Facial Emotion Recognition for Visually Impaired People Using Transfer Learning

Anandhu T G[1*], Areena Aji[2*], Jithin K A[3*], Sukanyathara J[4*], Rotney Roy Meckamalil[5*]
*Department of Computer Science and Engineering
Mar Athanasius College of Engineering, Kothamangalam, Kerala
[1]anandhutg032@gmail.com, [2]areenaaji12@gmail.com, [3]jithinka23@gmail.com [4]sukanyathara.j@gmail.com
[4]rotney.rotney@gmail.com

***Abstract*—Individuals with visual impairment often face challenges in social interactions, specifically at recognizing emotional cues. The proposed framework tackles this issue head-on by devising a Facial Emotion Recognition(FER) system, by employing an advanced Transfer Learning approach within Convolutional Neural Networks (CNNs). By leveraging the dataset FER-2013 [13], the proposed system aims to transcend the limitations of traditional emotion recognition methods. Transfer learning allows the model to benefit from pre-trained knowledge on vast datasets, making it more efficient and effective in capturing complex facial features associated with different emotions. This approach is designed to offer better accuracy and generalization capabilities than other conventional methods. During training, the system will be designed to comprehensively capture the intricacies of facial expressions, enabling it to not only identify individuals but also interpret subtle changes in their emotional states throughout conversations. An innovative audio output system will be integrated into the FER system to provide a smooth and accessible experience for visually impaired users, allowing for a better understanding of social dynamics. By emphasizing transfer learning, this framework is designed to be efficient and robust, potentially revolutionizing emotional understanding for visually impaired individuals and setting a new standard in the field by showcasing the superior performance achievable through advanced machine learning techniques. Ultimately, this research aims to bridge the social gap for the visually impaired by fostering inclusivity, independence, and safety in their daily life.**

***Index Terms*—visually impaired, facial emotion recognition, transfer learning, convolutional neural networks, computer vision, facial recognition.**

I. Introduction

Vision is critical for understanding our surroundings, and its loss significantly hinders social interaction. People with visual impairments struggle to perceive nonverbal cues[1], especially emotions expressed through facial features. To address this challenge, we propose a facial emotion recognition system with real-time audio feedback. This system will enhance social interaction for Visually Impaired People (VIPs) by accurately interpreting emotional cues in real time.

In this paper, we propose a Transfer Learning based Convolutional Neural Network(CNN) for Facial Emotion Recognition. By noticing small changes in facial features like raised eyebrows, frowns, and smiles, the system is designed to predict the emotional states of individuals during social interactions. The VIPs can now better understand how others are feeling in social situations, thanks to this new technology, which is a big step forward in assistive technologies.

In addition to recognizing emotions in real-time, the system incorporates facial recognition. This allows users to identify familiar faces during social interactions, enriching their experience by enabling them to instantly recognize friends or acquaintances. By integrating facial recognition algorithms, the system can generate discreet notifications when a known person is present, facilitating smoother interactions and fostering a stronger sense of connection in everyday life.

In summary, this Facial Emotion Recognition System represents a huge advancement in social inclusion for visually challenged people. By leveraging machine learning and facial recognition, the technology enables users to manage real-time social interactions more successfully. This newfound ability to understand emotions and recognize familiar faces promotes a more meaningful and connected social experience, ultimately contributing to a more equal and inclusive world.

II. Literature Survey

A. Facial Emotion Recognition and Encoding Application for the Visually Impaired.

In the paper titled "Facial Emotion Recognition and Encoding Application for the Visually Impaired" [12], the authors have developed a socially assistive app tailored for those who are visually impaired. The application leverages a transfer learning Facial Expression Recognition (FER) model, which is embedded within a mobile application, to recognize facial expressions. The approach includes utilizing deep learning and app development technologies in order to develop an application capable of recognizing different human expressions. The recognized gesture is then relayed to the user through tactile feedback. The results of this research indicate that the application improves the social awareness of blind individuals by aiding them in interpreting the facial expressions of those they are engaging with. Nevertheless, possible disadvantages could involve restrictions in practical use, like different lighting situations affecting the accuracy of recognition, and the necessity for ongoing improvements to the model to guarantee top performance in various

circumstances.

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

A new method for recognizing facial emotions by utilizing a CNN that has been custom-trained and incorporating partial transfer learning was introduced through "A Lightweight Facial Emotion Recognition System Using Partial Transfer Learning for Visually Impaired People" [2]. The process consisted of training a CNN on a specific dataset and then applying the acquired features to a different dataset. The researchers created a portable facial expression recognition system with wireless capabilities, designed specifically for individuals who are visually impaired. The system showed a significant enhancement compared to the existing technology, achieving the top accuracy of 82.1% on the improved FER2013 dataset for Facial Expression Recognition in 2013. However, the document fails to address potential restrictions or disadvantages of the suggested system. Future research could concentrate on assessing the system's performance in real-life situations, its usability for people with different levels of visual impairment, and its effectiveness in various social contexts for a thorough understanding.

C. Multimodal Emotion Recognition Based on Cascaded Multichannel and Hierarchical Fusion.

A multimodal emotion recognition framework, named CMC-HF1, is suggested in [3]. This system utilizes visual, speech, and text cues as inputs from various modes. The approach includes three successive channels based on deep learning that extract features for each mode, enhancing information extraction and improving recognition performance within each mode. An enhanced fusion module was also implemented to facilitate interactions between the three modes and enhance recognition and classification accuracy even further. The researchers performed tests to assess two standard datasets, IEMOCAP and CMU-MOSI1. The outcomes indicated improved accuracy in both datasets compared to current cutting-edge techniques. Potential disadvantages could involve difficulties in coordinating and matching various forms, along with the requirement for extensive and varied datasets to guarantee the applicability of the framework in different situations and with different groups.

D. Hybrid Facial Emotion Recognition Using CNN-Based Features

"Hybrid Facial Emotion Recognition Using CNN-Based Features" [4] introduces a combination method for identifying facial emotions by utilizing characteristics taken from a pre-existing Convolutional Neural Network (CNN). The process includes obtaining acquired features from a pre-trained CNN and assessing various machine learning (ML) algorithms for classification purposes. It examines the effects of using different ML algorithms instead of the typical SoftMax classifier on the FC6, FC7, and FC8 layers of Deep Convolutional Neural Networks (DCNNs). Tests were performed on two popular CNN models, AlexNet and VGG-16, with a dataset of masked facial expressions (MLF-W-FER dataset). The findings indicate that Support Vector Machine (SVM) and Ensemble classifiers are more effective than the SoftMax classifier on both AlexNet and VGG-16 architectures. These algorithms were successful in enhancing accuracy by 7% to 9% in individual layers. Challenges arise from the intricate process of incorporating various ML algorithms into the recognition pipeline and the necessity for extensive trials to determine the best algorithm for a specific dataset.

E. An Adaptive Convolutional Neural Network Model for Human Facial Expression Recognition

Creating a new model that can transfer features between datasets, with the goal of simplifying training and saving computational resources while still achieving high accuracy in facial expression recognition tasks was engaged in "An Adaptive Convolutional Neural Network Model for Human Facial Expression Recognition" [5] . The authors tackled the issue of multivalued classification of motor units by developing a technique to adjust the parameters of the CNN model. They tried out different specialized CNN designs and pre-trained models on the ImageNet dataset in their experiments. They were able to improve the accuracy of recognizing facial expressions on human face images by utilizing transfer learning methods. Yet, significant adjustments to parameters and validation were needed to guarantee top performance on various datasets and expression fluctuations.

F. Hybrid CNN-SVM Classifier for Human Emotion Recognition Using ROI Extraction and Feature Fusion

The authors suggest a hybrid method that merges Convolutional Neural Network (CNN) and Support Vector Machine (SVM) for emotion recognition in their paper titled "Hybrid CNN-SVM Classifier for Human Emotion Recognition Using ROI Extraction and Feature Fusion" [14]. The process includes preparing images, identifying Regions of Interest (ROI), and utilizing feature fusion methods such as Local Binary Pattern and Gabor feature extraction. The goal of the hybrid model is to enhance accuracy by preprocessing images to improve quality and contrast, then extracting and merging features to generate a strong representation for classification. Nonetheless, the precision is influenced by both the learning rate and the quantity of layers within the CNN model. The article proposes that upcoming studies could investigate more advanced hybrid methods or optimization techniques in order to predict different emotional states with greater accuracy.

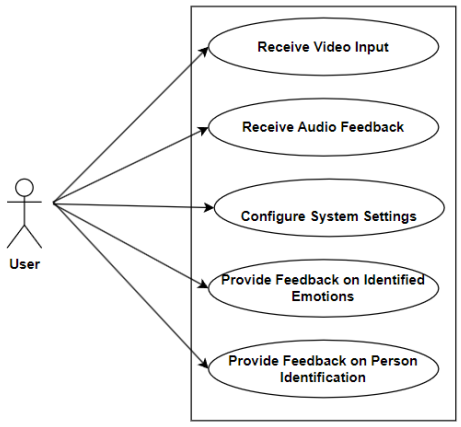

Fig. 1. **Use Case Diagram of Proposed Model**

III. Proposed Methodology

This section details the methodology for developing a Facial Emotion Recognition System developed using Convolutional Neural Networks by implementing Transfer Learning, to enhance real-time social interactions for visually impaired individuals. We present the system design (A), model architecture (B), data acquisition (C), model training (D), and known-face detection (E).

A. System Design

The proposed Facial Emotion Recognition(FER) System prioritizes portability and real-time functionality for social interactions. A miniature camera mounted on a user-worn device captures video frames during conversation. These frames undergo a multi-step processing pipeline to extract relevant information and convey emotions to the visually impaired user (VIP) via an earphone.

First, individual frames are extracted from the captured video stream. These frames are converted to gray scale for efficient processing. Face detection algorithms then identify and locate faces within each frame. The detected faces are cropped, resized, and normalized to ensure consistency for the emotion recognition model.

Following preprocessing, the system employs a Convolutional Neural Network (CNN) model [11], outlined in Section B, to classify emotions. This model is trained to recognize and categorize facial expressions into seven unique emotions: anger, disgust, fear, happiness, sorrow, surprise, and neutral. Upon successful classification, the recognized emotional state is converted into an audio message and delivered to the VIP through an earphone. This real-time audio feedback provides valuable social cues, enhancing the user's ability to navigate social interactions effectively.

B. Model Architecture

The proposed FER System leverages transfer learning to harness the power of a pre-trained deep learning model. This approach significantly reduces training time and computational resources compared to building a model from scratch. The system utilizes the MobileNetV2 architecture, known for its efficient feature extraction capabilities honed on the vast ImageNet dataset. By employing MobileNetV2 as a foundation, the FER system capitalizes on its ability to identify and extract relevant visual patterns from facial images.

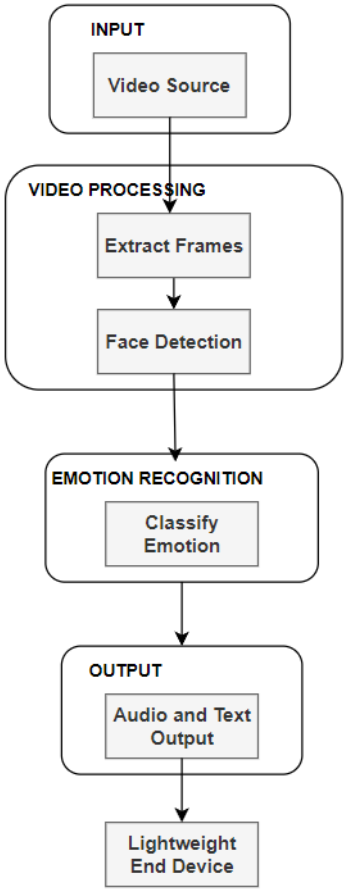

Fig. 2. **Flowchart of emotion recognition model**

To adapt MobileNetV2 for the specific task of emotion recognition, the system adopts a two-step customization process. The first step involves initializing the pre-trained model while excluding its classification head. This preserves the valuable feature extraction layers, allowing the system to benefit from the extensive image recognition knowledge embedded within MobileNetV2. Subsequently, the system appends custom layers specifically tailored for emotion recognition.

These custom layers consist of two Dense layers designed for feature extraction and classification. The first Dense layer, equipped with 128 neurons and a ReLU activation function, facilitates the extraction of high-level features from facial expressions. Following this, a second Dense layer with 64 neurons and ReLU activation further refines the extracted features. This step enhances the model's ability to differentiate

subtle emotional cues crucial for accurate recognition. Finally, a terminal Dense layer with 7 neurons and softmax activation is integrated. This layer predicts the probability distribution across the seven emotion classes: anger, disgust, fear, happiness, sadness, surprise, and neutral.

The sequential architecture of the FER leverages the strengths of MobileNetV2 for feature extraction while incorporating custom layers specifically designed for emotion recognition. This approach not only expedites the training process but also improves overall performance, leading to more accurate and efficient emotional state predictions from facial images

C. Dataset Acquisition

The Facial Expression Recognition 2013 (FER2013) dataset [13] serves as the basis for developing and testing the proposed Facial Emotion Recognition System [10]. This publicly available dataset contains 35,897 grayscale facial photos classified into seven different emotions: anger, contempt, fear, happiness, sorrow, surprise, and neutral. Each image has a centered and cropped face with a resolution of 48x48 pixels.

While FER2013 provides a varied variety of emotions, one key problem is the dataset's inherent class inequality. The amount of photos varies greatly between categories, with some emotions, such as "disgust," having far fewer examples (about 600) than others, such as "happy," which has nearly 5,000 images. To solve this imbalance and ensure that the suggested system learns well from all emotions, the project uses a data selection approach. The proposed FER system makes use of about 20,000 photos from the FER2013 dataset, which were carefully selected to ensure a more equal representation of all seven emotion classes. This strategy ensures that the system is trained on a dataset that more accurately represents the distribution of emotions seen in real-world circumstances.

D. Model Training

The suggested Facial Emotion Recognition system was trained on a laptop with an Intel Core i5-12500H processor and 16 GB of memory. To maximize training and validation, the dataset was divided in an 80:20 ratio. This signifies that 80% of the photos were used to train the model, with the remaining 20% reserved for validation purposes.

The training phase took place for 25 epochs, which allowed the model to effectively learn from the input data. During training, a precise setup was used to assure peak performance. The design employed the Adam optimizer and cross-entropy loss function [9], which are well-known strategies for training deep learning models. This setting improved the model's capacity to achieve high levels of accuracy.

E. Known-Face Detection

The proposed FER system extends its functionality beyond emotion recognition to encompass known-face detection. This feature significantly enhances the social interaction experience for visually impaired users (VIPs) by identifying familiar faces encountered in real-time.

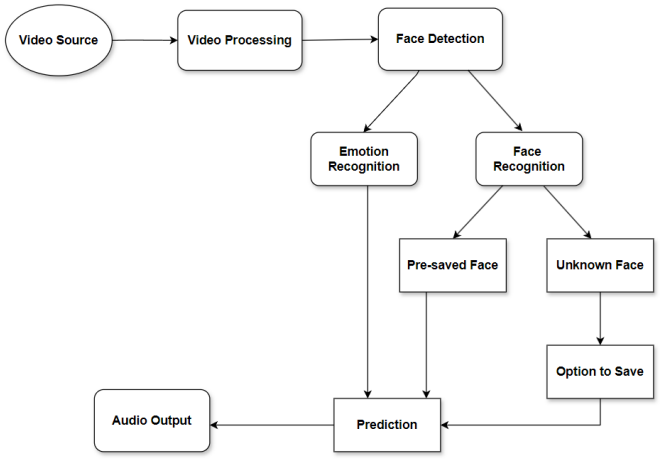

Fig. 3. **Design of proposed FER system**

The system leverages the *face_recognition* python library to facilitate known-face detection. VIPs can create a personalized database containing facial images and corresponding names of friends, family, or acquaintances. During operation, the camera captures video frames, and the system executes two critical tasks concurrently:

1) **Facial Emotion Recognition:** As described previously, the system analyzes facial expressions to identify emotions in real-time.
2) **Known-Face Identification:** The captured faces are compared against the VIP's database of known faces. This comparison occurs simultaneously with the emotion recognition process, maximizing efficiency.

If a match is detected, the system delivers helpful social clues by producing an audio output that includes the recognized person's name as well as their emotional condition. This combined information helps VIPs handle social encounters with increased effectiveness.

For unidentified faces, the system has an additional feature that promotes an inclusive social environment. When the FER system encounters an unknown face, it can take a real-time snapshot with the VIP's permission. This photo can then be added to the VIP's database, along with the new individual's name, allowing them to recognize them in future encounters. This functionality simulates real-world introductions and allows the VIP to broaden their social circle using the FER system.

IV. Result

The suggested Facial Emotion Recognition System was tested on the FER2013 dataset. This evaluation aimed to assess the model's ability to accurately classify emotions from unseen facial images.

Testing revealed varying degrees of accuracy across the seven emotion classes. The model achieved the highest recognition rates for happiness (77%), neutral (73%), disgust (69%)

Fig. 4. **Accuracy Curve during training of proposed model**

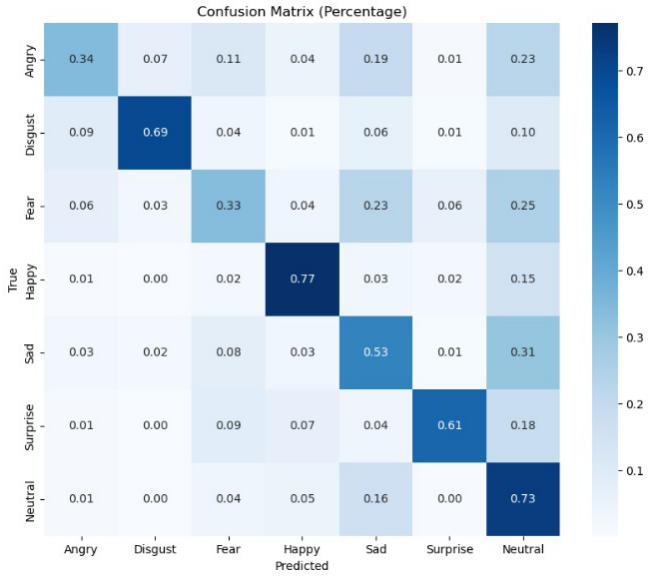

Fig. 5. **Confusion Matrix after applying proposed model on FER2013 testing set**

and surprise (61%) emotions. However, the accuracy for anger (34%), fear (33%), and sadness (53%) was lower.

These performance differences can be attributed to a variety of factors. For instance, facial expressions vary slightly between individuals, making it challenging to establish a universal recognition standard. Second, emotions can be complicated and multidimensional, and people may experience an array of feelings concurrently. This complexity might result in facial expressions that share certain characteristics, thereby confusing the model [7] [8]. A raised brow, for example, might be used to represent anger, disgust, or fear. These inherent issues in facial expressions limit any model's predictive potential.

It's important to note that the primary focus of this project lies in the real-time functionality of the FER, not solely on achieving the highest possible accuracy on a benchmark dataset. The true test of the system lies in its ability to deliver accurate results in real-world scenarios using frames captured from a webcam. Initial testing under controlled conditions with a high-quality camera and proper lighting yielded promising results. However, further investigation and testing in real-time use are necessary. This will enable us to continuously improve the system and optimize its performance according to different lighting conditions, webcam variations, and the specific needs and preferences of visually impaired users.

V. FUTURE SCOPE

Future advancements in facial emotion recognition systems with audio output for people with visual impairments through transfer learning have enormous promise to improve social interactions. To increase the capabilities of these systems, a number of topics might be investigated.

1) Including Multimodal Inputs: To gain a more comprehensive knowledge of social cues beyond facial emotions, use touch or scent inputs.
2) Real-time learning algorithms: Algorithms can be used to update emotional recognition models in response to user feedback, which will enhance accuracy and provide more individualized interpretations
3) Transfer Learning Techniques: Increase system performance by using pre-trained models to handle a bigger database of faces and emotions.
4) Features for customization: Increase user happiness by letting consumers specify their preferences for emotional cues.
5) Ethical Considerations: When developing new systems, give data protection and bias mitigation top priority
6) Collaborative Interfaces: Provide inclusive interfaces that facilitate smooth communication between those with visual impairments and those who are sighted.

By concentrating on these areas, transfer learning-based facial emotion recognition with audio output for visually impaired people seeks to empower users and promote a more inclusive society.

VI. CONCLUSION

In conclusion, a major advancement in helping people with visual impairments has been made with the development of a facial emotion recognition system with audio output through the use of transfer learning techniques. This system demonstrates the efficacy of merging different technologies for real-time emotional interpretation, while also improving social interactions and emotional comprehension. We have successfully enhanced user experiences, encouraging natural interactions and boosting confidence and independence among visually impaired users by lowering computational burdens and offering aural feedback.

In order to advance facial expression recognition systems, it will be essential to continuously refine and integrate emerging technologies like multi-modal interfaces, real-time learning algorithms, and privacy considerations. Because these developments facilitate communication and foster real human connections, they have the potential to build societies that are more accepting and compassionate. Therefore, by using technological innovation to empower underprivileged

user groups and promote holistic well-being, our continued efforts in this field are beneficial.

REFERENCES

[1] D. Phutela, "The importance of non-verbal communication," IUP J. Soft Skills, vol. 9, no. 4, p. 43, 2015.

[2] Shehada, D., Turky, A., Khan, W., Khan, B., Hussain, A.(2023). A Lightweight Facial Emotion Recognition System Using Partial Transfer Learning for Visually Impaired People. IEEE Access, 11, 36961-36969.

[3] Liu, Xia, Zhijing Xu, and Kan Huang. "Multimodal Emotion Recognition Based on Cascaded Multichannel and Hierarchical Fusion." Computational Intelligence and Neuroscience 2023 (2023)

[4] Shahzad, H. M., Bhatti, S. M., Jaffar, A., Akram, S., Alhajlah, M., Mahmood, A. (2023). Hybrid Facial Emotion Recognition Using CNN-Based Features. Applied Sciences, 13(9), 5572.

[5] Arsirii O.O., Petrosiuk D.V. "An adaptive convolutional neural network model for human facial expression recognition". Herald of Advanced Information Technology. 2023; Vol. 6 No. 2. 128–138. DOI:

[6] R. Kishore Kanna, Bhawani Sankar Panigrahi, Susanta Kumar Sahoo, Anugu Rohith Reddy ,Yugandhar Manchala, Nirmal Keshari Swain (2024). "CNN Based Face Emotion Recognition System for Healthcare Application"

[7] 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.

[8] T. Gremsl and E. Hödl, "Emotional AI: Legal and ethical challenges," Inf. Polity, vol. 27, no. 2, pp. 1–12, Apr. 2022.

[9] 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.

[10] 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.

[11] 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.

[12] Pushpalatha, M.N., Meherishi, H., Vaishnav, A. et al. Facial emotion recognition and encoding application for the visually impaired. Neural Comput Applic 35, 749–755 (2023).

[13] Facial Expression Recognition 2013 (FER2013) Dataset
Available: https://www.kaggle.com/datasets/msambare/fer2013

[14] Vaidya, K.S., Patil, P.M. Alagirisamy, M. Hybrid CNN-SVM Classifier for Human Emotion Recognition Using ROI Extraction and Feature Fusion. Wireless Pers Commun 132, 1099–1135 (2023).