See You

1st Jasmine Jolly
Dept. of Computer Science and Engineering
Christ College of Engineering,
Thrissur, India
jasminejolly@cce.edu.in

2nd Afzal M A
Dept. of Computer Science and Engineering
Christ College of Engineering,
Thrissur, India
afzalma261@gmail.com

3rd Chrislyne Pathrose
Dept. of Computer Science and Engineering
Christ College of Engineering,
Thrissur, India
chrislyne.pathrose@gmail.com

4rd Sarang Sajith
Dept. of Computer Science and Engineering
Christ College of Engineering,
Thrissur, India
sarangsaraz@gmail.com

5th Vaishnav T S
Dept. of Computer Science and Engineering
Christ College of Engineering,
Thrissur, India
vaishnavts33@gmail.com

***Abstract*—To help the blind community, a solution is to identify the individual in front of them and also detect the emotion of the same person. In this system, using facial recognition technologies, the individuals are identified, and with the help of machine learning, emotion detection is made possible. Face detection is a type of bio metric technique that refers to computerised systems automatically detecting a face by looking at it. It is a popular feature used in biometrics, digital cameras, and social tagging. Face detection and recognition have gained more research attention in the last few years. Emotion recognition, which is a part of affective computing, draws a lot of attention from researchers because of its broad applications. Unlike previous approaches to recognising human emotional states that relied on facial expression, speech, or gesture, some researchers believe that contextual information from the scene has potential. In conclusion, by combining both technologies, an effective and accurate solution can be obtained, and it will be very convenient for the blind to navigate through daily life with such ease identifying individuals.**

***Index Terms*—CNN, Emotion Recognition, Machine Learning, Deep Learning**

I. Introduction

It is difficult for people who are blind and visually impaired to recognize people in a variety of social interactions. Sole reliance on voice recognition may be difficult in some circumstances, and impossible in other circumstances. The inability to identify people is a disadvantage for blind people in many professional and educational situations.

Data science and Artificial Intelligence are revolutionizing the world through technical transformations. We can observe many machine learning applications in our day to day lives but one of the greatest applications of machine learning is to classify individuals based on their emotions. Such human recognition and emotion detection techniques are highly effective in increasing the popularity and attractiveness of products and services.

People in real life display their emotions on their faces to display their psychological processes and attitudes in social interactions. This project's main goal is to identify the emotion that an input image with a single facial expression belongs to. Emotion recognition can be precisely divided into the classification of basic emotion and the classification of compound emotion due to the complexity of interpreting a human face.

The emotion recognition problem requires an algorithm to perform feature extraction and categorical classification, much like all other classification tasks.

We need to develop a model that can categorize the input depending on the feature after extracting a certain feature from the data in order to classify an emotion.

These are the steps:

1) Extracting Features: The standard procedure is to identify faces and extract Action Units (AU) from them; some emotions comprise combinations of AUs that are coded as features.
2) Feature Analysis: A traditional classifier can use an unsupervised algorithm or a supervised method. Support Vector Machine is a classic example of a supervised algorithm, and Principle Component Analysis (PCA).
3) Generation of Results: Based on the analysis of facial features ,emotion is detected.

II. Motivation

Visually Impaired Persons (VIPs) make up a significant portion of the population and are present in all regions of the world, and without additional interventions this number is expected to increase significantly. The Visually Impaired (VIP) has difficulty performing activities of daily living.

Opportunities to work, learn, move around, interact with the environment and search for common objects (indoor / outdoor) alone or with help. The main challenges for VIP are object detection and recognition, currency identification, textual information (characters, symbols) and translation, mobility / navigation, and security. One of the most difficult tasks for the visually impaired is to identify people and understand their personalities. The inability to recognize a person without auditory or tactile cues severely limits the social interaction of the blind and jeopardizes it for safety reasons. Blind people adjust to normal life using traditional aids such as walking sticks and hounds. White canes are preferred because they are easy to use, inexpensive, and widely accepted. From the blind community. However, these ancillary elements quickly reach their limits when faced with the wide variety of situations that can occur in today's urban scenes. Moreover, the white cane cannot provide additional information to users such as the degree of danger of the encountered obstacles or recognition of persons that are present in the scene. So being able to propose a system that enables them to recognize human faces and detect their emotions improves the access, integration, and independence of the blind in all situations.

III. Literature Survey

The first paper "DEEP-SEE FACE: A Mobile Face Recognition System Dedicated to Visually Impaired People" introduces the DEEP-SEE FACE framework, an assistive device designed to improve cognition, interaction, and communication of visually impaired (VI) people in social encounters. The proposed approach jointly exploits computer vision algorithms (region proposal networks, ATLAS tracking and global, and low-level image descriptors) and deep convolutional neural networks in order to detect, track, and recognize, in real-time, various persons existent in the video streams.

In the next paper "Design and evaluation of a real-time face recognition system using convolutional neural network " presents a design of a real-time face recognition system using CNN. The proposed system is designed using an open-source library called Keras running on top of the tensor flow.

In the next paper "An improved face recognition algorithm and its application"presents a new method using local binary pattern(LBP) algorithm combined with advanced image processing techniques such as contrast adjustments, Bilater filter, histogram equalisation and image bending, thus improving accuracy of the overall system.

The next paper "Research on 3D face recognition method based on LBP and SVM" proposes a 3-D facial recognition method to improve speed and accuracy of face recognition that combines LBP and SVM. First LBP algorithm is used to extract the feature information from the 3D face depth image. Then SVM algorithm is used to classify the information and recognize face.

The next paper "Brain inspired face recognition: A computational framework" presents a new proposal for an efficient model for face recognition which is inspired by face recognition mechanism of the brain. Uses 3 main features: histogram of oriented gradients (HOG), local binary pattern(LBP) and principal components(PC's) extracted from target images.

The next paper "Design of a mobile face recognition system for visually impaired persons" presents the design and implementation of a face recognition system for the visually impaired through the use of mobile computing. This system is assisted by a server base support system.

In the next paper" Combining Data-driven and Model-driven Methods for Robust Facial Landmark Detection", This paper proposes an effective and robust approach for facial landmark detection by combining data- and model-driven methods. Firstly, a Fully Convolutional Network (FCN) is trained to compute response maps of all facial landmark points.

In the next paper "Facial Recognition in Public Areas",we are comparing or matching a face of a person who we want to detect, with the video which is recorded through CCTV. There are certain algorithms to detect faces from video like through HAAR cascades, eigenface, fisher face, etc. open-source computer vision library is used for facial recognition.

The next paper "Facial Emotion Recognition of Students using Convolutional Neural Network" an implements emotion recognition system in education by analysing students' facial expressions based on Convolutional Neural Network (CNN). It consists of multistage image processing to extract feature representations. The system includes three phases: face detection, normalization and emotion recognition that should be one of these seven emotions: neutral, anger, fear, sadness, happiness, surprise and disgust.

In the next paper "Intelligent facial expression recognition and classification using optimal deep transfer learning model" a novel IFER-DTFL technique is developed for the detection and classification of facial expressions. The proposed IFER-DTFL technique involved a three-stage process such as Mask RCNN-based face detection, Adam optimizer-based hyperparameter optimization, DenseNet121-based feature extraction, and WKELM-based classification.

In the next paper "Improved Deep CNN-based Two Stream Super Resolution and Hybrid Deep Model-based Facial Emotion Recognition" The proposed system is the facial emotion recognition methodology that includes 3 phases of operation, which are the Super-resolution phase, Facial emotion recognition as well as Classification phase

The next paper "Human Emotion Detection using Machine Learning Techniques"A monitoring system for elderly people which is based on technology involving recognizing emotions from video images. Our proposed system includes video analysis technology which includes the data from a video adopted to realize monitoring elders' living conditions in real-time.

the next paper "Context-Aware Emotion Recognition Based on Visual Relationship Detection" the emotional state prediction method based on visual relationship detection between the main target and the adjacent objects from the background. Specifically, we utilize both the spatial and semantic features of objects in the scene to calculate the influences of all context-related elements and their properties of impact (positive, neg-

ative, or neutral) on the main subject by a modified attention mechanism. After that, the model incorporates those features with scene context and body features of the target person to predict their emotional states. Our experimental results achieve state-of-the-art performance on the CAER-S dataset and competitive results on the EMOTIC benchmark.

The next paper Intelligent facial emotion recognition based on Hybrid whale optimization algorithm and sine cosine algorithm utilises 2 Optimization Algorithm and incorporates them in Multi-SVNN to detect facial emotions.the Sine Cosine Algorithm in the Whale Optimization Algorithm optimization (WOA-SCA)process to enhance global convergence and improve enactment. The paper intends an enhanced approach to identifying the human appearances dependant on their routine developments, for which the person expression is in use as of the ADFES dataset.

The next paper Emotion recognition in conversations with emotion shift detection based on multi-task learning propose a multi-task learning model ESD–ERC, which comprises the auxiliary task of Emotion Shift Detection (ESD) and the main task of ERC. The model exploits a shared BERT-based encoder to extract the unified emotion semantic representations, obtains the emotion shift representations through ESD based on Bi-directional Long Short-Term Memory Neural Network and feeds the emotion semantic representations concatenated with the emotion shift representations into the context level Transformer with positional encoding for ERC.• The proposed ESD–ERC is the first work to effectively exploit emotion shift to improve the accuracy of ERC. Then ESD–ERC adopts a multi-task learning framework and employs ESD as the auxiliary task, providing explicit guidance for the main task ERC.Multi-task learning model ESD–ERC, which exploits Emotion Shift Detection (ESD) as an auxiliary task to assist in the completion of ERC.

The next paper Amygdala-Inspired Affective Computing: to Realize Personalized Intracranial Emotions with Accurately Observed External Emotions uses CNN, the MLNN to recognize facial emotions using an amygdala-inspired affective computing framework.The goal is to realize personalized affective computing capable of recognizing both the time-related process emotion and instantaneous emergency emotion of humans. Similar to the amygdala, the instantaneous emergency emotion is first computed more quickly in a low-redundancy convolutional neural network compressed by pruning and weight sharing with hashing trick. Then, the real-time process emotion is identified more accurately by the memory-level neural networks.

In the next paper "CNN-Based Object Recognition and Tracking System to Assist Visually Impaired People" The proposed system comprises a raspberry pi digital signal processing (DSP) board with GSM and a global positing system (GPS) module, headphones, and a camera. DSP captures a live feed from the video camera and passes it to the object detection and recognition module (CNN model).

In the next paper" Sight-to-Sound Human-Machine Interface for Guiding and Navigating Visually Impaired People", a sight-to-sound human-machine interface (STSHMI), a novel machine vision guidance system that enables visually impaired people to navigate with instantaneous and intuitive responses. The proposed STS-HMI system extracts visual context from scenes and converts them into binaural acoustic cues for users to establish cognitive maps. A series of experiments were conducted to evaluate the performance of the STS-HMI system in a complex environment with difficult navigation paths. The experimental results confirm that the STS-HMI system improves visually impaired people's mobility with minimal effort.

In the next paper Emotional Voice Conversion Using Multitask Learning with TextTo-Speech utilizes multitask learning with text-to-speech (TTS). By using multitasking learning, VC is expected to capture linguistic information and preserve training stability. This method does not require an explicit alignment for capturing abundant text information. Furthermore, it uses emotional voice conversion. Given a style reference speech, the style encoder extracts only emotional information and removes linguistic contents. The extracted emotion embedding is then injected into the decoder to generate various emotions.

In the next paper Design and Construction of Electronic Aid for Visually Impaired People electronic device to assist visually impaired people with obstacle-free pathfinding. It consists of six ultrasonic sensors, a wet floor detector sensor, a step-down button, microcontroller circuits, four vibration motors, and a battery for the power supply. An ultrasonic sensor is used to send out high-frequency sound waves and record the time it takes for the reflected sound waves to return. An object may be directly in front of the transmitter or at an angle for the signal to be reflected and received by the ultrasonic sensor.

IV. Proposed Model

This project proposes to develop a system to aid the blind and visually impaired in ALL types of social interactions. By using face recognition technology and emotion detection, the system will help to identify people and detect their behaviour. The aim is to design a system to aid the visually impaired in understanding their surroundings.

First, the system recognises the face of the individual and then detects the emotion((Angry, Disgust, Fear, Happy, Sad, Surprise, Neutral).) on the said face. Finally, the detected emotion which is shown as text gets converted into speech and announced to the visually impaired person. This system helps people who are blind or have low vision, to recognize people and understand their emotions in a variety of social interactions.

1.Facial Recognition using CNN

A Convolutional Neural Network (CNN) model uses different layers like the convolutional, pooling and fully connected layers to detect a face by learning to recognize patterns in the pixel values of images. It detects faces by learning to recognize patterns in the input images that correspond to facial features, such as eyes, nose, mouth, and overall face shape. A

dataset can be manually created by taking multiple images of an individual. For face recognition, the dataset should include images of individuals' faces, along with their corresponding names The images should be of good quality, with clear and consistent lighting, and the faces should be aligned and centred in the images. The CNN learns to recognize patterns in the feature maps that are indicative of a face, such as the location of the eyes, nose, mouth, and other facial landmarks.

2. Emotion Detection

Emotion detection using the FER (Facial Expression Recognition) 2013 dataset involves training a machine learning model to classify facial expressions into different emotion categories. The FER 2013 dataset is a benchmark dataset for facial expression recognition, containing 35,887 grayscale images of size 48x48 pixels, labeled with seven emotion categories: anger, disgust, fear, happiness, sadness, surprise, and neutral.To detect emotion in an image, the algorithm first processes the image to extract relevant features, such as facial expressions and patterns. This is done using the convolutional and pooling layers of the CNN.The extracted features are then used to make a prediction on the emotion category using the fully connected layers of the CNN. The output of the fully connected layers is a probability distribution over the seven emotion categories, with the highest probability corresponding to the predicted emotion.

The prediction is then compared to the actual label of the image, and the accuracy of the algorithm is evaluated.

3.Text to Speech

Text-to-speech can be particularly useful for blind individuals, as it allows them to perceive and understand emotions in text-based content. Once the emotion has been detected the next step is to convert the text into spoken words.Finally the audio output is received by the user. Blind individuals can listen to the audio output to understand the emotion.

V. Implementation details

To train our CNN architecture, we used the FER2013 database . Faces in the database have been automatically normalized to 48×48 pixels. The FER2013 database contains 35887 images (28709 training images, 3589 validation images and 3589 test images) with 7 expression labels. The number of images for every emotion is represented in the table below.

Emotion label	Emotion	Number of image
0	Angry	4593
1	Disgust	547
2	Fear	5121
3	Happy	8989
4	Sad	6077
5	Surprise	4002
6	Neutral	6198

Fig. 1. The number of image for each emotion of FER 2013 dataset

CNN Implementaion

We used OpenCV library to capture live frames from web camera and to detect the emotion of individuals using the Haar Cascade method. Finally included text to speech to this so that the visually impaired user is able to understand the emotion of the individual. Haar Cascades uses the Adaboost learning algorithm .The Adaboost learning algorithm chose a few number of significant features from a large set in order to provide an effective result of classifiers. We built a Convolutional Neural Network model using TensorFlow.

VI. Conclusion

The system recognizes faces from individuals from input images using Haar-like detector and classifies them into seven facial expressions: surprise, fear, disgust, sad, happy, angry and neutral. The proposed model achieved an accuracy rate of 68 percent on FER 2013 database.The facial expression recognition system can help the visually impaired to be able to detect emotion, recognise the face and allows him to understand the emotion by text to speech One of the many challenges faced by the visually impaired is their inability to recognize the faces of individuals and understand the personality of people when they encounter them in their daily lives. The proposed system improves the access, integration, and independence of blind or severely visually impaired individuals in the workplace or educational settings. Through this system of facial recognition and emotion detection, becomes able to assist and work with the environment and make it a better place for the blind.

Objectives:-

• Create a system that aids the visually impaired

• Enables the blind to be able to detect the faces and emotions of people

Acknowledgment

We want to thank everyone who gave us the opportunity to finish this thesis in the most sincere way possible. We would also want to express our gratitude to Ms. Jasmine Jolly, our project coordinator for the Final year, whose contribution of energising suggestions and support helped us to plan our project, particularly in drafting this paper. Additionally, we would like to express our gratitude to the department head, Dr. Remya K. Sasi,who invested her full effort in guiding the team in achieving the goal. We must be grateful for the advise provided by other departmental faculty members, especially in regards to our paper presentation, which has enhanced our presentation skills as a result of their comments and suggestions.

References

[1] Design and construction of electronic aid for visually impaired people. https: //ieeexplore.ieee.org/document/8291597, 2018

[2] A mobile face recognition system dedicated to visually impaired people. https: //ieeexplore.ieee.org/document/8466782, 2018.

[3] Facial emotion recognition of students. https://ieeexplore.ieee.org/ document/8942386, 2019.

[4] Combining data-driven and model-driven methods for robust facial landmark detection.
https: //www.researchgate.net/ publication/311222440 2020.

[5] 3d face recognition. https://www.sciencedirect.com
[6] Converts sight to sound for visually impaired people. https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=9216032, 2020.
[7] Design and evaluation of real-time face recognition system. https://www.sciencedirect.com/science/article/pii/S1877050920311583?ref=pdfdownloadfr=RR-2=77516b9c6bed854f, 2020.
[8] Emotional voice conversion using multitask learning with text-to-speech. https://ieeexplore.ieee.org/document/9053255, 2020
[9] Human emotion detection using machine learning techniques. https: //www.researchgate.net/publication/341782389HumanEmotion DetectionusingMachineLearningTechniques, 2020.
[10] An improved face recognition system and its applications. https://www.sciencedirect.com/science/article/pii/S2590005619300141?ref=pdfdownloadfr=RR-2=77516c7349d8854f, 2020.
[11] Quantitative personality prediction. https://ieeexplore.ieee.org/ document/9143398, 2020.
[12] to realize personalized intracranial emotions with accurately observed external emotion. https://ieeexplore.ieee.org/abstract/document/8820765, 2020.
[13] Context aware emotion recognition. https://ieeexplore.ieee.org/ document/9461727, 2021.
[14] Intelligent facial expression recognition and classification. https://www.sciencedirect.com/science/article/pii/S0262885622002128, 2021.
[15] Prediction of the big five personality traits using static facial images. https: //ieeexplore.ieee.org/document/9420736, 2021.
[16] Brain inspired face recognition. https://www.sciencedirect.com/science/article/abs/pii/S1389041722000687?fr=RR-2ref=pdfdownload=77516d6e8944854f, 2022.
[17] Cnn-based object recognition and tracking system to assist visually impaired people. https://ieeexplore.ieee.org/abstract/document/9698080, 2022.
[18] Emotion recognition in conversations with emotion shift detection based on multi-task learning. https://www.sciencedirect.com/science/article/pii/S0950705122004117, 2022.
[19] Intelligent facial emotion recognition based on hybrid whale optimization algorithm and sine cosine algorithm. https://www.sciencedirect.com/science/ article/pii/S0141933122002484?via3Dihub, 2022.
[20] Two stream super resolution and hybrid deepmodel-based facial emotion recognition. https://www.sciencedirect.com/science/article/pii/S0952197622004766, 2022.