Deep Learning-Based Facial Emotion Recognition for Driver Healthcare

Goutam Kumar Sahoo*,1, Santos Kumar Das*,2 and Poonam Singh*,3
*Department of Electronics and Communication Engineering, National Institute of Technology Rourkela, India
Email: [1]goutamkrsahoo@gmail.com, [2]dassk@nitrkl.ac.in, [3]psingh@nitrkl.ac.in

***Abstract*—This study proposes deep learning-based facial emotion recognition (FER) for driver health care. The FER system will monitor the emotional state of the driver's face to identify the driver's negligence and provide immediate assistance for safety. This work uses a transfer learning-based framework for FER which will help in developing an in-vehicle driver assistance system. It implements transfer learning SqueezeNet 1.1 to classify different facial expressions. Data preprocessing techniques such as image resizing and data augmentation have been employed to improve performance. The experimental study uses static facial expressions publicly available on several benchmark databases such as CK+, KDEF, FER2013, and KMU-FED to evaluate the model's performance. The performance comparison only showed superiority over state-of-the-art technologies in the case of the KMU-FED database, i.e., maximum accuracy of 95.83%, and the results showed comparable performance to the rest of the benchmark databases.**

***Keywords*—Deep Learning, Facial Emotion Recognition (FER), Driving Assistance, Transfer Learning, Driver Healthcare.**

I. Introduction

Driver behavior monitoring is vital in diagnosing driver's health conditions. Healthcare in real-time driving monitoring will establish an intelligent transport system (ITS) for safe driving. The advancement of information and communication technology (ICT) creates scope for implementing driver assistance systems based on the human-computer interface (HCI) [1]. Driver healthcare is an essential and most crucial factor of ITS and plays a vital role in setting up safe driving in smart cities. With the increase in the number of motor vehicles, road accidents are also increasing. Most road accidents occur due to driver's fault like a distraction due to use of mobile, aggressive driving, impairments due to alcohol consumption and drugs [2]–[5]. The monitoring of driver status can be established using driving patterns obtained through steering wheel movements, driver's physiological data captured through body sensors, and in-vehicle image data captured through vehicle dashboard camera [6]. Facial expression recognition (FER), eye-closure analysis, head pose estimation, etc., will help identify driving anomalies. Facial expressions are based on six basic emotions: happiness, surprise, anger, sadness, fear, and disgust [7], [8]. These basic emotions involve non-verbal communicative signals that do not occur very frequently in regular personal interactions and are extremely difficult to record truly spontaneous instances [9]. However, having these six basic emotions deliver a powerful message to the surroundings for health care and safety [10].

Recently different studies show state-of-the-art results of widely used FER approaches based on deep neural network (DNN) techniques. DNN based methods eliminate the feature extraction process, and its deep network structure enables high performances. However, the limitations of the DNN approach are large dataset requirements, a large number of hyper-parameters tuning, expensive processing costs, and high-end system requirements. The computer-based facial emotion recognition is a big challenge for real-time applications. The HCI requires an intelligent system previously trained with publicly available datasets. Most FER systems use a laboratory-controlled dataset that suffers due to illumination variation, occlusion, low resolution and variation in facial expressions. Thus FER using the real-world datasets collected from the internet is problematic due to variance in face poses [11]. Facial expressions can reflect mental states and may express non-verbal communication to convey emotional information [12]. This paper uses several publicly available static image datasets for performance evaluation. The study presents a facial expression recognition framework that effectively uses the features extracted from the fully connected layers of the SqueezeNet pre-trained model. A series of experiments with a broad range of benchmark databases such as the Karolinska Directed Emotional Faces (KDEF) Database [13], the extended Cohn-Kanade dataset (CK+) [14], and the Facial Expression Recognition Challenge (FER2013) database [15], are evaluated and compared. This study also evaluates the Keimyung University Facial Expression of Drivers (KMU-FED) database consisting of images captured in an actual driving environment [16], [17].

Table I provides a comprehensive review of different datasets and approaches followed for facial emotion recognition. The second column represents the approaches and architecture used. The third column provides the FER datasets used to evaluate the performance accuracy. The main contributions of this paper are as follows:

1) Implementation of transfer learning based framework for driver facial emotion recognition using "pre-trained SqueezeNet" model.
2) Performance evaluation using a number of benchmark FER Database such as CK+, KDEF, and the wild dataset FER2013.
3) Performance evaluation using static in-vehicle driving image database, i.e., the KMU-FED.
4) Result analysis and comparisons with the state-of-the-art methods.

2022 National Conference on Communications (NCC) | 978-1-6654-5136-9/22/$31.00 ©2022 IEEE | DOI: 10.1109/NCC55593.2022.9806751

TABLE I: Study of Different Approaches Used for Facial Emotion Recognition.

Reference	Approach*	Database	Test Accuracy (%)
Y. Zhou and B. E. Shi [18]	Transfer learning AlexNet	KDEF	86.43
		RaFD	96.75
	Transfer learning AlexNet+FOS	KDEF	88.27
		RaFD	97.75
Sajjanhar *et al.* [19]	Transfer learning InceptionV3	CK+	76.52
	Transfer learning VGG19		86.2
	Transfer learning VGG-Face		91.37
Fei *et al.* [12]	Transfer Learning AlexNet	KDEF	87.8
		CK+	94.7
		FER2013	56.4
A. Krishnadas and S. Nithin [20]	Transfer learning VGG16	FER2013	57.3
	Machine learning SVM		34
	Deep learning CNN		57
Aggarwal *et al.* [21]	Transfer learning MobileNet	FER2013	70
	ML techniques SVM		51
	ML techniques KNN		41
Leone *et al.* [22]	Transfer Learning VGG-16	KMU-FED	94.27
Fei *et al.* [23]	Transfer Learning MobileNet and SVM	KDEF	88.7

*FOS: Face-Occupancy-based Feature Selection, CNN:Convolutional Neural Networks, KNN:K-Nearest Neighbor, SVM:Support Vector Machines.

5) Proposition is a FER system for driving safety assistance useful for in-vehicle embedded system applications.

The remaining parts of the article are organized as follows. Section II presents the problem definition and proposed methodology. Section III discusses the simulation results, and the conclusive remarks are provided in Section IV.

II. PROBLEM DEFINITION AND PROPOSED METHODOLOGY

Typically, distracted driving activities are performed on the driver's sitting posture. Currently, deep CNN frameworks are widely adopted for image-based approaches that operate activity detection tasks. Deeper architecture usually gives better performance; however, it requires a large training dataset and storage resources for efficient computation [24]. A work by H. Ma *et al.* [24], presented a facial expression recognition method based on a lightweight deep CNN model for classifying emotions with fewer parameters to deploy them on resource constrained devices, such as mobile devices. The Squeeze-and-Excitation (SE) module and the network slimming strategy of LA-Net reduce the computational cost and the number of parameters with better generalization ability and robustness, which motivates us to work on CNNs with smaller architectures. Our work uses a modified SqueezeNet 1.1 deep learning architecture to make it suitable for deployment in embedded devices for an intelligent FER framework set up to provide distracted driving assistance.

The system architecture that formulates the proposed work problem is shown in Fig. 1. The inputs to the system will be the images of the driver's frontal face captured using the in-vehicle webcam. The image pre-processing techniques such as image resizing, data augmentations, etc., are then applied to meet the input requirement of the proposed system. The system then extracts the in-depth features from the images for facial emotion classification. Happy, surprise, anger, sad,

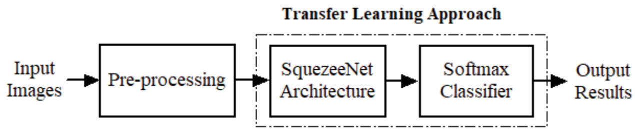


Fig. 1: System Overview of the Proposed Work.

fear, and disgust are the six primary emotions considered for recognizing facial expressions [8]. The transfer learning-based model reduces the computational complexity and can be implemented in embedded systems for real-time in-vehicle applications.

A. Transfer Learning for CNN Feature extraction

The main goal of this study is to investigate the performance of convolutional neural networks (CNN) in the facial expression recognition task. The bigger CNN architectures have limitations of communication overhead among servers while deployed in real-world scenarios of autonomous driving. Hence, it creates the scope to bring out a small CNN architecture, which have less communication overhead for frequent operations with distributed CNN training. One such model called SqueezeNet, a CNN with small architecture for deployment in embedded systems as a replacement for AlexNet. It has almost 50× fewer parameters than AlexNet and performs 3× faster [25]. The custom architectural design for embedded devices mainly uses the transfer learning technique with fewer parameters. However, in real-time complex environment each frame of the captured video requires preprocessing to meet proper driver emotion recognition requirements.

The transfer learning SqueezeNet is implemented with the last layer modification of the original SqueezeNet 1.1 architecture as shown in Fig. 2. The SqueezeNet design strategies is comprised mainly of fire modules, which are

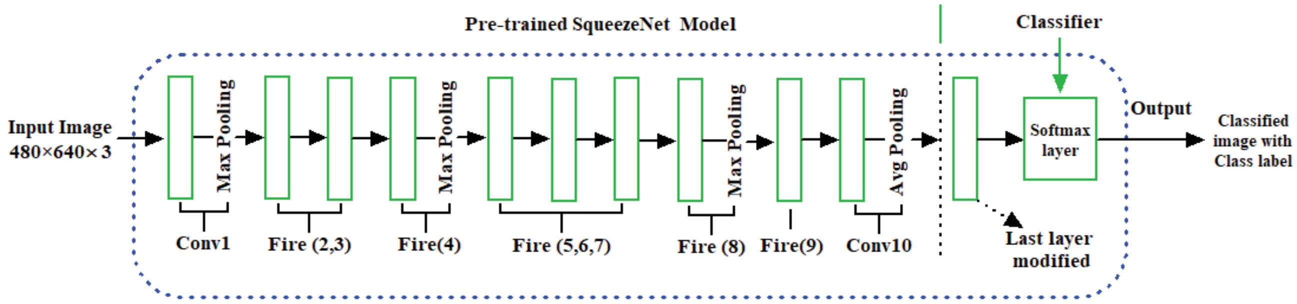


Fig. 2: Transfer learning SqueezeNet Model.

the building blocks out of which CNN architectures are built. The architecture begins with a convolution layer (conv1), followed by 8 fire modules (fire2-9), ending with a final convolution layer (conv10). A fire module comprises a squeeze convolution layer (which has only 1×1 filters), feeding into an expand layer that has a mix of 1×1 and 3×3 convolution filters. The last layer modification is used to classify all six primary facial emotions classification from the publicly available FER datasets. Transfer learning model reduces time compared to training from scratch and enables fast prediction as required for real-time applications.

B. Methodology

The FER process for an in-vehicle application involves facial image acquisition, image pre-processing, feature extraction using the pre-trained SqueezeNet CNN model, and classification. The stepwise approach of FER methodology is presented in Algorithm 1. The workflow diagram is presented in Fig. 3.

Algorithm 1 : Facial Emotion Recognition Algorithm using Transfer Learning SqueezeNet CNN

1: Load the FER image dataset.
2: Apply the image pre-processing techniques.
3: Split the dataset into training and testing samples.
4: Build the model using pre-trained SqueezeNet network parameters.
5: Feature extraction from the dense layer of the model.
6: Test image data to the model for emotion classification.
7: Calculate the performance measurement parameters.

1) Image Acquisition: Driver facial image acquisition is the first step in the FER system development. An in-vehicle dashboard camera or webcam will perform the data acquisition tasks. These images are then stored in a designated folder for each topic of the driver's facial expression to make the database. The real-time FER implementation needs clear and quality datasets for accurate recognition. However, in-vehicle implementation is still challenging due to environmental condition variations. We have used the publicly available datasets as our image database for experimental work.

2) Image Pre-processing: The image databases contain images of different orientations, imbalanced datasets, and different sizes. The images are then resized to feed the input data size of $480 \times 640 \times 3$ pixels for the SqueezeNet CNN model framework. Preprocessing techniques such as adaptive histogram equalization and image-inpainting are used to remove the darkness and automatically repair the damaged area by image interpolation.

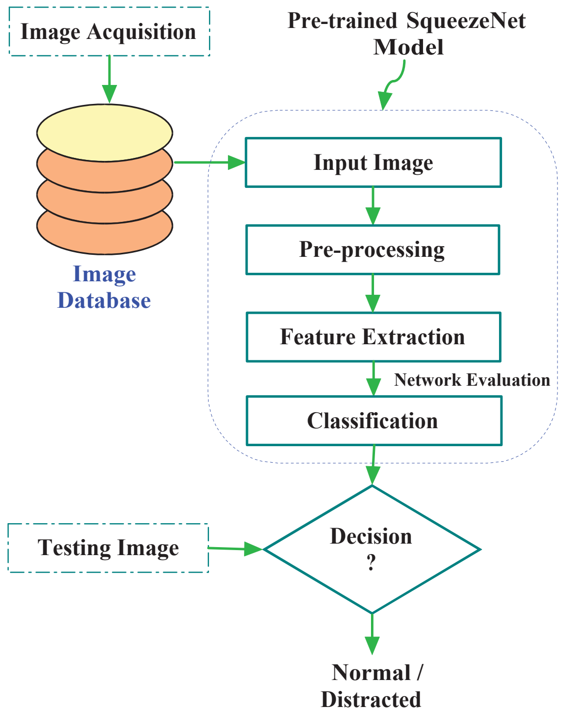


Fig. 3: SqueezeNet Transfer Learning Framework.

3) Loading and Training Summary of the Model: The training step of system modelling follows a 80-20% data split where 80% of the data were used for training, and the remaining 20% were used for testing. The system loads the pre-trained SqueezeNet model and trains the network using batch gradient descent with batch size of 64 for 55 epochs.

4) Feature Extraction: Features extraction method extracts the feature set from the acquired data. The last dense layer of the CNN network is trained to extract features to find a set of vectors that efficiently represent facial expressions.

5) Classification: The process of identifying a set of categories which belongs to a new observation based on a training set is the classification task. The CNN classifier uses the input images to classify the seven facial emotion classes. All other classes of FER except the normal class can be treated as distracted driving activity. The prediction of FER work can be formulated into normal and distracted based on driver facial expressions.

III. Results and Discussion

All experiments were conducted on the Google Colab platform. Python and PyTorch-fastai libraries are used to build the deep learning model to evaluate the FER and classification task. The performance of the pre-trained SqueezeNet CNN transfer learning system is assessed based on the performance metrics, i.e., recognition accuracy, error rate, and ROC-AUC curve.

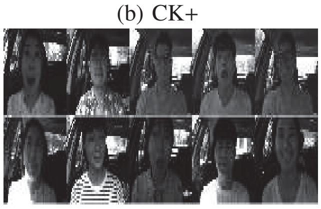
A. Databases

Table II summarizes the datasets used in this work. Many benchmark FER databases are publicly available and mainly captured under controlled conditions in an indoor laboratory environment. FER is challenging to implement a model trained with laboratory datasets to be used in the actual driving environment. One dataset that uses in-vehicle settings is the Keimyung university facial expression of drivers (KMU-FED) and helps to prepare the model for real-time outdoor application [17]. Some samples of the datasets are shown in Fig. 4. The KDEF Dataset contains 4900 facial

(a) KDEF

(b) CK+

(c) FER2013



(d) KMU-FED

Fig. 4: Sample Images of Different Databases.

expression images in total, from which all front view images of all seven categories are 980 images [13]. The KDEF dataset is a relatively balanced sample and training and validation datasets are selected randomly. The CK+ Dataset is large in number and has many similar images of grayscale with seven facial expressions [14]. The FER-2013 grayscale dataset images are quite small, unbalanced and one of the largest wild dataset used in this work. As per our knowledge, the only in-vehicle dataset is the KMU-FED dataset which has six basic expressions of 1101 images.

B. Experimental Results and Analysis

This experimental study evaluates the transfer learning SqueezeNet network for the FER classification task. The test accuracy performance measurement is shown in Table III for 55 epochs with mini-batch size of 64. From the training performance, the highest accuracy is found to be 95.83% for the KMU-FED dataset. The performance accuracy for the CK+ dataset, KDEF, and FER2013 are found to be 91.39%, 86.86% and 61.09% respectively.

The experimental results and performance analysis will be discussed in following sections with the training performances curves. The model training process shows the performance with model loss parameter, error rate, accuracy, and the ROC-AUC curve. The proposed transfer learning SqueezeNet CNN model training accuracy and loss results can be observed from the training performance curves presented for the best model performances with KMU-FED dataset. Figures 5 to 7 presents the best model performance results.

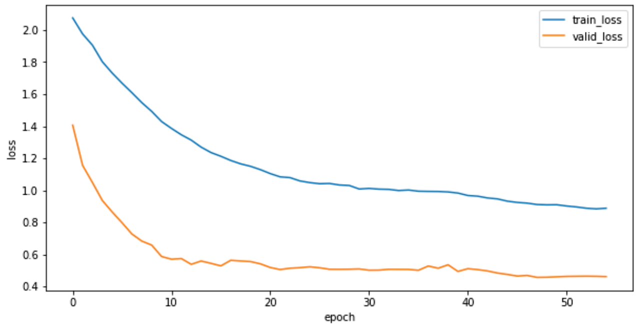


Fig. 5: Learning Performance on Training.

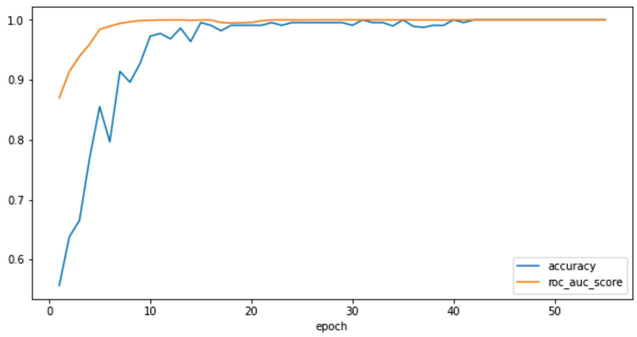


Fig. 6: The AUC-ROC Curve.

Fig. 5 shows the model training and validation loss performance. It shows a good fit as the training and validation loss decreases with a minimal gap between the two final loss values. The models average accuracy performance reaches to 95.83% and can be seen from Fig. 6. The receiver operator characteristic (ROC) metric gives the probability curve between the true positive rate (TPR) and false positive rate (FPR). Summary of ROC metric is represented as area under the curve (AUC) to classify normal and distracted driving due to emotion variations. The average ROC-AUC performance achieved to be 99.31% and can be visualized from Fig. 6. The models performance is evaluated on 55 epochs with the 80-20% data splits. The error performance

TABLE II: Detailed of FER Datasets Studied in the Experiment.

Datasets	Images	Identities	Collection Condition	Expression Distribution	Images Size	Images Type
KDEF [13]	4900	70	Lab	Seven Basic Expressions	562×762	JPG
CK+ [14]	593	123	Lab	Seven Basic Expressions	48×48	PNG
FER2013 [15]	35887	N/A	Web	Seven Basic Expressions	48×48	JPG
KMU-FED [17]	1101	12	In-vehicle Enviornment	Six Basic Expressions	1600×1200	JPG

TABLE III: Performance Comparison in Terms of Accuracy.

Methodology Used	Database			
	KDEF	**CK+**	**FER2013**	**KMU-FED**
Transfer Learning (SqueezeNet)	86.86%	91.39%	61.09%	95.83%

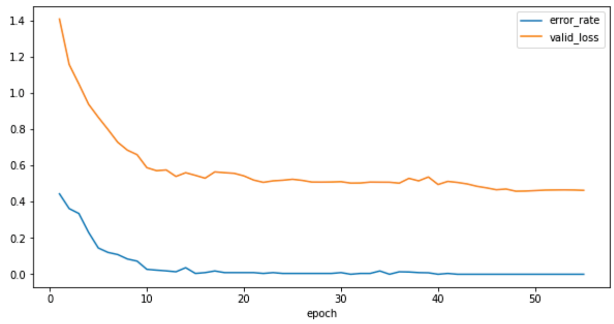


Fig. 7: The Training Error Rate.

curve with test loss is shown in Fig. 7 to judge over-fitting, and the good fit between the curves can be observed.

C. K-fold Cross Validation

The cross-validation (CV) evaluation process divides the data by K-equal or almost equal folds. Here stratified k-fold cross-validation is used. Model accuracy is obtained by finding the average accuracy of each iteration. The accuracy performance of the FER model has been evaluated for 5-fold, 7-fold and 10-fold for 30 epochs on the publicly available KMU-FED dataset. The mean accuracy of K-fold CV is compared as shown in Fig. 8. The bar graphs show that the predicted highest performance accuracy is 83.43%.

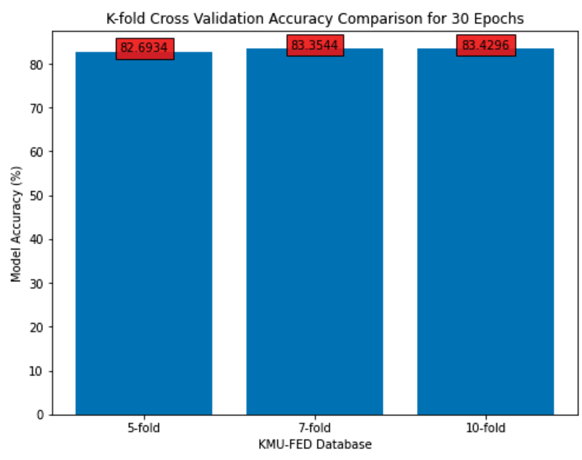


Fig. 8: K-Fold CV Performance Using KMU-FED Dataset.

D. Comparative Analysis of Results

The facial emotion recognition methodologies have been studied in the literature, and a comparative summary of the various model's performance accuracy is presented in Table I. The transfer learning CNN models are focused in this work for emotion recognition. The state-of-the-art technologies applied with available benchmark datasets are presented for performance comparisons. Table IV presents the performance comparison results based on transfer learning approaches.

TABLE IV: Performance Comparison of Transfer Learning Based Facial Emotion Recognition.

Reference	Model Used	Database Used	Test Accuracy (%)
Sajjanhar *et al.* [19]	InceptionV3	CK+	76.52
	VGG19	CK+	86.2
	VGG-Face	CK+	91.37
Fei *et al.* [12]	AlexNet	CK+	94.7
Proposed	Modified SqueezeNet	CK+	91.39
Y. Zhou and B. E. Shi [18]	AlexNet	KDEF	86.43
	AlexNet+FOS	KDEF	88.27
Fei *et al.* [12]	AlexNet	KDEF	87.8
Fei *et al.* [23]	MobileNet and SVM	KDEF	88.7
Proposed	Modified SqueezeNet	KDEF	86.86
Fei *et al.* [12]	AlexNet	FER2013	56.4
A. Krishnadas and S. Nithin [20]	VGG16	FER2013	57.3
Aggarwal *et al.* [21]	MobileNet	FER2013	70
Proposed	Modified SqueezeNet	FER2013	61.09
Leone *et al.* [22]	VGG-16	KMU-FED	94.27
Proposed	**Modified SqueezeNet**	**KMU-FED**	**95.83**

As per Table IV, Sajjanhar *et al.* [19] used pre-trained deep CNN models for facial expression recognition using publicly available face databases. The FER accuracy reported were 76.52%, 86.2%, and 91.37% based on the pre-trained model InceptionV3, VGG19, and VGG-Face, respectively. The performance of the proposed work on the pre-trained SqueezeNet network shows the improved result on the CK+ dataset compared to [19], whereas our proposed model's performance result is comparable to the accuracy achieved by Fei *et al.* [12]. In work by Y. Zhou and B. E. Shi [18], used an automatic selection schemes on pre-trained AlexNet CNN for facial expression classification using KDEF database. The performance accuracy reported were 86.43% and 88.27% for transfer learnig AlexNet and AlexNet+FOS model respectively. Similarly, work by Fei *et al.* [12] uses transfer learning AlexNet CNN architecture improves the performance accuracy to 87.8% when evaluated on KDEF dataset. The work by Fei *et al.* [23], uses transfer learning MobileNet and SVM technique for FER and the highest performance accuracy reported was 88.7%. However our proposed model on KDEF datasets performs well in achieving a comparable accuracy of 86.86%.

In work by Aggarwal *et al.* [21], an attempt was made to compare the performances of machine learning and deep learning applied to human facial emotion recognition. Use transfer learning MobileNet network shows performance accuracy of 70% with the most challenging and imbalanced

FER2013 dataset; however, the machine learning model performs very poorly. The transfer learning performances on other FER datasets like CK+, KDEF, and KMU-FED are not available. Another work by A. Krishnadas and S. Nithin [20] classifies the emotional state of the driving behavior using machine learning and deep learning techniques on the FER2013 wild dataset. The performance accuracy achieved using the transfer learning VGG16 model was 57.3%, whereas deep learning CNN and machine learning SVM techniques achieved 57% and 34%, respectively. Similarly, another work by Fei *et al.* [12] uses transfer learning AlexNet CNN and LDA architecture for facial emotion recognition. The overall recognition accuracy performance presented was 56.4%. However, our proposed work on the FER2013 dataset shows improved performance as compared to [12], [20] and performs poorly compared to [21].

The works presented by M. Jeong and B. Chul Ko [16], Jeong *et al.* [17] and Leone *et al.* in [22] used the real driving scenario KMU-FED dataset for FER recognition. The techniques WRF, LMRF, and Transfer learning VGG19 were used, and the performance accuracy reported were 94.7%, 95.1%, and 94.27%, respectively. The performance of the proposed work on the KMU-FED dataset using the transfer learning SqueezeNet network shows the best result of 95.83% recognition accuracy. The proposed model is evaluated on four benchmark databases, and the obtained performance accuracy can be generalized to facial emotion recognition of a human driver.

IV. Conclusions

This paper implements a SqueezeNet transfer learning framework for facial emotion recognition to extract features for emotion classification. The performance is evaluated using different datasets with various real-time in-vehicle challenges such as illumination variation, imbalance datasets for different classes, emotions sharing the same expression, etc. Publicly available datasets such as CK+, KDEFE, FER2013 and KMU-FED were used for the classification task in this work. The performance of the proposed work is compared with state-of-the-art models. The SqueezeNet transfer learning model only achieved superior performance on the KMU-FED dataset, i.e., maximum accuracy of 95.83%, and showed comparable performance to the rest of the benchmark database. A camera-based embedded FER system deployment would be helpful to capture images of driver's faces and process them to predict emotions and assist drivers with their health care.

References

[1] C. Bisogni, A. Castiglione, S. Hossain, F. Narducci, and S. Umer, "Impact of deep learning approaches on facial expression recognition in healthcare industries," *IEEE Trans. Indust. Informatics*, 2022.

[2] H. Gjerde, P. T. Normann, A. S. Christophersen, S. O. Samuelsen, and J. Mørland, "Alcohol, psychoactive drugs and fatal road traffic accidents in norway: a case–control study," *Accident Anal. Prevention*, vol. 43, no. 3, pp. 1197–1203, 2011.

[3] A. Das, H. Gjerde, S. S. Gopalan, and P. T. Normann, "Alcohol, drugs, and road traffic crashes in india: a systematic review," *Traffic Injury Preven.*, vol. 13, no. 6, pp. 544–553, 2012.

[4] Lacey *et al.*, "2007 national roadside survey of alcohol and drug use by drivers: drug results." United States. National Highway Traffic Safety Administration, Tech. Rep., 2009.

[5] J. G. Ramaekers, G. Berghaus, M. van Laar, and O. H. Drummer, "Dose related risk of motor vehicle crashes after cannabis use," *Drug and alcohol dependence*, vol. 73, no. 2, pp. 109–119, 2004.

[6] C. Marina Martinez, M. Heucke, F. Wang, B. Gao, and D. Cao, "Driving style recognition for intelligent vehicle control and advanced driver assistance: A survey," *IEEE Trans. Intell. Transp. Syst.*, vol. 19, no. 3, pp. 666–676, 2018.

[7] C. Darwin, *The expression of the emotions in man and animals by Charles Darwin*. John Murray, 1872.

[8] P. Ekman, "An argument for basic emotions," *Cognition & emotion*, vol. 6, no. 3-4, pp. 169–200, 1992.

[9] M. Valstar, M. Pantic *et al.*, "Induced disgust, happiness and surprise: an addition to the MMI facial expression database," in *Proc. 3rd Int. Workshop EMOTION (satellite of LREC): Corpora Res. Emotion Affect*. Paris, France., 2010, p. 65.

[10] D. Keltner and P. Ekman, "Emotion: An overview." 2000.

[11] J. Shao and Y. Qian, "Three convolutional neural network models for facial expression recognition in the wild," *Neurocomputing*, vol. 355, pp. 82–92, 2019.

[12] 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.

[13] D. Lundqvist, A. Flykt, and A. Öhman, "The Karolinska directed emotional faces (KDEF)," *CD ROM Depart. Clinical Neuroscience, Psychology Sec., Karolinska Inst.*, vol. 91, no. 630, pp. 2–2, 1998.

[14] P. Lucey, J. F. Cohn, T. Kanade, J. Saragih, Z. Ambadar, and I. Matthews, "The Extended Cohn-Kanade Dataset (CK+): A complete dataset for action unit and emotion-specified expression," in *Proc. IEEE Comput. Society conf. Comput. Vis. Pattern Recogn. Works.*, 2010, pp. 94–101.

[15] *Facial Expression Recognition 2013 Dataset (FER2013).*, [Online]. Available: https://www.kaggle.com/c/challenges-in-representation-learning-facial-expression-recognition-challenge.

[16] M. Jeong and B. C. Ko, "Drivers facial expression recognition in real-time for safe driving," *Sensors*, vol. 18, no. 12, p. 4270, 2018.

[17] M. Jeong, J. Nam, and B. C. Ko, "Lightweight multilayer random forests for monitoring driver emotional status," *IEEE Access*, vol. 8, pp. 60 344–60 354, 2020.

[18] Y. Zhou and B. E. Shi, "Action unit selective feature maps in deep networks for facial expression recognition," in *Proc. IEEE Int. Joint Conf. Neural Netw. (IJCNN)*, 2017, pp. 2031–2038.

[19] A. Sajjanhar, Z. Wu, and Q. Wen, "Deep learning models for facial expression recognition," in *Proc. IEEE Digital Imag. Comput. Techniques Appl. (DICTA)*, 2018, pp. 1–6.

[20] A. Krishnadas and S. Nithin, "A comparative study of machine learning and deep learning algorithms for recognizing facial emotions," in *Proc. IEEE 2nd Int. Conf. Electron. Sustainable Commun. Syst. (ICESC)*, 2021, pp. 1506–1512.

[21] A. Aggarwal, S. Garg, R. Madaan, and R. Kumar, "Comparison of different machine learning and deep learning emotion detection models," in *Intell. Comput. Commun. Syst.* Springer, 2021, pp. 401–408.

[22] A. Leone, A. Caroppo, A. Manni, and P. Siciliano, "Vision-based road rage detection framework in automotive safety applications," *Sensors*, vol. 21, no. 9, p. 2942, 2021.

[23] Z. Fei, E. Yang, L. Yu, X. Li, H. Zhou, and W. Zhou, "A novel deep neural network-based emotion analysis system for automatic detection of mild cognitive impairment in the elderly," *Neurocomputing*, vol. 468, pp. 306–316, 2022.

[24] H. Ma, T. Celik, and H.-C. Li, "Lightweight attention convolutional neural network through network slimming for robust facial expression recognition," *Signal, Imag. Video Process.*, pp. 1–9, 2021.

[25] F. N. Iandola, S. Han, M. W. Moskewicz, K. Ashraf, W. J. Dally, and K. Keutzer, "SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and < 0.5 MB model size," *arXiv preprint arXiv:1602.07360*, 2016.