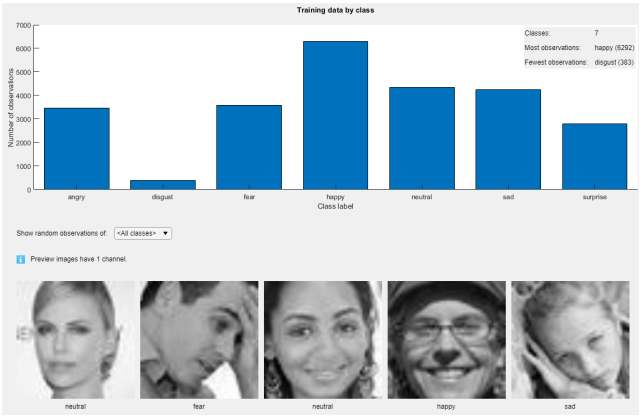

Fig. 6. Different emotional class details of FER2013 database.

B. Experimental Results and Analysis

Using the FER2013 dataset, this experimental study compares transfer-learning VGG networks and CNN models for the FER classification problem. Model training and testing performance parameters are obtained. Measuring the model training time shows how long it takes for our DL model to be trained on the training dataset. The measure of model fit time or training time is a floating number and is represented in seconds. Similarly, computational speed is the amount of time our trained model takes to process a set of new data for prediction. According to the model's training results, the 10-layer CNN had the highest accuracy, which was 68.34%. However, the pre-trained VGG16 model shows lower performance than both CNN models. Table I gives a comparative performance on the number of trained and non-trained parameters and the speed of the model's performance. It can be concluded that the 10-layer CNN model shows higher performance accuracy with a slightly higher computational time cost than the 6-layer CNN.

1) Result analysis using CNN with 6 layer: The 6-layer deep CNN model's performance, which was trained across 300 epochs, yielded a FER accuracy of 66.67 with a mini-batch size of 32. In this model, the data augmentation is done first which means that the existing images are rotated horizontally, vertically or at a rotation angle of 20 degrees. The model shows the trend of increasing performance from the training curves. The test accuracy shows good performance which can be seen in Fig. 7. Similarly, model loss performances are plotted and can be seen in Fig. 8. The model loss performance curve can be seen decreasing towards zero and a low loss value indicates good model performance.

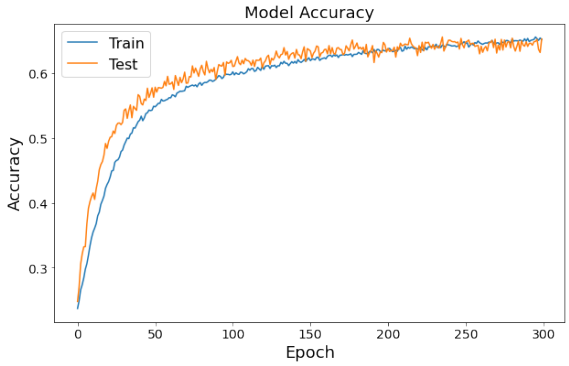

Fig. 7. Accuracy performance of 6-layer CNN model.

2) Result analysis using CNN with 10 layer: The 10-layer deep CNN model's performance, which was trained across 300 epochs, yielded a FER accuracy of 68.34% with a mini-batch size of 32. The model shows the trend of increasing performance from the training curves. The test accuracy shows good performance which can be seen in Fig. 9. Similarly, model loss performances are plotted and can be seen in Fig. 10. The model loss performance curve can be seen decreasing towards zero and a low loss value indicates good model performance.

TABLE I
FACIAL EXPRESSION RECOGNITION MODEL PERFORMANCE COMPARISON BASED ON TRAINING PARAMETERS.

Model used	Trainable parameter	Non-trainable parameter	Training speed (Sec)	Computational speed (Sec)	Test accuracy (%)
6-layer CNN	6903367	512	5448.50	15.98	66.67
10-layer CNN	1202759	1024	5569.54	16.57	68.34
Pre-trained VGG16	175623	14714688	78749.55	50.12	63.68

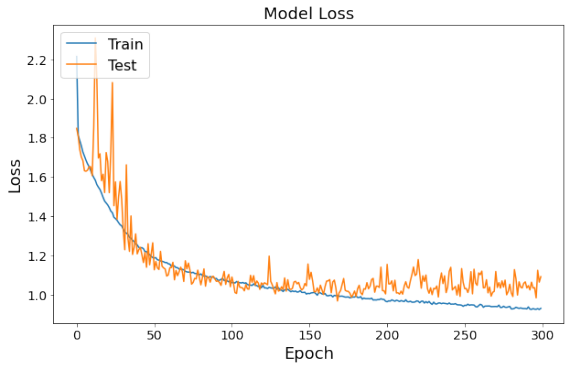

Fig. 8. Loss Performance of 6-layer CNN model.

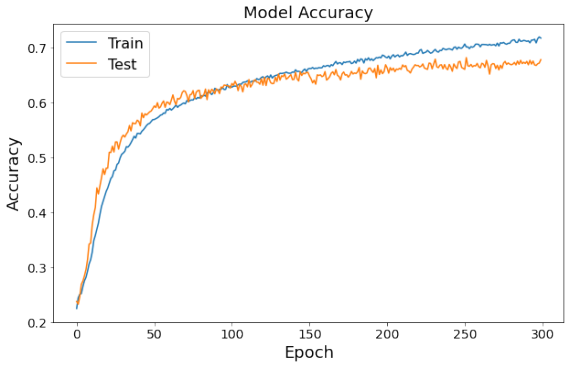

Fig. 9. Accuracy performance of 10-layer CNN model.

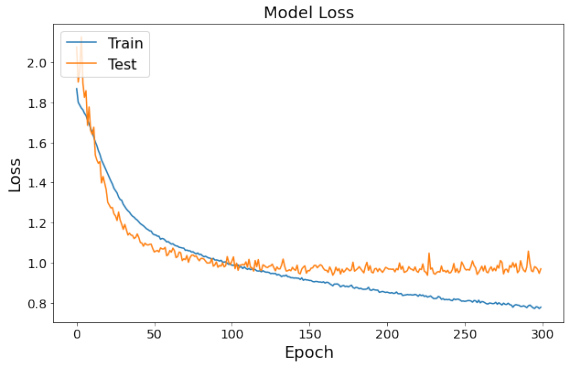

Fig. 10. Loss Performance of 10-layer CNN model.

Fig. 11 shows the confusion matrix for the best model with 10-layer CNN architecture. From the Fig. 11, the rows correspond to the true/actual class (the target class) and the columns correspond to the predicted class (output class). Diagonal cells represent correctly classified, and off-diagonal cells correspond to incorrectly classified observations. If we analyze the facial expression *'angry'* class of the FER2013 dataset, out of the total 491 images of the facial expression, 505 was predicted (296 images correctly predicted and 209 wrongly predicted). Similarly, if we analyze the facial expression *'disgust'* class, out of the total 55 images of the facial expression, 29 was predicted (20 images correctly predicted and 09 wrongly predicted). Similarly, we can observe the performance matrices for all other classes. The overall test accuracy is found to be of 68.34% from the confusion matrix.

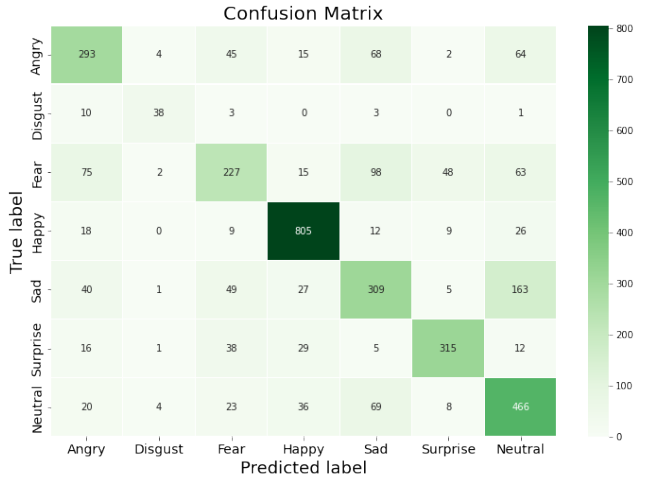

Fig. 11. Confusion matrix of 10-layer CNN model.

3) Result analysis using pre-trained VGG16 model: The performance of the pre-trained VGG16 model on the FER task achieved an accuracy of 63.68% when the model was trained for 100 epochs with a mini-batch size of 32 using FER2013 dataset. The model training and validation curves show the accuracy performance, which can be seen in Fig. 12. Similarly, model loss performances are plotted and can be seen in Fig. 13. The model training and validation loss curves show the performance of the model. The validation loss value increases which can lead to poor performance.

C. Comparison with SOTA models

The approach of recognising facial emotions has been researched in the literature. Performance comparison of different models such as CNN model from scratch and transfer learning-based CNN model is evaluated for sentiment recognition. Table II presents performance comparison results with state-of-the-art technologies applied to the FER2013 benchmark dataset.

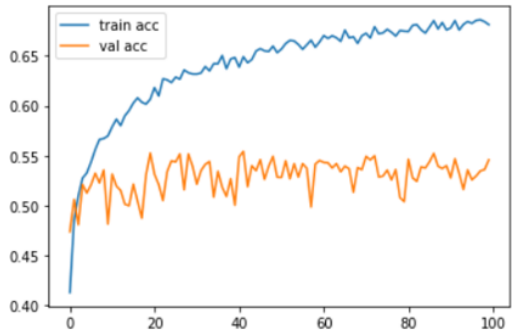

Fig. 12. Accuracy performance of pre-trained VGG16 model.

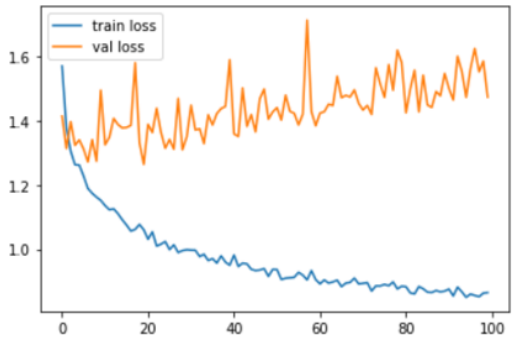

Fig. 13. Loss Performance of pre-trained VGG16 model.

TABLE II
FER PERFORMANCE COMPARISON USING FER2013 DATABASE.

Reference	Model Used	Test Accuracy(%)
A. Krishnadas and S. Nithin [13]	Pre-trained VGG-16	57.3
Gunawan *et al.* [14]	CNN	57.4
Bhatti *et al.* [15]	RELM	62.7
Chand *et al.* [12]	Xception	68.57
Bodapati *et al.* [16]	FERNet	69.57
Yan *et al.* [6]	ResNET-50	64.28
Panagiotis *et al.* [17]	GoogleNet	65.2
Proposed	6-layer CNN	66.67
	10-layer CNN	68.34
	Pre-trained VGG16	63.68

*RELM: Regularized Extreme Learning Machine.

The work presented by A. Krishnadas and S. Nitin [13] to classify the emotional state of driving behavior evaluated on FER2013 wild dataset for using artificial intelligence techniques. The deep learning CNN and machine learning SVM algorithms achieved 57% and 34%, respectively, while the performance accuracy obtained using the transfer learning VGG16 model was 57.3%. Similarly, Gunawan *et al.* [14] used a deep learning CNN model for video-based FER, which yielded a performance accuracy of 57.4% on the FER2013 database. Mini-batch size of 256 and data segmentation of 80-20% is used to work on Google Colab platform. However, when used on the FER2013 dataset as described above, the proposed technique outperforms the existing studies [13]–[15]. In addition, several works on FER are reported in Table II using different models that show comparable results with the proposed works.

IV. Conclusion

In this work, we have evaluated pre-trained VGG16 and custom CNN models on the FER2013 dataset with 6 layers and 10 layers. The FER2013 dataset is complex and imbalanced. The use of data augmentation successfully improved the accuracy. To validate the model's efficacy, the performance of the presented models is also compared with SOTA techniques. The deep model with 10-layer CNN achieves a maximum accuracy of 68.34 percent. Deep CNN model in FER can be helpful in implementing in-vehicle embedded system for driver assistance for a safe drive. In the future, other benchmark datasets, particularly data captured using a real vehicle driving environment, may be used with developed models for performance evaluation.

References

[1] S. Li and W. Deng, "Deep facial expression recognition: A survey," *IEEE transactions on affective computing*, 2020.
[2] P. Ekman and W. V. Friesen, "Constants across cultures in the face and emotion." *J. Personality Soc. Psychology*, vol. 17, no. 2, 1971.
[3] D. Matsumoto, "More evidence for the universality of a contempt expression," *Motivation and Emotion*, vol. 16, no. 4, pp. 363–368, 1992.
[4] Nandyala *et al.*, "Emotion analytics for advanced driver monitoring system," SAE Technical Paper, Tech. Rep., 2019.
[5] S. Suchitra, P. Sathya, P. Balachandran, and M. Faustina, "Intelligent driver warning system using deep learning-based facial expression recognition," *Scopus*, vol. 8, no. 3, pp. 831–838, 2019.
[6] B. Yan, Z. Xiao, P. Yuan, K. Cai, and Q. Chen, "Facial expression recognition with convolutional neural networks via a data augmentation strategy," 2021.
[7] Eyben *et al.*, "Emotion on the roadnecessity, acceptance, and feasibility of affective computing in the car," *Advances in human-computer interaction*, vol. 2010, 2010.
[8] *Facial Expression Recognition 2013 Dataset (FER2013).*, [Online]. Available: https://www.kaggle.com/c/challenges-in-representation-learning-facial-expression-recognition-challenge.
[9] Z. Fei, E. Yang, D. D.-U. Li, S. Butler, W. Ijomah, X. Li, and H. Zhou, "Deep convolution network based emotion analysis towards mental health care," *Neurocomputing*, vol. 388, pp. 212–227, 2020.
[10] G. K. Sahoo, S. K. Das, and P. Singh, "Deep learning-based facial emotion recognition for driver healthcare," in *Proc. IEEE National Conference on Communications (NCC)*, 2022, pp. 154–159.
[11] L. Pham, T. H. Vu, and T. A. Tran, "Facial expression recognition using residual masking network," in *Proc. IEEE 25th Int. Conf. Pattern Recog. (ICPR)*, 2021, pp. 4513–4519.
[12] S. Chand, A. Singh, R. Bhatia, I. Kaur, and K. Seeja, "Real-time facial emotion recognition using deep learning," in *Intelligent Computing and Communication Systems*. Springer, 2021, pp. 219–226.
[13] A. Krishnadas and S. Nithin, "A comparative study of machine learning and deep learning algorithms for recognizing facial emotions," in *Proc. 2nd Int. Conf. Electro. Sustainable Commun. Syst. (ICESC)*, 2021, pp. 1506–1512.
[14] T. S. Gunawan, A. Ashraf, B. S. Riza, E. V. Haryanto, R. Rosnelly, M. Kartiwi, and Z. Janin, "Development of video-based emotion recognition using deep learning with google colab," *TELKOMNIKA*, vol. 18, no. 5, pp. 2463–2471, 2020.
[15] Y. K. Bhatti, A. Jamil, N. Nida, M. H. Yousaf, S. Viriri, and S. A. Velastin, "Facial expression recognition of instructor using deep features and extreme learning machine," *Computational Intelligence and Neuroscience*, vol. 2021, 2021.
[16] J. D. Bodapati, U. Srilakshmi, and N. Veeranjaneyulu, "Fernet: a deep cnn architecture for facial expression recognition in the wild," *Journal of The institution of engineers (India): series B*, vol. 103, no. 2, pp. 439–448, 2022.
[17] P. Giannopoulos, I. Perikos, and I. Hatzilygeroudis, "Deep learning approaches for facial emotion recognition: A case study on fer-2013," in *Adv. Hybrid. Intell. Methods*. Springer, 2018, pp. 1–16.