EMOTION DETECTION USING ADABOOST AND CNN

1st Dr Sumathi Pawar
Information Science & Engineering
Nitte(DU), NMAMIT, Nitte
Karkala, Karnataka, India
pawarsumathi@gmail.com*

2nd Suma K
Computer Science & Engineering
Mangalore Institute of Technology and Engineering
Moodbidri, Karnataka, India
Sumaknair07@gmail.com

Abstract: **Recent advances in Artificial Intelligence techniques are largely to blame for the rise in interest in the recognition of facial expressions. The process of distinguishing different human emotions using facial expressions is known as emotion detection. The system picked up on emotions like sadness, happiness, rage, fear, surprise, neutrality, and contempt. This paper is focussed on detecting 7 emotions through Haar-cascade, Adaboost and Convolutional Neural Networks algorithms. Compared to other existing systems, CNN's emotion recognition improved accuracy while reducing space complexity. Lie detectors, robotics, and artistic expression are some more uses of this system.**

Keywords: CNN, Emotion Detection, KNN, SVM

I. INTRODUCTION

Emotion recognition is the process of identifying a face expression. This procedure calls for a feature extractor and classifier that provide labels based on feature categories. The "Emotion Detection" technology primarily recognises and identifies the emotions that people display through their facial expressions. This system uses facial expression photos as its source to identify emotions using classifiers. After the emotion is identified, the result is displayed as a label along with the category that the emotion would belong to.

This system builds a model from a facial expression image that classifies the emotions into the following basic human emotions: disgust, neutral, surprise, fear, angry, glad, and sad. Transfer learning is used to build the model, and the accuracy of the model will determine how well the results are appraised.

This system utilises the CNN algorithm, which takes an image as input, assigns weights to various elements of the image, and condenses the information into a flexible form for processing while preserving key information needed for reliable predictions. This system focuses on identifying different human emotions and uses the Convolution Neural Network algorithm to categorise face expressions based on facial landmarks. Emotions can typically be divided into seven categories: disgust, anger, happiness, fear, surprise, surprise, sadness, and neutral.

This system uses facial expressions to identify emotions, and the identified emotion is shown as a label along with the category to which it belongs. The Convolutional Neural Networks used by the system also produce accurate results. The literature review is the main focus of section 2, the methodology is the main focus of section 3, the implementation procedures are shown in section 4, and the results and conclusion are given in sections 5 and 6.

II. LITERATURE REVIEW

[1]The plan of an artificial intelligence (AI) system which identifies emotions from facial expressions is shown in this research. It recognises emotions from photos using CNN architecture with deep learning model. For the Japanese Female Facial Emotion (JAFFE) and Facial Emotion Recognition Challenge (FERC-2013) datasets, the utilized approach here confirmed accuracy rates of 70.14 percent and 98.65 percent, correspondingly.

[2]Understanding human feelings requires an understanding of emotions. The work found that many techniques are used for assessing emotions, including linguistic, visual, physiological and aural cues. This study has looked at speech because it is the most natural way to express emotion. This system makes use of the Emotional Speech by Song (RA VDESS) database and Ryerson Audio-Visual Database. These database has audio files of many people voicing various emotions. A comparison of diverse machine learning algorithms, utilizing speech data, was conducted to aid in the classification of different emotions.

[3] Human-Computer Interaction (HCI) method includes speech analysis and extraction of signals from speech called as speech emotion recognition (SER) is used. In contemporary times, deep learning approaches have emerged as an alternative to conventional Speech Emotion Recognition (SER) techniques.

[4] The lips and eyes are key regions of the face that hold significant importance in identifying facial expressions. The framework suggests capturing features of face in real-time and processing them to see if there are any facial training faces that resemble

2023 IEEE 2nd International Conference on Data, Decision and Systems (ICDDS) | 979-8-3503-1416-8/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICDDS59137.2023.10434694

any of the real-time input. The output of the suggested method is to show the name of the identified person and categorise the facial expression using the probability of the emotion.

[5] A field called Emotion Recognition of Human Face (HFER) looks at how diverse facial expressions reveal how people are feeling. Transfer Learning, Bidirectional Convolutional LSTM, Multiple Pipelines are compared using a variety of universal human emotions, including disgust, sadness, happiness, surprise, anger, fear, , neutrality and contempt. Image data sets that contain images of static with a diverse set of emotions are employed in the procedures.

[6] This research suggests a brand-new image-based face expression identification system. Within this investigation, two principal approaches were implemented: face detection and facial expression recognition. To mitigate the impact of appearance alterations, the face identification technique such as utilized Haar-like features and reset the region of interest are utilized.

[7]This system initially use the FER2013 databases to train the CNN network model, and in testing step, which used JAFFE and CK+ dataset. Accuracy scored was 92.05% and 98.13% respectively. For assessing the system's effectiveness, the system also track facial expressions in real-time. For experimental findings, ConvNet is able to reach 96% training accuracy. This demonstrates a remarkable advancement compared to previous models.

[8]Facial expression-based automatic emotion recognition is an intriguing study area which is presented and given as number of fields, including health and safety. Deep learning has been remarkably successful, and its various designs are now being used to improve performance.

[9] The learning method for FER needs training image examples which is assigned a specific emotion category. Authors undertaken an experimental investigation on the identification of facial emotions[10]. The fundamental steps of the FER system are included in the emotion recognition system. These involve taking an image, preprocessing it, finding faces, extracting features and classifying it. After the emotions are categorised, the system focuses on real-time webcam photos. This system can identify stressed people and then recommend music therapy to help them relax. The commonly acknowledged emotions taken into account for the experiments include joy, sorrow, surprise, fear, disgust, and anger.

[11] According to Human psychology, given data is analysed quickly when it is generated. This system designed for testing the attitude of human, whether it is negative or positive. Authors analysed difficulties involved in their work.

[12] The study includes EEG databases and EEG preparation techniques, featured techniques for feature extraction and selection. The system analysed the classification of both emotions including deep learning based and machine learning based. There is also a look at current trends in feature research and suggestions for new research areas.

III. METHODOLOGY

A detailed methods are discussed in this section and emotion detection process is tasked into following steps.

A. Face Recognition

Using Harr-cascade with OpenCV the face region is detected. Following steps shows required procedure.

- Load the image:
- Convert image to grayscale:
- Detect faces
- Draw rectangles around detected faces
- Display the image with faces highlighted

The supplied photos are initially put through pre-processing. Three steps are involved in pre-processing. Images must first be converted to grayscale in the first stage. In this instance, the coloured image is used as input and transformed to grayscale for simpler prediction. The Haarscade method is used to detect faces once the images have been made grayscale, as illustrated by the cascade of filters in equations 1 and 2.

A cascading set of phases is used to construct the Haar Cascade classifier. Each stage uses the AdaBoost algorithm to train a set of weak classifiers (Haar-like features). The cascade classifier attempts to swiftly reject non-face regions in a picture, hence it is created to be computationally efficient. It accomplishes this by using increasingly complex classifiers at each stage, and if an image region fails a stage, it is automatically dismissed as a non-face region without further processing.

B. Adaboost Algorithm

The AdaBoost algorithm iteratively updates sample weights and selects weak classifiers which perform better than random guessing. By combining multiple weak classifiers, AdaBoost constructs strong classifier that often has better generalization routine than individual weak classifiers, making it a powerful ensemble method for classification tasks.

Training dataset: $\{(x_1, y_1), (x_2, y_2), ..., (x_i, y_i)\}$, where x_i represents the input features and y_i is the corresponding binary class label (+1 or -1).

Number of iterations: T (the total number of weak classifiers to be trained).

1. Initialize Sample Weights: At the beginning, each sample is given an equal weight (w_i) such that $\sum w_i = 1$. These weights are used to control the importance of each sample in the training process.
2. Train Weak Classifier: At each iteration t (from 1 to T), training of a weak classifier $h_t(x)$ is done using training dataset with the current sample weights.
3. Compute Error: The weighted error (ε_t) of the weak classifier $h_t(x)$ is calculated as follows:

$$\varepsilon = \Sigma_i w_i * \ I(y_i \neq \ h_t(x_i)) \quad (1)$$

(Here indicator function I() returns '1' if the condition is true or '0' if false. The error ε_t represents the weighted misclassification rate of the weak classifier.)
4. Compute Classifier Weight: The weight (α_t) of the weak classifier $h_t(x)$ is computed as follows:

$$\alpha_t = 0.5 * ln\left(\left(1-\varepsilon_t\right)/\varepsilon_t\right) \quad (2)$$

5. Update Sample Weights: The sample weights are updated based on the classifier's performance (α_t) and whether the sample is correctly classified by the weak classifier ($I(y_i = h_t(x_i))$). The updated sample weights (w_i) are given by:

$$w_i\left(new\right) = w_i * exp\left(-\alpha_t * y_i * h_t\left(x_i\right)\right)/Z \quad (3)$$

Where Z is a normalization factor to ensure that the updated weights sum to 1

$$Z = \Sigma_i w_i * exp\left(-\alpha_t * y_i * h_t\left(x_i\right)\right) \quad (4)$$

Weights assigned in such a way that classification takes place correctly according to given weights.

6. Combine Weak Classifiers: The ultimate strong classifier, H(x), is derived by aggregating the predictions of all the weak classifiers, each with its corresponding weight.

$$H\left(x\right) = \sin\left(\Sigma_t * \alpha_t * h_t\left(x\right)\right) \quad (5)$$

(Where sign () function returns +1 if the argument is zero or more than 0, and -1 otherwise.)

7. Prediction: To make predictions on new unseen data, the final strong classifier H(x) in which the samples are classified as either +1 or -1, depending on the majority vote of the weak classifiers weighted by their respective α_t values.

$$h_{LP} = \left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right) \ = \frac{1}{\sqrt{2}}\{1,1\} \quad (6)$$

$$h_{HP} = \left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right) \ = \frac{1}{\sqrt{2}}\{1,-1\} \quad (7)$$

C. *Emotion Detection*

Following steps are used in Emotion Detection.

- Data Collection and Preparation
- Data Augmentation
- Model Architecture
- Training
- Model Evaluation and Fine-Tuning

After completing the face detection stage, the image is segmented into four sections (the right eye, left eye, nose, and mouth). Then, landmarks and feature data are extracted from the input photos using feature extraction methods. The results of the recognition are produced using a classifier after the features have been extracted. The classification process is the last step in the emotion detection system. After identifying an emotion, the classifier classifies the expression with one of the following categories: disgust, surprise, sad, glad, angry, fear, or neutral. Section IV has a full description of the implementation.

This method makes use of the dataset Fer-2013 from Kaggle. It features grayscale portraits of 48 by 48 pixel faces with descriptions of their moods. This dataset contains 7 emotions: disgusted = 0, disgusted = 1, fear = 2, joyful = 3, sad = 4, surprised = 5, and neutral = 6. Three columns make up this dataset: usage, mood, and pixels. While the pixels column contains pixels in the form of a string, separated by spaces, the usage field denotes training or testing. Emotions are listed in the emotion column as integers.

D. *Convolutional Neural Network*

CNN's deep learning algorithm has the ability to capture photos and recognise one image from another. A CNN's main job is to convert images into a more digestible format while maintaining the essential details needed for precise predictions and facilitating processing simplicity.

- **Convolution Operation:** Convolution entails computing element-wise multiplications and summations while a filter—also referred to as a kernel—being dragged across the input data (such as an image), is the fundamental operation of CNNs. The convolution of an input picture I with a filter K is mathematically given by:

$$(I * K)(x, y) = \sum_i \sum_j I(x + i, y + j) * K(i, j)$$
(8)

• Stride: The amount of the filter's shift following each convolution operation is determined by the stride. The filter S locations are moved at each step when a stride of S is used.

• Padding: To maintain the input image's spatial dimensions after convolution, padding entails adding additional pixels. This assists in avoiding situations when the output is much smaller than the input.
• Pooling Operation: By choosing a representative value from a collection of nearby pixels, pooling decreases the spatial dimensions of the input. Max-pooling and average-pooling are two common pooling techniques.
• Activation Function: ReLU (Rectified Linear Unit) activation functions add non-linearity to the CNN model. ReLU aids the network in learning complex patterns by having a mathematical definition of f(x)=max(0,x).
Each neuron in a completely connected layer is linked to every neuron in the layer above it and the layer below it. Matrix multiplication and addition are used to calculate the output of a completely connected layer.
Backpropagation is a significant mathematical procedure used during the training of neural networks to compute gradients and update the weights of the network. The error is spread back across the network using the chain rule of calculus.

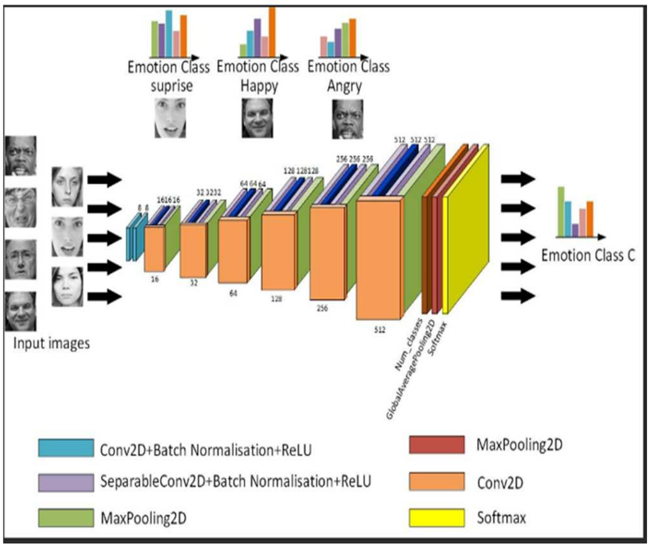

Fig 1 Convolutional Neural Networks(CNN)[14]

As shown in figure 1, the photographs on CNN have gone through a lengthy processing process. The convolutional layer receives the original input image. The features of the dataset are extracted using this layer. The supplied photographs are subjected to a series of teachable filters. The output layer gives a feature map. The essential features are enhanced with the use of the activation feature. Activation is used to condense the data set and speed up calculation. The output layer completes the classification or regression operation using feature maps that are transferred from hidden layers.

IV STEPS OF IMPLEMENTATION

Figure 2 depicts the implementation of this system's outline.

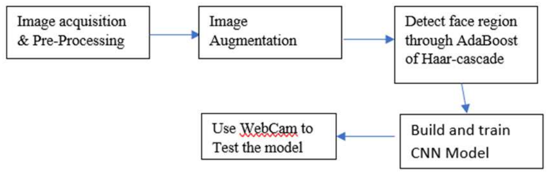

Fig 2 Implementation steps

A. Preprocessing Data

• Grayscale Conversion: To convert a coloured image to greyscale, system can use the luminance method, which takes the average of the RGB values for each pixel:

$$Gray=(0.299*R)+(0.587*G)+(0.114*B) \quad (8)$$

• Resizing: When resizing an image to a particular width and height while preserving its aspect ratio, the system can employ various interpolation methods, such as bilinear or bicubic interpolation. These methods help maintain the visual quality of the image during the resizing process.

• Normalization: This process is used to change the pixel values to a specific range [0, 1] to facilitate training neural networks or other machine learning algorithms. The formula for min-max normalization is:

$$X_{normalized} = (X - X_{min}) / (X_{max} - X_{min}) \quad (9)$$

• Standardization: Standardization transforms pixel values to a mean of 0 and a standard deviation to 1. This is often used in machine learning models that require input features with similar scales.

$$X_{standardized} = (X - mean) / standarddeviation \quad (10)$$

• Noise Reduction: Various techniques like Gaussian blurring or bilateral filtering are be used to minimize the noise of the image. The formulas for some filters can be quite complex, involving convolution operations.

• Histogram Equalization: This is a technique employed to increase the contrast of an image by reallocating its values, thereby making the image more visually distinct and vibrant. The formula involves computing the cumulative distribution function (CDF) of the image and then scaling the pixel values accordingly.

• Edge Detection: Techniques like Sobel operator or edge detection technique of Canny is used to find the edges in an image. These methods involve convolving the image with specific filter kernels.

- Thresholding: Thresholding is employed to transform a gray scale image into a binary image by assigning specific values (e.g., 0 and 1) to pixel values above or below a predefined threshold.

B Image Augmentation

[9]Image augmentation applies numerous transformations to existing pictures to create new, slightly altered versions of the original data. Augmentation helps in increasing the size and diversity of the training dataset, which can recover the model's generalization and robustness.

The mathematics behind image augmentation can vary based on the specific transformations applied. Some common image augmentation methods for facial emotion detection include:

- Rotation: Rotating the image by a certain angle (θ) can be represented using an affine transformation matrix:

$$\begin{pmatrix} cos(\theta) & -sin(\theta) & 0 \\ sin(\theta) & cos(\theta) & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

- Scaling: Scaling the image can be represented by a scaling matrix:

$$\begin{pmatrix} S_x & 0 & 0 \\ 0 & S_y & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

Here S_x and S_y are scaling factors in the x and y directions.

- Translation: Translating an image can be represented by a translation matrix:

$$\begin{pmatrix} 1 & 0 & T_X \\ 0 & 1 & T_Y \\ 0 & 0 & 1 \end{pmatrix}$$

- Shearing: Shearing the image in the x or y direction can be represented using shearing matrices:

Shearing in x-direction:

$$\begin{pmatrix} 1 & Shx & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

C. Fitting the Generator to the Data

After fitting the data to picture generator, it is required to generate data with a batch size 64 using the batch size setting.

- Randomization: To increase the diversity of augmented data, randomization plays a vital role. Randomly selecting transformation parameters (e.g., random angles for rotation, random scales, random flips, etc.) adds variability to the generated data.
- Normalization: It's important to normalize the image data previously feeding it into the model. This usually involves scaling of pixel values in the range [0, 1] or [-1, 1] depending on the input requirements of the model.
- One-Hot Encoding: According to categorical emotion labels (e.g., happy, sad, angry, etc.), the system required to transform them into numerical values or one-hot encode them before providing them to the model during training.
- Generator Function/Class: This generator function or class is responsible for generating augmented data batches on-the-fly during training.

D. Emotion Detection Model Training and Testing

KerasApIs are used for implementing the model. The mathematics behind compiling a facial emotion detection model involve specifying the loss function, optimization algorithm and evaluation metrics. If the model's accuracy will not increase after a few epochs, the learning rate is changed.

For 100 epochs of training, it takes at least 20 minutes and image captured using Webcam is tested by passing through this model. Accuracy of the performance is calculated.

V. RESULT

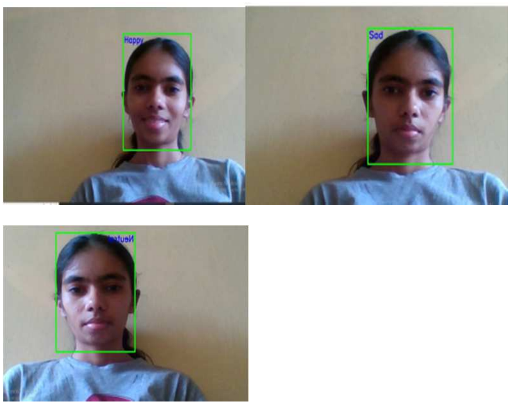

Fig 3 Images of emotion detection with labels.

This system is tested with image of a student showing different emotions and system is able to identify emotions of any student with 7 given feelings.

Table 1 CF of Emotion Prediction

	Average of Predicted Wrong	Average of Predicted Properly
Actual yes	FP 8	TP 142
Actual no	TN 22	FN 3

Table 1 gives confusion matrix for different emotions which is calculated with average of predicted correctly and incorrectly is given.

Loss Function: The loss function calculates the difference between the intended result and the projected result. Typical loss functions for classification problems include mean squared error, categorical cross-entropy, etc.

This is a multi class problem and equation 11 is used to calculate cross entropy loss. The formula is:

$$Categorical\ Cross\_Entropy\_loss = -\sum_i y_{true,i} * \log(y_{pred,i}) \quad (11)$$

Hear 'I' is each class and after averaging the loss of each class this system got 14% loss and 86% of average accuracy.

Accuracy is calculated using equation 12

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (12)$$

Figure 4 shows graph of accuracy of emotion detection for each emotion and according to the result it is received different accuracy for different types of emotions. It is observed that average accuracy of the emotion detection is more than 80% for all types of emotions.

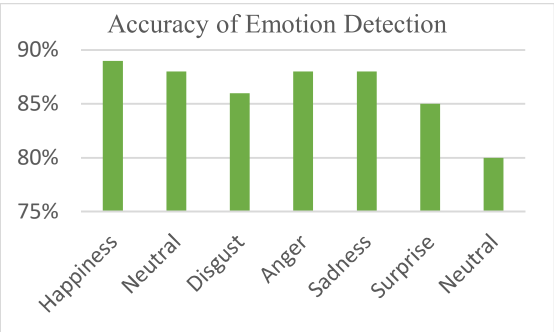

Fig 4 Accuracy of Emotion Detection

VI. CONCLUSION AND FUTURE WORK

The experiment demonstrates that the Adaboost and CNN-based model outperforms all other suggested models. With cutting-edge accuracy, this model categorises the photos based on the expressions on the faces. Furthermore, transfer learning and sophisticated feature extraction methods can improve the model's accuracy. As a future work, the model will be trained and evaluated on a larger range of data in order to not only fit the training dataset or lab condition photos but also to improve accuracy for recognising wild emotions. In addition, the model will be anticipated to bring real-world applications in education to enhance classroom interactions and learning.

References

[1]RaveendraBabu Vempati, Lakhan Dev Sharma, "A systematic review on human emotion recognition through electroencephalogram signals and artificial intelligence", Results in Engineering, Volume 18, 2023.

[2]R. Arya, D. Pandey, A. Kalia, B. J. Zachariah, I. Sandhu and D. Abrol, "Speech based Emotion Recognition using Machine Learning", *2021 IEEE Mysore Sub Section International Conference (MysuruCon)*, Hassan, India, 2021, pp. 613-617.

[3] R. A. Khalil, E. Jones, M. I. Babar, T. Jan, M. H. Zafar and T. Alhussain, "Speech Emotion Recognition Using Deep Learning Techniques: A Review," in *IEEE Access*, vol. 7, pp. 117327-117345, 2019.

[4] N. Pratheeksha Hegde, C. Shetty, B. Dhananjaya, Deepa, N. Rashmi and H. Sarojadevi, "Face and Emotion Recognition in Real Time using Machine Learning" 2022 7th International Conference on Communication and Electronics Systems (ICCES), Coimbatore, India, 2022, pp. 1018-1025.

[5] S. Yuvaraj, J. V. Franklin, V. S. Prakash, A. Anandaraj and R. Subha, "A Survey on Human Face Emotion Recognition using Machine Learning Models", *2023 9th International Conference on Advanced Computing and Communication Systems (ICACCS)*, Coimbatore, India, 2023, pp. 1993-1996.

[6] S. Kim, G. H. An and S. -J. Kang, "Facial expression recognition system using machine learning," *2017 International SoC Design Conference (ISOCC)*, Seoul, Korea (South), 2017, pp. 266-267.

[7] Debnath, T., Reza, M.M., Rahman, A. *et al.* "Four-layer ConvNet to facial emotion recognition with minimal epochs and the significance of data diversity," *Sci Rep* **12**, 6991 (2022).

[8] Wafa Mellouk, Wahida Handouzi, "Facial emotion recognition using deep learning: review and insights," Procedia Computer Science, Volume 175, 2020, Pages 689-694.

[9] "Recognizing human facial expressions with machine learning| Thoughtworks," https://www.thoughtworks .com/insights/articles/recognizing-human-facial-expressions-machine-learning, Last Retrieved on 1/10/2022.

[10] Renuka Deshmukh, Shilpa Paygude, Vandana Jagtap, "Facial Emotion Recognition System through Machine Learning approach," ICCS 2017, May 2017.

[11] Nandwani, P., Verma, R. "A review on sentiment analysis and emotion detection from text. Soc. Netw. Anal. Min," 11, 81 (2021).

[12] A. Jaiswal, A. Krishnama Raju and S. Deb, "Facial Emotion Detection Using Deep Learning," *2020 International Conference for Emerging Technology (INCET)*, Belgaum, India, 2020, pp. 1-5.

[13] B.C. Ko, "A Brief Review of Facial Emotion Recognition based on Visual Information", Sensors, Vol. 18, No. 2, 2018, pp. 401-421.

[14] Sumathi Pawar, K. Pandith, M. G. Rao, N. N. Chiplunkar and R. Samaga, "Sign Language Recognition Using Keypoints Through DTW," 2022 Fourth International Conference on Cognitive Computing and Information Processing (CCIP), Bengaluru, India, 2022, pp. 1-6.

[15] face emotion recognition using CNN,OpenCV and Python, https://45.153.231.124/face-emotion-recognition-using-cnn-opencv-and-python/, Posted on October 13.2023.