IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 13 March 2023, accepted 28 March 2023, date of publication 3 April 2023, date of current version 18 April 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3264268

RESEARCH ARTICLE

A Lightweight Facial Emotion Recognition System Using Partial Transfer Learning for Visually Impaired People

DINA SHEHADA[1,2], (Member, IEEE), AYAD TURKY[3], WASIQ KHAN[4], (Senior Member, IEEE), BILAL KHAN[5], (Senior Member, IEEE), AND ABIR HUSSAIN[1,4], (Senior Member, IEEE)

[1]Department of Electrical Engineering, University of Sharjah, Sharjah, United Arab Emirates
[2]Department of Engineering and IT, University of Dubai, Dubai, United Arab Emirates
[3]Department of Computer Science, University of Sharjah, Sharjah, United Arab Emirates
[4]School of Computer Science and Mathematics, Liverpool John Moores University, L3 3AF Liverpool, U.K.
[5]School of Computer Science and Engineering, California State University at San Bernardino, San Bernardino, CA 92407, USA

Corresponding author: Dina Shehada (dinashehada@ieee.org)

ABSTRACT The inability to perceive visual and other non-verbal cues for individuals with visual impairment can pose a significant challenge for their correct conversational interactions and can be an impediment for various daily life activities. Recent advancements in computational resources, particularly the computer vision capabilities can be utilized to design effective applications for visually impaired people (VIP).Among various assistive technologies, automated facial impression recognition with real-time accurate interpretation can be proven useful to tackle the above problem. Using such approach, facial emotions (e.g., sad, happy) can be robustly recognized and conveyed to the associated individuals. In this paper, a partial transfer learning approach is adopted utilizing a custom trained Convolutional Neural Network (CNN) for facial emotion recognition. A novel model that transfers features from one dataset to another is proposed. This model enables the transfer of features learned from a small number of instances to solve new challenging instances. Using the proposed approach based on a newly trained CNN, a portable lightweight facial expression recognition system with wireless connectivity and high detection accuracy was constructed and targeted specifically for VIP. The proposed recognition model provides a notable improvement over the current state-of-the-art, by providing the highest recognition accuracy of 82.1% on the enhanced Facial Expression Recognition 2013 (FER2013) dataset. Moreover, with only 1.49M parameters, the model is operable on edge devices with limited memory and processing power. Overall, three labeled emotions happy, sad, surprise were recognized by the model with high accuracy whereas a relatively lower accuracy rate for anger, disgust, fear was noticed with higher misclassification labels for sad.

INDEX TERMS Partial transfer learning, deep learning, visually impaired people, convolutional neural network.

I. INTRODUCTION

Vision plays a crucial role in comprehending our surroundings, but loss of vision can hinder a person's ability to live a typical life. The World Health Organization (WHO) reports that 285 million people globally have a visual impairment, with 39 million being blind and 246 million experiencing low vision [6]. People suffering from visual impairment or eye conditions require support to overcome daily tasks, such as emotional, navigating, and exploring new surroundings. The term visually impaired people refers to people suffering from any kind of vision loss that can range from partial to complete loss of sight. According to a study conducted in 2020 [5], the total number of visually impaired people (VIP) around the world is 1.1 Billion [34]. This number is increasing every year as shown in Figure 1:

The associate editor coordinating the review of this manuscript and approving it for publication was Yiming Tang.

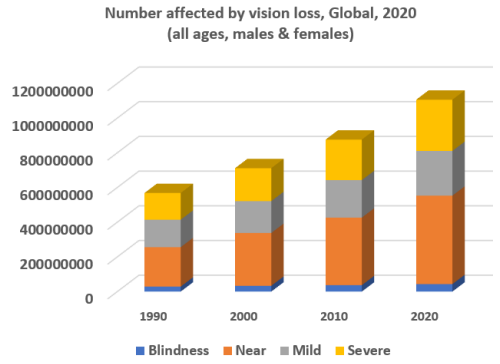


FIGURE 1. Number of people affected by visual impairment across the world [34].

Emotions are mental state that is expressed through the facial expressions of a person. The face is considered the prime perceptual stimulus [32]. The facial muscles have the ability to form more than 40 expressions. However, according to physiologists, emotions evoke a small number of basic expressions namely: joy, sadness, fear, anger, disgust, surprise, interest, and contempt [9]. These emotions are naturally expressed by humans and are taken as a common reference. These facial expressions are a form of nonverbal communication that can support or even replace verbal communication [27].

VIP face many challenges every day. Exploring unfamiliar environments is one of the biggest challenges; navigating safely from one place to another can be quite a difficult task. Particularly the relevant outdoor infrastructures (e.g., roads, footpath etc.) do not consider VIP in common practice. Likewise, visual impairment poses a significant social challenge such as the inability to perceive visual and other non-verbal cues, which significantly affects their quality of life (QoL) in terms of social aspects and engagements. To aid the social interactions for VIP, technology along with machine intelligence can play a significant role. An intelligent Facial Emotion Recognition (FER) model can be implemented to recognize the correct emotions of the person (such as happiness, or sadness), and convey this information to the VIP. Considering this, we propose an FER-based systems utilizing machine learning and computer vision techniques to identify the correct facial emotions.

In this paper, we propose Partial Transfer Learning based Convolutional Neural Network (CNN) for facial emotion recognition. This paper presents a novel model for transferring features from one dataset to another. This model enables the transfer of features learned from a small number of instances to solve new challenges instances. The proposed system is useful for the VIP to address the limitations of the existing systems, by providing a portable, lightweight, and accurate assistive system.

The paper contribution can be summarized as follows:

- Propose a wireless, portable solution for VIP.
- Propose A Partial Transfer Learning based CNN for facial emotion recognition.
- Different from the current systems which utilize computation complex solutions, the proposed solution uses lightweight computations and real-time detection.
- The model's generalization ability and robustness were verified against multiple datasets. A notable performance improvement over the current state-of-the-art models, achieving the highest recognition accuracy of 82.1% on the Facial Expression Recognition 2013 (FER2013) FER2013 dataset.

The remainder of the paper is organized as follows. In Section II, related works are reviewed. Section III, explains our proposed methodology and design. Implementation results and discussion are provided in Section IV. Section V concludes with the paper's findings and discusses future work.

II. RELATED WORKS

In recent years, some studies have been performed on facial expression recognition. However, facial expression recognition still faces great challenges. In this section, we discuss the most recent machine learning techniques introduced for automatic recognition of facial expressions.

A deep learning-based model for FER is proposed in [19]. The model is based on CNN. It classifies a person's face image into seven different emotions including sad, fear, happy, anger, neutral, disgust, and surprise. The model was trained and tested over FER2013 dataset [10] with 35,685 grayscale images, where 80% of the proposed dataset was used for training and the remainder was used for testing. Random Search algorithm was used to optimize the hyper-parameters of the CNN. This model achieved an accuracy of 66.7%. The recognized emotions are conveyed in text and audio formats.

Vulpe-Grigorasi & Grigore [35] proposed another CNN based FER system with parameter optimization. This model is tested over the FER2013 dataset, and achieved an accuracy of 72.16 % and a loss value of 0.97 when trained with a learning rate of 0.001, 128 batch size for 750 epochs. The model has 517,3959 parameters and a size of 59 MB.

Minaee et al. [26] proposed an attention CNN based FER model. The model focuses attention on specific parts of the face that are believed to have a higher impact on the classification such as the eyes and mouth. Spatial transformer [15] is used to extract the parameters aggregated with features extracted by the CNN layer and passed to the dense layer. The model utilized 28,709, 3500, and 3589 images for training, validating and testing, respectively, using the FER2013 dataset. Classification accuracy was found to be 70.02 % on the testing set. For the extended Cohn–Kanade (CK+) dataset [22], the authors used 420, 60, and 113 images for training, validation, and testing respectively. The achieved accuracy was 98.0%. The model was also tested on FERG [2] and JAFFE [24] datasets and achieved an accuracy of 99.3% and 92.8%, respectively.

Another FER system for enhancing online teaching is proposed in [13]. The model aims to identify the facial expressions of students to evaluate their concentration in class.

A real-time video is taken of the students then Multi-Task Cascaded Convolutional Neural Networks (MTCNN) algorithm is used to extract the frame with the student's face. Image is passed to the model combining VGG16 [14] and ECANet [36]. The VGG16 structure is utilized for feature extraction. The extracted features are passed to the ECANet network. Two feature maps are concatenated and input into the classification model. The model was trained and tested on different datasets. It achieved an accuracy of 67.40% over FER2013 dataset and 99.18% on CK+ dataset, with a learning rate of 0.0001, batch size was 128, and 300 epochs. Similarly, the work in [31] also proposes a similar approach but based on Multiple Branch Cross-Connected Convolutional Neural Network (MBCC-CNN), and achieved an accuracy of 71.5% over FER2013, 98.48% on CK+, 88.1% on FER+ [4] and 87.34% on RAF [20], [21] datasets.

A lightweight model (DenseNet) is proposed in [39] for emotion recognition. The CNN based model detects the face using histograms of gradients (HOG) [40]. DenseNet reduces the number of parameters to be trained. The accuracy of the model was 71.73%, tested over FER2013 and trained for 250 epochs.

Another lightweight approach is proposed in [17] for real-time expression detection. The model also uses CNN architecture along with log level threshold quantization (LLTQ) method to reduce the number of operations and overhead. The model size is 0.39 MB, and the number of operations is about 28M integer operations (IOPs). Although the model achieved a considerably high testing accuracy of 86.5% on FER+ dataset, it consumes high power [37].

Saurav et al. [29] integrate two CNN models to create a dual feature extraction model with 1.08M parameters. The model is lightweight and suitable to embedded systems. On the FER2013 dataset the model achieved an accuracy of 72.77%, while on CK+ dataset the accuracy was 98.54%.

EmNet [30] is another FER system, made of two separate CNN models to predict emotions. The results are combined through fusion techniques. The achieved accuracy was 74.1% on FER2013 dataset with 4.81M parameters and 19.3MB model size. The model was also tested on RAF and SFEW [1] and achieved an accuracy of 84% and 53% on these datasets, respectively.

A lightweight emotion recognition (LER) system is proposed in [38]. The model incorporates compression techniques into the connected dense layer to eliminate redundant parameters. Three different models, DenseNet-1, DenseNet-2, and DenseNet-3 are proposed. The models were trained and tested over FER2013, and FER+. DenseNet-2 with 218,839 parameters, achieved the highest accuracy of 71.55% and 85.68% on FER2013 and FER+, respectively. A new dataset FERFIN is also created as an enhanced version of FER2013 dataset with less noise and corrected labels. The model achieved an accuracy of 85.89% on FIRFIN dataset.

Burrows et al. [7] propose three different CNN and Generative Adversarial Networks (GANs) based models that classify facial expressions. The authors created a combined dataset from six different datasets to have around 41K images used for testing and training. The first model classifies the images into seven emotions; sad, fear, happy, anger, neutral, disgust, and surprise. It achieved an accuracy of 58.71%. The second model classifies into three classes namely, negative, neutral, and positive, and achieved an accuracy of 73.71%. The third model classifies into negative and positive and achieved an accuracy of 77.83%. The three models have about 1.5M parameters and were trained for 70 epochs.

A deep learning FER system for the blind is proposed in [16]. The CNN based architecture classifies the facial image into seven classes of emotions. The model was also tested on FER2013 dataset and achieved an accuracy of 67.18%, with 150 training epochs. The model is incorporated into an android application that captures the image, classifies it, and reads the predicted class label to the user.

A FER model for VIP is provided in [23]. The model classifies emotions into three categories; positive, neutral, and negative. It incorporates ResNet model [12] for feature extraction. Extracted features are combined through Gated Recurrent Network (GRU) which is a type of Recurrent Neural Networks (RNNs). Multi-layer Perceptron (MLP) classification system is used to classify the emotion. The model was tested on CK+ dataset and achieved an accuracy of 87%. The developed tool displays the probability of each category with an Emoji that represents the predicted emotion. The authors indicated that they are planning to use three-stage signal with a Braille display [18].

Another FER system for VIP is proposed in [3]. The system uses Support Vector Machine (SVM) to classify emotions into three main categories; sad, happy, and surprise. The model was trained on JAFFE dataset combined with some newly added images. The system is incorporated into a desktop application that conveys the classified emotion in audio. However, some important details about the model accuracy, error, and sample size are not provided in the paper [33].

Many generic FER systems are proposed in the literature but very few are proposed for VIP. Generic systems are useful in the sense that they can be utilized in various applications. However, such solutions may be unsuitable for VIP. Most of the current FER systems proposed in the literature were implemented on desktop [3], [7], [13], [17], [19], [23], [30], [31], or other powerful devices [29]. Furthermore, some of them produce a huge overhead [13], [19], [26], [31], [35]. This might ensure good performance but will not suit the nature of use of our targeted VIP users. With VIP, system portability and ease of use are very crucial. Moreover, VIP assistive solutions need to be lightweight and operable on edge devices with limited storage and power. All of these limitations make the existing systems not convenient to the VIP. In the next section, a FER system designed for VIP is proposed. The proposed system addresses all the limitations aforementioned in the related works.

III. PROPOSED METHODOLOGY

A. PROPOSED SYSTEM DESIGN

In the present work, a FER system based on Partial Transfer Learning and CNN is proposed to recognize facial emotions with higher accuracy. It improves the social interaction activities for VIP through real-time labeling and audio interpretation of various visual expressions. As illustrated in Figure 2, the proposed model can be utilized in a wearable device with a camera and a lightweight set of packages (Raspberry Pi) designed to operate in real-time.

An embedded portable camera is mounted as part of the recognition system that acquires images from a conversational interaction which is then processed to recognize facial emotions in real-time. Subsequently, the recognized emotion is conveyed via an audio device. Audio speech is fed to the VIP via a typical earphone. A low cost and lightweight embedded device with Raspberry Pi was used as the main processing unit that has proven useful for various real-world applications where performance in real-time scenarios can be challenging. The design of FER system followed the workflow shown in Figure 3 with a high level overview provided in Figure 3. From the captured video, frames are first extracted and processed through several steps including grayscale transformation, face detection, cropping, image resizing, and normalization. Subsequently, emotion recognition is performed on the processed image via the proposed CNN model. The recognized emotion is then conveyed in audio format. Emotion classification was conducted based on six classes: anger, disgust, fear, happiness, sadness, and surprise. Figure 4 provides a high level overview of the proposed FER system.

Algorithm 1, 2, and 3, describe the detailed steps of the proposed system.

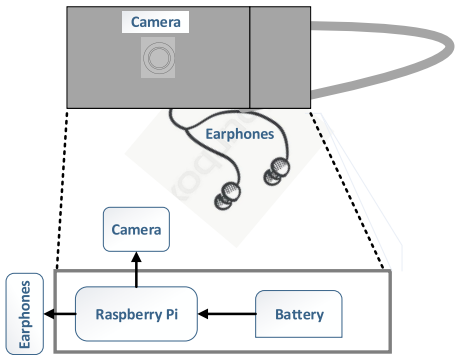


FIGURE 2. The design of the proposed FER system.

B. PROPOSED MODEL ARCHITECTURE

A custom CNN model was constructed and trained in the proposed FER system with the internal layer structure given in Figure 5. As described in Figure 5, the proposed CNN model consisted of four convolutional, three max-pooling, five dropout, a flatten, and three fully connected (dense) layers with a total of 1,486,854 (1.49M) trainable parameters

Algorithm 1 Image Prepossessing

Let S be a set of captured images from video F
Let $F \subset S$, where:
$F = \{\forall S \subset I_{NxM}\}$
Let ET be a set of emotion images where:
$ET = \{happy, sad, angry, surprised, fear, digust\}$
Let $SC \subset F$, where:
$SC = \{\forall\ sc \in F\}$ & sc is a cropped image
Let N to be a set of normalized images
$N = \{\forall\ sc \in SC \exists n, m = \frac{sc}{255}\}$

Algorithm 2 Network Training

Define CNN to be a deep learning model
Let M be the set of metrics used to evaluate the model
$M = \{Accuracy, Recall, F1score, Precicion\}$
Let $training$ & $testing$ be set of images
$training$ & $testing \subset EFER2013$ where:
$training = \{n \in N, \&size(training) = 70\%\ of\ EFER2013\}$
$testing = \{\forall n \in testing, n \in N\ \&\ n \notin training\}$
$et = CNN(training)$
$(m, et) = CNN(testing)$ where $m \in M$

of a size of only 0.24 MB. The above lightweight system was focused to ensure its compatibility with Raspberry Pi processing unit with limited storage and computational power.

IV. IMPLEMENTATION RESULTS

A. DATASETS

Model training and validation were accomplished using two popular facial expression recognition datasets, FER2013 [10], and CK+ [22]. Model validation was then followed by tests on unseen real subjects from a custom developed dataset of 66 images of facial images expressing 6 different emotions.

The FER2013 dataset, constructed using Google image search, contains a total of 35,887 facial expression images with a resolution of 48 × 48 pixels. FER2013 is a diverse dataset of images with faces representing seven different emotions happy, angry, sad, fear, surprise, disgust, neutral as well as non-facial and text images. Additionally, there are images with noisy input (sleepy faces) and missing labels [25], [28]. Furthermore, FER2013 exhibits greater diversity including images with facial occlusion, partial faces, and faces with eyeglasses. However, FER2013 suffers from imbalanced classes with the following distribution happy: 25%, sad: 16.9%, angry: 13.8%, surprised: 11.2%, fear: 14.3%, disgust: 1.5%. To overcome class imbalance issue with additional mislabeled images, an enhanced version

Algorithm 3 Transfer Learning

Let $CNN_{trainied}$ be a deep learning model that was trained on EFER2013
Define $T = \{t \in N, \&\ N \subset CK+\}$
$(m, et) = CNN_{trained}(T)$ where $m \in M$

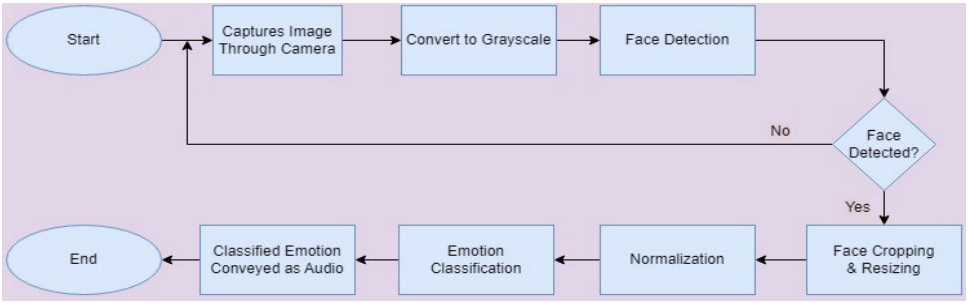


FIGURE 3. Flowchart of the proposed FER system.

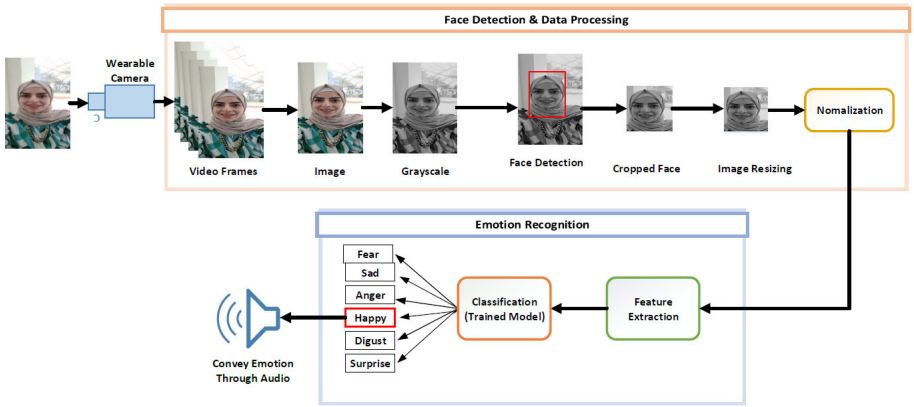


FIGURE 4. Overview of the proposed FER system.

of FER2013 (EFER2013) [25] dataset was used which an updated balanced version of FER2013 after cleaning bad images and generating new images using CycleGAN for the minority class.

As the proposed model aims to classify the facial image into six emotions, anger, fear, disgust, happy, sad and surprise, the neutral class samples were omitted and 24,839 samples were used.

The newly available EFER2013 dataset was used to train and validate the proposed CNN. The second dataset analyzed in the proposed study (CK+) consists of 327 recorded videos of 123 subjects, each labeled with one of seven expression classes anger, contempt, disgust, fear, happy, sad, surprise. A total of 981 facial expression images were extracted from the above videos, each with a resolution of 48 × 48 pixels.

In CK+ dataset, 327 videos recorded for different 123 subjects were labeled with one of seven expression classes: anger, contempt, disgust, fear, happy, sad, and surprise. 981 facial expression images with a resolution of 48 × 48 pixels were extracted from these videos and were considered in this study. It is noted that the CK+ is comparatively a smaller dataset with only 981 images of 7 classes of expressions, thus posing a challenge to train a robust recognition model. Therefore, images from CK+ dataset were only used to test and validate the original model that was trained on EFER2013 dataset.

B. MODEL TRAINING & TESTING

Model training and validation was accomplished on cloud-based virtual environment with Nvidia K80/T4 Graphical Processing Unit (GPU) and a RAM of 12 GB. Model training and test split was performed based on 80:20 ratio with 80% and 20% of the images used for model training and validation, respectively. The training phase took place for 45 epochs with a batch size of 128, with a convergence threshold set to 10-4 as recommended in the literature [13]. Model performance was optimized utilizing Adam optimizer with cross-entropy loss function [14]. Figure 6 shows the accuracy curves for the training and testing phases across the different epochs. With the above configuration, the training and validation accuracy of the model reached the maximum of 93.3% and 82.1%, respectively which was higher than the benchmarked approaches from the literature. Moreover, the overall classification recall, precision, and F1 score values were, 81.9%, 81.9%, and 81.8%, respectively. Figure 7 shows the confusion matrix after applying the proposed model on 20% of the EFER2013 dataset.

The second dataset (CK+) was used as the test set for the model initially developed based on FER2013 dataset with the confusion matrix shown in Figure 8. As shown in the Figure, the three types of emotions happy, sad, surprise were classified by the model with higher relative accuracy with 54% and only 19% of samples for anger and disgust

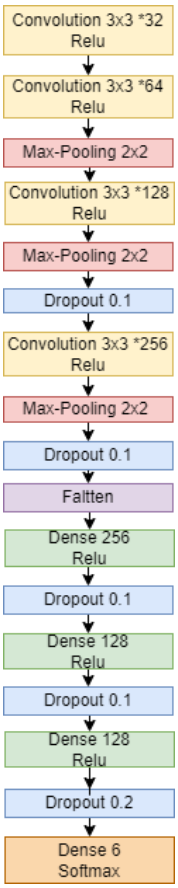


FIGURE 5. Proposed FER model architecture.

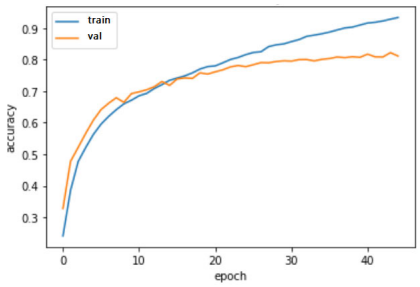


FIGURE 6. Accuracy curve during training and testing of the proposed model.

correctly classified, respectively. Furthermore, only 8% of the fear samples were correctly classified as the majority of the samples were misclassified as sad. This is partly due to higher congruence in the extracted features for fear, anger, sadness. The overall accuracy for the test set (CK+) was 65.8% with 60.6%, 70.7%, and 54.3% for recall, precision, and F1 score, respectively.

As the results demonstrate the proposed model recognizes sad, happy, and surprise emotions with high accuracy. On the other hand, the model recognition accuracy is relatively lower for the other emotions, anger, fear, and sad. However, most of the misclassified samples from these three classes were

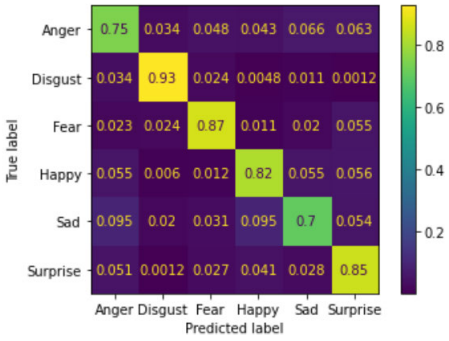


FIGURE 7. Confusion matrix after applying the proposed model on EFER2013 testing set.

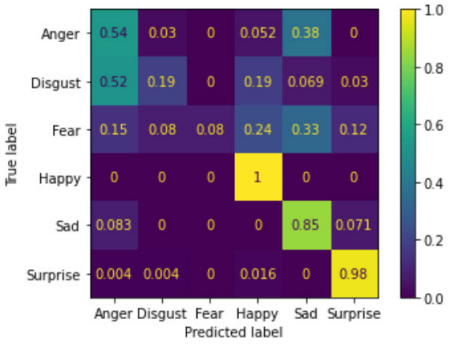


FIGURE 8. Confusion matrix after applying the proposed model on CK+ testing set.

misclassified as either angry or sad. This is not surprising as facial expressions can slightly vary among individuals, and may mix different emotional states experienced at the same time. Moreover, some share some of the facial features. For example, pulled up eyebrow is one feature that can exist in anger, disgust, and fear faces. These challenges pose a great limitation on the model predictive capability [8], [11]. Some works in the literature [3], [7], [23], solved this problem by minimizing the number of classes or combining these classes under one class.

Some samples from EFER2013 and CK+ datasets with the actual and predicted labels are shown in Figures 9 and 10.

C. COMPARISON WITH THE RELATED WORKS

To benchmark the proposed solution with the existing FER systems, a set of criteria was developed to evaluate model performance according to various evaluation metrics that include:

- **Approach:** Type of approach used in the FER system.
- **Accuracy:** Achieved accuracy of the proposed system. This evaluates its ability to correctly recognize emotions.
- **Parameters:** Total number of parameters used in the proposed model. This element is important as it gives an indication about the amount of overhead that the model can produce.
- **Model size:** Size of the proposed model.

FIGURE 9. Samples from EFER2013 dataset with the actual and predicted labels.

FIGURE 10. Samples from CK+ dataset with the actual and predicted labels.

TABLE 1. Comparison between the proposed FER system and the related works.

Model	Approach	Accuracy	Parameters	Model Size	Application	Number of Classes
Latitha, et al.,2021 [19]	CNNs	66.7% on FER2013	-	-	General	7
Vulpe-Grigorasi & Grigore, 2021 [35]	CNNs with Random Search	72.16% on FER2013	5.2M	59 MB	General	7
Minaee, et al., 2021 [26]	Attentional CNNs	70.02% on FER2013 98.0% on CK+ 99.3% on FERG 92.8% on JAFFE	-	-	General	7
Hou, et al., 2022 [13]	MTCNN	67.40% on FER2013 99.18% on CK+	-	-	Education	7
Shi, et al., 2021 [31]	MBCC-CNNs	71.52% on FER2013 98.48% CK+ 88.10% FER+ 87.34% RAF	-	-	General	7
Zhou, et al., 2020 [39]	HOG-CNNs	71.73% on FER2013			General	7
Kim, et al., 2021 [17]	FPGA-Based CNNs	86.5% on FER+	-	0.39 MB	General	7
Saurav, et al., 2022 [29]	Dual CNNs	72.77% on FER2013 98.54% on CK+	1.08M		General	7
Zhao, et al., 2020 [38]	CNN with Compression Techniques	71.55% on FER2013 85.68% on FER+ 85.89% on FERFIN	0.21M	-	General	7
Saurav, et al. 2021 [30]	Dual CNNs with Fusion Schemes	74.1% on FER2013 84% on RAF 53% on SFEW	4.81M	19.3 MB	General	7
Burrows, et al., 2021 [7]	CNNs and GANs	58.71% (7 classes) 73.71% (3 classes) 77.83% (2 classes)	1.5M	-	General	7, 3,2
Joseph & Mathews, 2021 [16]	CNNs	67.18% on FER2013	-	-	VIP	7
Lutfallah, et al., 2022 [23]	ResNet with GRU & RNN	87% on CK+	-	-	VIP	3
Ashok & John, 2018 [3]	SVM	-	-	-	VIP	3
Proposed Model Trained on FER2013	**CNNs**	**82.1% on FER2013 66.7% on CK+**	**1.49M**	**0.24 MB**	**VIP**	**6**
Proposed Model Trained on CK+	**CNNs**	**98.5% on CK+**	**1.49M**	**0.24 MB**	**VIP**	**6**

Note: Entries with "-" in the table indicate the absence of information from the selected studies.

- **Application:** Target audience of the proposed model.
- **Number of Classes:** Model capability to perform recognition on the number of emotions.

A detailed comparison of the proposed FER system with models from the literature is summarized in Table 1. Results show that the proposed model achieved the highest validation accuracy on FER2013 dataset with a value of 82.1%. Moreover, with only 1.49M parameters comprising of the smallest model size, implementation of the proposed FER system in wearable edge devices, thus leading to its usability at a greater extent compared to other large models. Moreover, although the classification accuracy for CK+ dataset of 891 unseen samples was only 66.7%, the proposed model can easily be fine-tuned on a newly added smaller dataset to significantly improve its accuracy. Such an approach was already adopted where the initially developed model based on EFER2013 was tested on CK+ where partial of the CK+ dataset could also be used to tune model parameters for higher accuracy.

Additionally, in comparison with the previous works, the proposed model was also trained and tested on CK+ which achieved a validation accuracy of 98.5% which was higher or equivalent to most of the previous works.

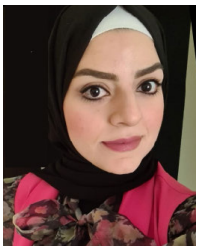
V. CONCLUSION

VIP cannot perceive visual and other non-verbal cues making normal conversations prone to misinterpretation. FER systems can be used to address this social challenge but most of the proposed FER systems have not been proposed for VIP specifically and therefore are not suitable to the nature of the use of the VIP. In this paper, a FER system based on Partial Transfer training and CNN was proposed. The VIP dedicated system addressed the limitations of the existing systems, by providing a portable, lightweight, and accurate assistive system. The proposed system was evaluated on FER2013 and CK+. The proposed model attained an accuracy of 82.1% on FER2013 dataset which is a notable improvement over the current state-of-the-art while maintaining a proper balance between recognition accuracy and computational efficiency. It also achieved an accuracy of 66.7% on the CK+ test set. Overall, the model can recognize happy, sad, and surprise emotions with high accuracy. It has a relatively lower accuracy rate for anger, disgust, and fear as it tends to mislabel some samples from these emotions as sad. Despite the good performance of the proposed model, there is still a range of points for further investigation and improvement. Our next plan is to train the proposed model on more samples from anger, disgust, and fear classes to enhance its recognition accuracy. Moreover, future efforts will concentrate on implementing the prototype for the designed FER system by deploying the trained FER model into Raspberry Pi to further test and enhance the proposed model.

REFERENCES

[1] *Facial Expressions in the Wild (SFEW/AFEW)*, IEEE, 2011.
[2] D. Aneja, A. Colburn, G. Faigin, L. Shapiro, and B. Mones, "Modeling stylized character expressions via deep learning," in *Proc. Asian Conf. Comput. Vis.* Cham, Switzerland: Springer, 2016, pp. 136–153.
[3] A. Ashok and J. John, "Facial expression recognition system for visually impaired," in *Proc. Int. Conf. Intell. Data Commun. Technol. Internet of Things*. Cham, Switzerland: Springer, 2018, pp. 244–250.
[4] E. Barsoum, C. Zhang, C. C. Ferrer, and Z. Zhang, "Training deep networks for facial expression recognition with crowd-sourced label distribution," in *Proc. 18th ACM Int. Conf. Multimodal Interact.*, Oct. 2016, pp. 279–283.
[5] R. Bourne, "Trends in prevalence of blindness and distance and near vision impairment over 30 years: An analysis for the global burden of disease study," *Lancet Global Health*, vol. 9, no. 2, pp. e130–e143, 2021.
[6] R. R. Bourne, "Magnitude, temporal trends, and projections of the global prevalence of blindness and distance and near vision impairment: A systematic review and meta-analysis," *Lancet Global Health*, vol. 5, no. 9, pp. e888–e897, 2017.
[7] H. Burrows, J. Zarrin, L. Babu-Saheer, and M. Maktab-Dar-Oghaz, "Real-time emotional reflective user interface based on deep convolutional neural networks and generative adversarial networks," *Electronics*, vol. 11, no. 1, p. 118, Dec. 2021.
[8] S. Du, Y. Tao, and A. M. Martinez, "Compound facial expressions of emotion," *Proc. Nat. Acad. Sci. USA*, vol. 111, no. 15, pp. E1454–E1462, Apr. 2014.
[9] P. Ekman, *Basic Emotions*. Hoboken, NJ, USA: Wiley, 1999, ch. 3, pp. 45–60.
[10] I. J. Goodfellow, D. Erhan, P. L. Carrier, A. Courville, M. Mirza, B. Hamner, W. Cukierski, Y. Tang, D. Thaler, and D.-H. Lee, "Challenges in representation learning: A report on three machine learning contests," in *Proc. Int. Conf. Neural Inf. Process.* Cham, Switzerland: Springer, 2013, pp. 117–124.
[11] T. Gremsl and E. Hödl, "Emotional AI: Legal and ethical challenges," *Inf. Polity*, vol. 27, no. 2, pp. 1–12, Apr. 2022.
[12] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.
[13] C. Hou, J. Ai, Y. Lin, C. Guan, J. Li, and W. Zhu, "Evaluation of online teaching quality based on facial expression recognition," *Future Internet*, vol. 14, no. 6, p. 177, Jun. 2022.
[14] Y. Huang, C. Dong, X. Luo, and Q. Dai, "Facial expression recognition algorithm based on improved VGG16 network," in *Proc. 6th Int. Symp. Comput. Inf. Process. Technol. (ISCIPT)*, Jun. 2021, pp. 480–485.
[15] M. Jaderberg, "Spatial transformer networks," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 28, 2015, pp. 2017–2025.
[16] J. L. Joseph and S. P. Mathew, "Facial expression recognition for the blind using deep learning," in *Proc. IEEE 4th Int. Conf. Comput., Power Commun. Technol. (GUCON)*, Sep. 2021, pp. 1–5.
[17] J. Kim, J.-K. Kang, and Y. Kim, "A resource efficient integer-arithmetic-only FPGA-based CNN accelerator for real-time facial emotion recognition," *IEEE Access*, vol. 9, pp. 104367–104381, 2021.
[18] A. Kunz, R. Koutny, and K. Miesenberger, "Accessibility of co-located meetings," in *Proc. Int. Conf. Comput. Helping People With Special Needs*. Cham, Switzerland: Springer, 2022, pp. 289–294.
[19] S. K. Lalitha, J. Aishwarya, N. Shivakumar, T. Srilekha, and G. C. R. Kartheek, "A deep learning model for face expression detection," in *Proc. Int. Conf. Recent Trends Electron., Inf., Commun. Technol. (RTEICT)*, Aug. 2021, pp. 647–650.
[20] S. Li and W. Deng, "Reliable crowdsourcing and deep locality-preserving learning for unconstrained facial expression recognition," *IEEE Trans. Image Process.*, vol. 28, no. 1, pp. 356–370, Jan. 2019.
[21] S. Li, W. Deng, and J. Du, "Reliable crowdsourcing and deep locality-preserving learning for expression recognition in the wild," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 2584–2593.
[22] 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. Soc. Conf. Comput. Vis. Pattern Recognit. Workshops*, Jun. 2010, pp. 94–101.
[23] M. Lutfallah, B. Käch, C. Hirt, and A. Kunz, "Emotion recognition—A tool to improve meeting experience for visually impaired," in *Proc. Int. Conf. Comput. Helping People With Special Needs*. Cham, Switzerland: Springer, 2022, pp. 305–312.
[24] M. Lyons, M. Kamachi, and J. Gyoba, "The Japanese female facial expression (JAFFE) dataset," Zenodo, Apr. 1998, doi: 10.5281/zenodo.3451524.
[25] F. M. A. Mazen, A. A. Nashat, and R. A. A. A. A. Seoud, "Real time face expression recognition along with balanced FER2013 dataset using CycleGAN," *Int. J. Adv. Comput. Sci. Appl.*, vol. 12, no. 6, pp. 1–12, 2021.
[26] S. Minaee, M. Minaei, and A. Abdolrashidi, "Deep-emotion: Facial expression recognition using attentional convolutional network," *Sensors*, vol. 21, no. 9, p. 3046, Apr. 2021.
[27] D. Phutela, "The importance of non-verbal communication," *IUP J. Soft Skills*, vol. 9, no. 4, p. 43, 2015.
[28] G. C. Porusniuc, F. Leon, R. Timofte, and C. Miron, "Convolutional neural networks architectures for facial expression recognition," in *Proc. E-Health Bioeng. Conf. (EHB)*, Nov. 2019, pp. 1–6.
[29] S. Saurav, P. Gidde, R. Saini, and S. Singh, "Dual integrated convolutional neural network for real-time facial expression recognition in the wild," *Vis. Comput.*, vol. 38, pp. 1083–1096, Feb. 2021.
[30] S. Saurav, R. Saini, and S. Singh, "EmNet: A deep integrated convolutional neural network for facial emotion recognition in the wild," *Int. J. Speech Technol.*, vol. 51, no. 8, pp. 5543–5570, Aug. 2021.
[31] C. Shi, C. Tan, and L. Wang, "A facial expression recognition method based on a multibranch cross-connection convolutional neural network," *IEEE Access*, vol. 9, pp. 39255–39274, 2021.
[32] E. W. Simon, M. Rosen, E. Grossman, and E. Pratowski, "The relationships among facial emotion recognition, social skills, and quality of life," *Res. Develop. Disabilities*, vol. 16, no. 5, pp. 383–391, Sep. 1995.
[33] P. Singh, R. Srivastava, K. P. S. Rana, and V. Kumar, "A multimodal hierarchical approach to speech emotion recognition from audio and text," *Knowl.-Based Syst.*, vol. 229, Oct. 2021, Art. no. 107316.

[34] J. D. Steinmetz, "Causes of blindness and vision impairment in 2020 and trends over 30 years, and prevalence of avoidable blindness in relation to vision 2020: The right to sight: An analysis for the global burden of disease study," *Lancet Global Health*, vol. 9, no. 2, pp. e144–e160, 2021.
[35] A. Vulpe-Grigorasi and O. Grigore, "Convolutional neural network hyperparameters optimization for facial emotion recognition," in *Proc. 12th Int. Symp. Adv. Topics Electr. Eng. (ATEE)*, Mar. 2021, pp. 1–5.
[36] Q. Wang, B. Wu, P. Zhu, P. Li, W. Zuo, and Q. Hu, "ECA-Net: Efficient channel attention for deep convolutional neural networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 11534–11542.
[37] G. Zhao, W. Wei, X. Xie, S. Fan, and K. Sun, "An FPGA-based BNN real-time facial emotion recognition algorithm," in *Proc. IEEE Int. Conf. Artif. Intell. Comput. Appl. (ICAICA)*, Jun. 2022, pp. 20–24.
[38] G. Zhao, H. Yang, and M. Yu, "Expression recognition method based on a lightweight convolutional neural network," *IEEE Access*, vol. 8, pp. 38528–38537, 2020.
[39] N. Zhou, R. Liang, and W. Shi, "A lightweight convolutional neural network for real-time facial expression detection," *IEEE Access*, vol. 9, pp. 5573–5584, 2021.
[40] W. Zhou, S. Gao, L. Zhang, and X. Lou, "Histogram of oriented gradients feature extraction from raw Bayer pattern images," *IEEE Trans. Circuits Syst. II, Exp. Briefs*, vol. 67, no. 5, pp. 946–950, May 2020.



DINA SHEHADA (Member, IEEE) received the B.Sc. and M.Sc. degrees in computer engineering from Khalifa University. She is currently pursuing the Ph.D. degree with the University of Sharjah. She has more than ten years of experience in the academic field. Her research interests include trust evaluation in social networks, formal verification methods, network and information security, image processing, machine learning, and secure IoT systems.

AYAD TURKY received the Ph.D. degree in computer science and IT from the School of Science, RMIT University, Melbourne, VIC, Australia. He is currently an Assistant Professor of computer science with the College of Computing and Informatics, University of Sharjah. He has published more than 30 papers in international journals and peer-reviewed conferences. His current research interests include the design and development of hyper-heuristic frameworks, deep learning, machine learning, evolutionary computation, and hybrid algorithms, with a specific interest in big data optimization problems, cloud computing, dynamic optimization, and data-mining problems.

WASIQ KHAN (Senior Member, IEEE) received the B.Sc. degree in mathematics, physics, and geography and the M.Sc. degree in computer science from Pakistan, and the M.Sc. degree in artificial intelligence for board games and the Ph.D. degree in speech analysis and intelligent reasoning from the University of Bradford, U.K. He received a Postgraduate Certificate in teaching and learning in higher education (PGCHEP). He is currently a Senior Academic in artificial intelligence and data sciences with the Department of Computer Science, Liverpool John Moores University, U.K. He is also a Visiting Professor of artificial intelligence with the University of Anbar, Iraq. He has been publishing the research outcomes in high-impact journals, peer-reviewed conferences, news blogs and media, scientific festivals, and public events. He is an active reviewer for top-ranked journals (e.g., IEEE TRANSACTIONS) and government funding bodies.

BILAL KHAN (Senior Member, IEEE) received the first M.Sc. degree in pervasive computing, the second M.Sc. degree in computer science, and the Ph.D. degree in computer science (artificial intelligence) with the thesis title "Game theoretic coalitional routing in cooperative vehicular ad hoc networks." From 2013 to 2020, he was an Assistant Researcher with the University of California, Los Angeles (UCLA), where he developed the most robust and comprehensive decision support systems as an online nanoinformatics platform to assist regulatory bodies in the management of nanotechnology. The nanoinformatics platform consists of machine learning and data mining models that are supported by the largest database of nanomaterials. He maintained the platform via a high-performance computing cluster that was designed from scratch (with 24 compute nodes and 115TB of storage space for high-performance computations and simulations). At UCLA, he also led the development of a cyber-infrastructure for a virtual water district for efficient use and consumption of water in small rural agricultural communities. In addition, he developed data-driven approaches (using machine and deep learning techniques) for water use patterns in small communities as online applications.

ABIR HUSSAIN (Senior Member, IEEE) received the Ph.D. degree from The University of Manchester (UMIST), U.K., in 2000, with a thesis "Polynomial Neural Networks for Image and Signal Processing." She is currently a Professor of image and signal processing with the University of Sharjah. She is also a Visiting Professor of machine learning with Liverpool John Moores University, U.K. She was involved with higher order and recurrent neural networks and their applications to e-health and medical image compression techniques. She has developed with her research students a number of recurrent neural network architectures. She is a Ph.D. supervisor and an external examiner for research degrees, including Ph.D. and M.Phil. students. She is one of the initiators and chairs of the development of e-Systems Engineering (DeSE) series. Her research interests include neural networks, signal prediction, telecommunication fraud detection, and image compression. She has published numerous refereed research papers in conferences and journals in the research areas.