Web application for Facial Expression, gender and age recognition

Mylavarapu Soumya
UG student
Department of Information Technology
Vardhaman College of Engineering
Hyderabad, India
soumyamylavarapu21@gmail.com

Alety Manish
UG student
Department of Information Technology
Vardhaman College of Engineering
Hyderabad, India
manishreddyalety2@gmail.com

Mrs N. Sujata Gupta
Assistant Professor
Department of Information Technology
Vardhaman College of Engineering
Hyderabad, India
sujatagupta1429@vardhaman.org

***Abstract*—In the subject of human-machine interaction, human facial recognition is crucial. Our goal is to anticipate the expression of a human face, gender, and age as quickly and accurately as possible in real time. Understanding human behaviour, detecting mental diseases, and creating synthetic human expressions are only few of the applications of automatic human facial recognition. Age, gender, and emotion prediction are all used to assist salespeople better understand their customers. Deep learning techniques such as Convolutional Neural Network are utilised to design model and predict emotion, age, and gender, using the Haar-Cascade frontal face algorithm to detect the face. This model can predict from both still photos and video in real time.The process of detecting the face, pre-processing, feature extraction, and prediction of expression, gender, and age is carried out in steps.**

***Index Terms*—Convolutional Neural Network, Haar-Cascade frontal face algorithm.**

I. Introduction

Our ambitions have risen since the arrival of modern technology, and they have no limitations. In today's world, there is a wide range of research going on in the field of digital imaging and image processing. The rate of progress has been exponential, and it continues to rise.

Facial expression is observable indications of a person's effective state, cognitive activity, intention, personality, and psychopathology, which serves as a communication role in interpersonal connections. The seven major emotions that may be easily classified in human facial expressions are happy, sad, surprise, fear, anger, disgust, and neutral[1]. Our facial emotions are expressed by the activation of various sets of face muscles. These seemingly insignificant, yet complicated, signals in an expression often transmit a lot of information about our mental state. Age and gender classification can be especially helpful in several real-world applications including security and video surveillance, electronic customer relationship management, biometrics, electronic vending machines, human-computer interaction, entertainment, cosmetology, and forensic art.

The goal of this project is to develop an Automatic Facial Expression, age and gender Recognition System through web application that can recognise and categorise human facial photographs with a variety of expressions into seven different expression classes, 2 classes of gender(Male and Female) and 8 classes of ages i.e. (0-2), (4-6), (8-13), (15-20), (25-32), (38-43), (48-53), 60+
[2]

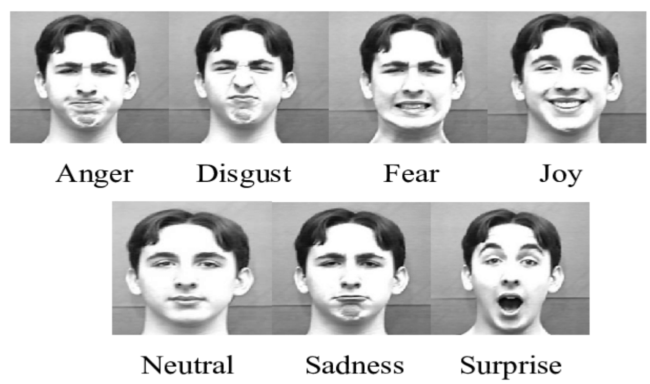

Fig. 1. Seven basic facial expressions

A. Problem definition

Face expressions communicate human emotions and intents, and developing an efficient and effective feature is an important part of the facial expression system. Nonverbal indicators are vital in interpersonal relationships, and facial expressions communicate them. In natural human-machine interfaces, as well as behavioural research and therapeutic practise, automatic facial expression detection can be a beneficial feature. An autonomous Facial Expression Recognition system must overcome challenges such as face identification and placement in a chaotic scenario, facial feature extraction, and facial emotion classification.

Convolution neural networks are used in this study to construct a facial expression recognition system. Anger, Disgust, Fear, Happy, Sad, Surprise, and 'Neutral' are the seven facial emotion categories that are used to classify facial photographs. The classifier is trained and tested using the Kaggle data set[1].

The expected gender can be either 'Male' or 'Female,' and the predicted age can be any of the following: (0–2), (4–6), (8–12), (15–20), (25–32), (38–43), (48–53), (60–100). (8 nodes in the final softmax layer). Because of elements such as cosmetics, lighting, obstacles, and facial expressions, determining an exact age from a single image is extremely difficult. As a result, rather than making this a regression problem, we make it a classification task.

B. Objective

The project's goal is to create a web application that uses a camera to capture a live human face and classify it into one of seven expressions, two ages, and eight age groups. We used the FER2013 dataset[3] for facial expressions and the Adience benchmark age and gender dataset[4] for age and gender recognition to build a CNN for facial expressions, age, and gender classification.

C. Limitations

The facial expression recognition system in this research accurately recognizes all of the expressions except disgust because the data set contains extremely few photos of that label and their training of that label is limited. It's exceedingly difficult to determine an accurate age from a single shot due to factors like lighting, impediments, and facial expressions. Random brightness and contrast variations added to RGB or grayscale train images improve gender and age group prediction accuracy in camera-based testing, according to our findings. In both training and test data, each photo has two labels: a gender label and an age group label. There are no gender or age labels on some of the photographs in the audience data collection.

II. Literature Survey

Facial Expressions Recognition Based on Cognition and Mapped Binary Patterns

A new expression recognition strategy based on cognition and mapped binary patterns is provided in this paper. They proved that eyes and mouth express greater emotion by referring to relevant studies of human cognition. To begin, the method relies on the LBP operator to extract facial outlines. Second, the pseudo-3-D model is used to divide the face area into six sub-regions based on facial emotion. The mapped LBP approach is used for feature extraction in the sub-regions and global facial expression images, and then two classifications, the support vector machine and softmax, are used with two types of emotion classification models, the basic emotion model and the circumplex emotion model. Finally, they compared the extension of the Cohn-Kanade (CK +) facial expression data set to the test data sets gathered from ten participants in a comparative experiment. The results of the experiments suggest that the approach can efficiently remove the image's confounding aspects. The circumplex emotion model produces significantly better results than the conventional emotional model.[5]

Artificial Neural Network Based Ensemble Approach for Multicultural Facial Expressions Analysis

The key problems to the facial expression identification system include variances in facial structure, facial appearance, and facial emotion representation due to cultural differences.These variations necessitate the need for multicultural facial expression analysis. This study presents several computational algorithms to handle these variations to get high expression recognition accuracy.[6]

For intercultural facial expression analysis, they presented an artificial neural network-based ensemble classifier. A multi-culture facial expression dataset is created by combining facial images from the Japanese female facial expression database, the Taiwanese facial expression image database, and the RadBoud faces database. The multicultural dataset's members come from four ethnic groups: Japanese, Taiwanese, Caucasians, and Moroccans. For facial feature representation, local binary patterns, uniform local binary patterns, and principal component analysis are used.

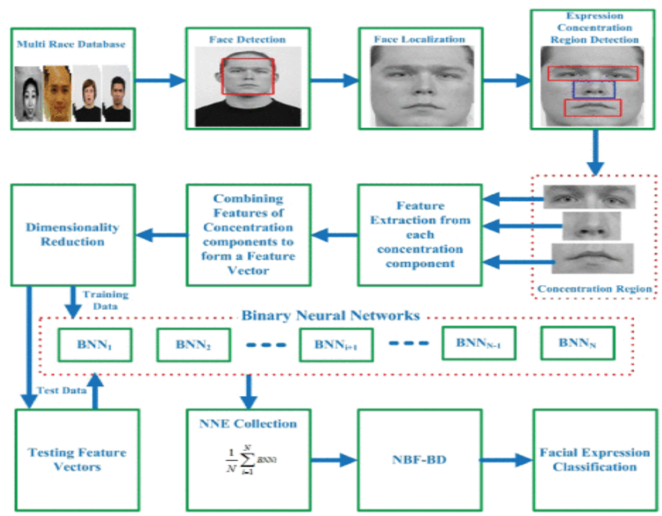

Fig. 2. Multi Culture Facial Expression Recognition Framework

Gender and Age recognition using speech

The following are some of the challenges that arise throughout the process of automatic speech estimate based on age: To begin, the speaker's weight, height, and mood affect speech, and these qualities may interact with age [1,-3]. Second, a large database is required for various speakers of various ages. Finally, because the majority of talks we hear in our daily lives take place in noisy situations, high-quality microphones and filters are essential. When different recording methods are used to record the same statement from the same speaker, the results vary.

Automatic gender and age recognition can be done in a variety of ways. Cepstral characteristics, such as Mel Frequency Cepstral Coefficients, are an example (MFCC). For the purpose of age recognition. With recorded data, MFCC is known for delivering poor gender and age categorization results. To avoid this issue, the MFCC features are improved by examining the parameters that influence the feature extraction process. MFCC has been employed in a variety of speech applications, including voice recognition and language recognition, and another acoustic characteristic that can be derived is format frequency[7].

Emotion Recognition using body movements

They propose a two-layer feature selection framework for emotion categorization based on a large number of body movement characteristics. The feature selection framework was utilised to correctly identify five basic emotions: happy, sad, fear, anger, and neutral. To eliminate unnecessary features in the first layer, a novel mix of Analysis of Variance (ANOVA) and Multivariate Analysis of Variance (MANOVA) was used. A binary chromosome-based genetic algorithm was presented in the second layer to pick a feature subset from the relevant array of features that increases emotion recognition rate. Different action scenarios were investigated, including walking and sitting actions, as well as an action-independent instance.

Based on the findings of the experiments, the suggested emotion recognition system outperformed all state-of-the-art systems in terms of emotion recognition rate. The suggested system demonstrated great accuracy by achieving recognition accuracy of 90.0 percent while walking, 96.0 percent while sitting, and 86.66 percent in an action-independent scenario[8].

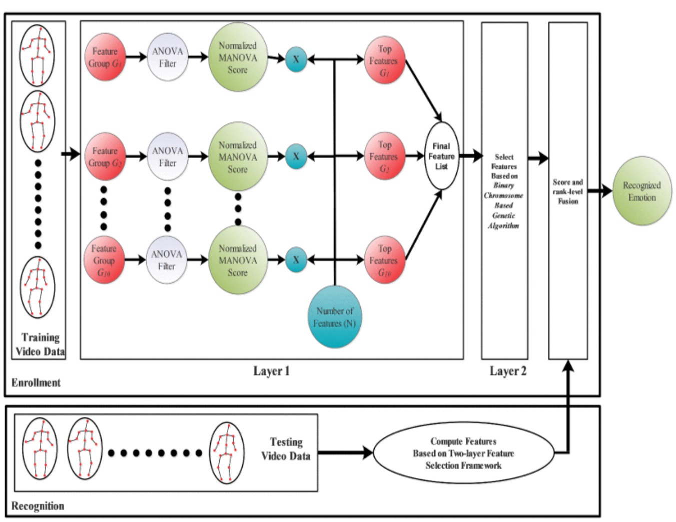

Fig. 3. An overview of the proposed framework for emotion recognition from body motion.

III. Methodology

A. *Convolutional Neural Network*

A Convolutional Neural Network (CNN) is a Deep Learning method that can take an input image, assign priority to distinct objects in the image using learnable weights and biases, and distinguish one from the other[9, 10, 11]. When compared to other classification methods, the amount of pre-processing required by a ConvNet is significantly less. While basic approaches require hand-engineering of filters, ConvNets can learn these characteristics with enough training[12].

In its most basic form, a ConvNet architecture is a list of Layers that turn the image volume into an output volume containing the class scores.There are a few different types of Layers (the most common being CONV/FC/RELU/POOL). Each Layer takes an input 3D volume and, using a differentiable function, converts it to an output 3D volume.There may or may not be parameters for each Layer. Additional hyper parameters may or may not be present in each Layer (for example, CONV/FC/POOL have them, but RELU does not[13].

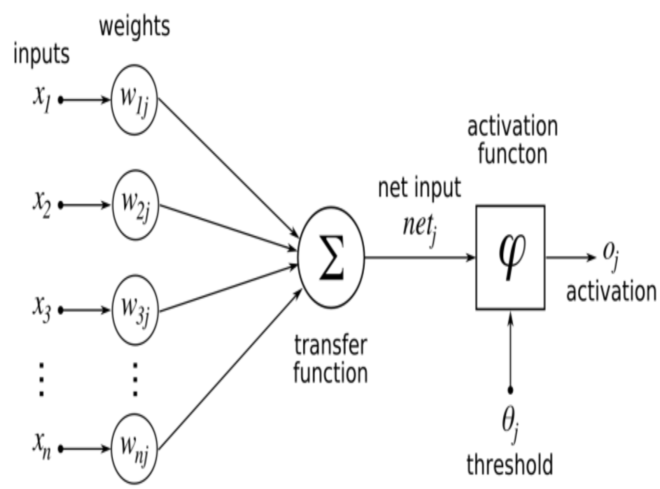

Fig. 4. Convolutional Neural Networks

B. *Haar Cascade Classifier*

Facial Recognition is everywhere. From the security camera on your front porch to the sensor on your iPhone X, facial recognition is everywhere. But, given the huge number of traits as input and the apparent similarities among humans, how does facial recognition operate to classify faces?

Haar classifiers were the first real-time face detector's classifiers. A Haar classifier, also known as a Haar cascade classifier, is a machine learning object recognition algorithm that can recognise objects in images and videos.Haar characteristics are used to determine how likely a given point is to be a part of an object. A combination of weak learners is used with boosting algorithms to give a strong prediction. Boosting methods are executed on multiple subsections of the input image using cascading classifiers[14, 15]. For Haar cascades, make sure to optimise against false negatives.Open CV can be used to implement Haar Cascade model. [16]

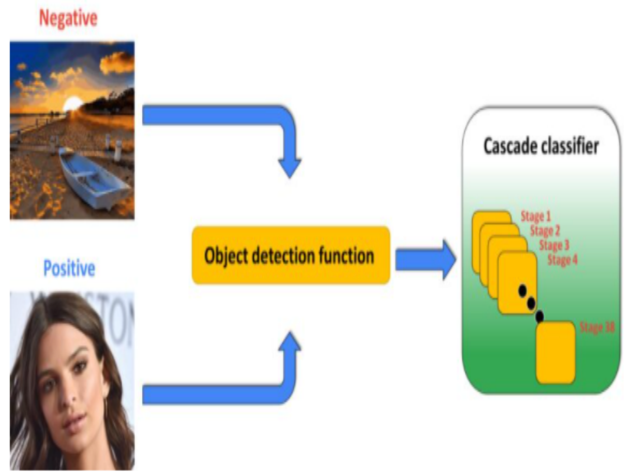

Fig. 5. Haar Cascade Classifier

C. Model

Deep learning is a common computer vision approach. Convolutional Neural Network (CNN) layers were chosen as the building blocks for our model architecture. When processing pictures, CNN's are known to mimic how the human brain functions.
A convolutional neural network's usual architecture includes an input layer, convolutional layers, some dense layers (also known as fully-connected layers), and an output layer. These are stacked layers that are organized linearly. The model is constructed as Sequential() in Keras, and further layers are added to build the architecture.

Input layer

Because the input layer's dimensions are pre-determined and fixed, the image must be pre-processed before being fed into it. For face detection in the image, we utilised Open CV, a computer vision package. OpenCV's haar-cascade frontalface default.xml file contains pre-trained filters and employs Adaboost to locate and clip the face quickly.

Convolutional layers

The numpy array is passed to the Convolution2D layer, where one of the hyper parameters is the number of filters. The collection of filters is one-of-a-kind, with weights that are produced at random. To construct a feature map, each filter, (3, 3) receptive field, slides across the source image with shared weights.In total, we have used 4 convolutional layers for expression classification and 3 convolutional layers for age and gender classification. Convolution produces feature maps that show how pixel values are improved. **Pooling** is a dimension reduction technique that is commonly used after one or more convolutional layers have been applied. When developing CNN's, this is a crucial stage because adding more convolutional layers can significantly increase processing time. We utilised the MaxPooling2D pooling approach, which uses (2, 2) windows across the feature map to maintain only the maximum pixel value and down-size image by 4.

Dense layers

Dense layers accept a large number of input features and transform them using trainable weights and layers. Forward propagation of training data is used to train these weights, followed by backward propagation of errors. By adjusting hyper-parameters like learning rate, we may regulate the training pace and complexity of the architecture.
Dropout is one way to avoid over-fitting and generalization on previously unknown data. During training, Dropout randomly picks a subset of nodes (typically less than 50 percent) and sets their weights to zero. This strategy efficiently controls the model's noise sensitivity during training while retaining the architecture's required complexity.

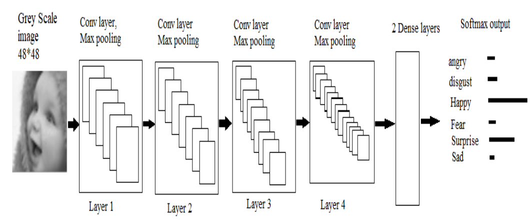

Fig. 6. Working model of expression

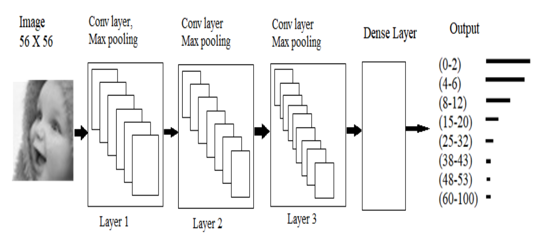

Fig. 7. Working model of age

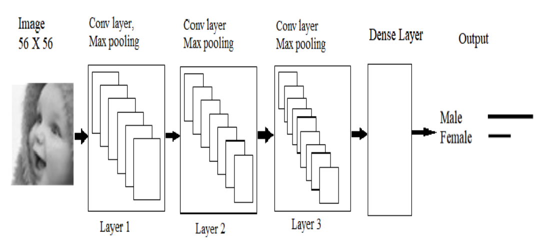

Fig. 8. Working model of gender

Output layer

At the output layer, we employed softmax instead of the sigmoid activation function. For each emotion class, this output appears as a likelihood. As a result, the model can display the probability composition of the emotions in the face in great detail. Finally, each image is classified into one of the seven basic expressions based on the emotion which has the highest probability, one of the two genders and one of eight classes of ages based on which has highest probability.

D. Experimental Results

Our model downsizes the training and testing images into 48x48 for expression and 56X56 for age and gender and then process. The model uses 4 convolutional layers and uses a haar cascade algorithm to detect the face and output the probabilities of all the seven expressions,two for gender and 8 for age and pick up the one with the highest probability for expression, age and gender .

Below are the steps used to lively detect the image through web camera and classify the detected face into one of the seven basic expressions[17], one of the two ages, one of the eight groups of age's[18].

1) Download the haarcascade_frontalface_default.xml and use the location where that is downloaded and store it in a variable (face_classifier).

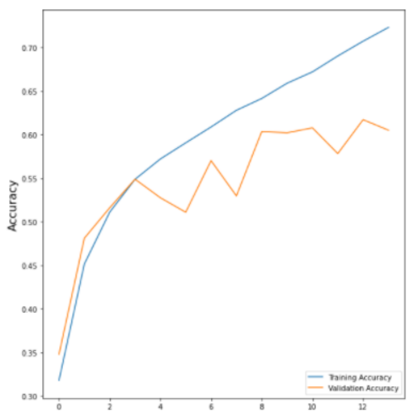

Fig. 9. Accuracy of model

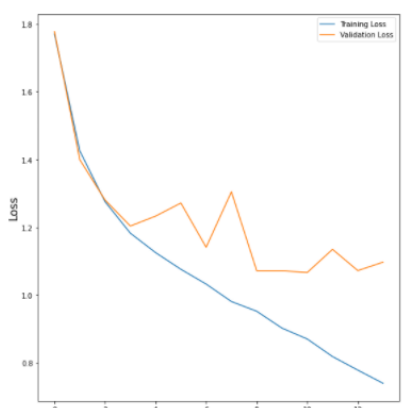

Fig. 10. Loss of model

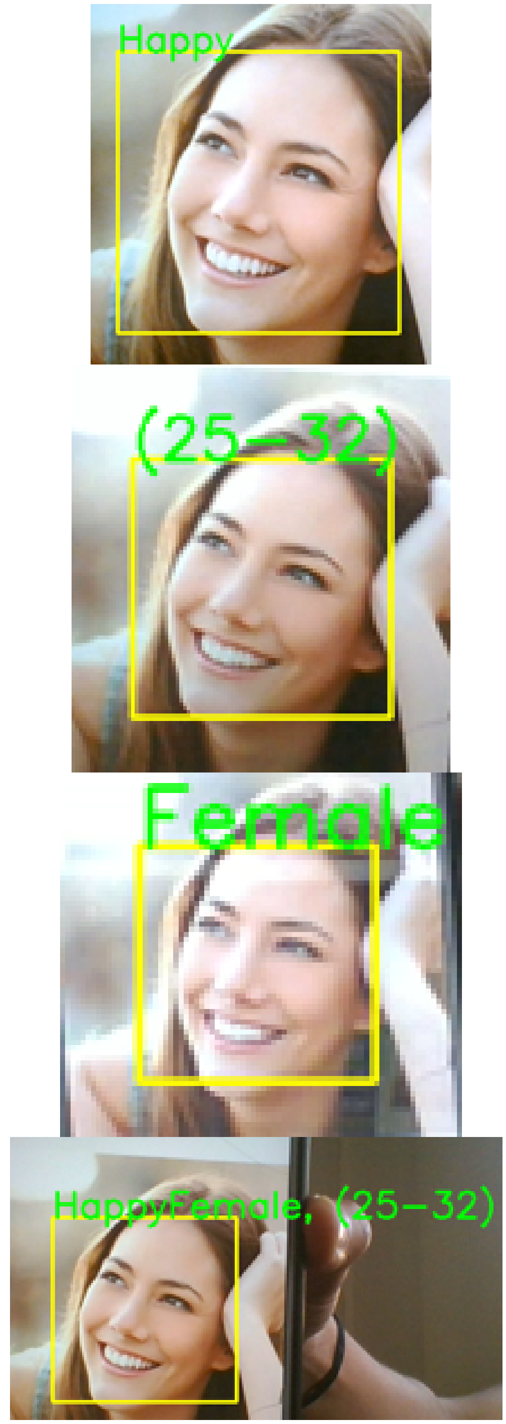

Fig. 11. Outputs

2) Use the location where the .h5 file created after running the model and store in a variable(classifier).
3) Use the openCV method VideoCapture(0) by cv2.VideoCapture, 0 is to capture from web camera of your currently working PC.
4) Draw a rectangle over the face
5) Get the label based on the emotion which gives the highest prop ability.

6) Put the text on the rectangle drawn
7) Press 'q' to exit.

E. Conclusion and Future work

When a model predicts an emotion or gender or age inaccurately, the correct label is frequently the second most likely feeling. The physiological properties of the human face that are relevant to diverse expressions such as happiness, sadness, fear, anger, surprise, and disgust are linked to geometrical structures that are reconstituted as the recognition system's basis matching template.

Future enhancements

In this experiment, we achieved an accuracy of around 70%, which is not terrible when compared to earlier models. However, there are several areas where we need to improve, such as

1) The number and layout of convolutional layers.
2) The quantity and arrangement of thick layers
3) In dense layers, the percentage of dropouts.

However, due to a lack of highly configured systems, we were unable to go deeper into dense neural networks because the system became extremely slow, and we will try to improve in these areas in the future. We would also like to train more databases into the system to improve the model's accuracy.

One can extend this project to take accept the input file as image and use it to classify one of the seven expression's, one of the two gender's and one of the eight ages group's.

References

[1] Heechul Jung, Sihaeng Lee, Sunjeong Park, Byungju Kim, Junmo Kim, Injae Lee, and Chunghyun Ahn. "Development of deep learning-based facial expression recognition system". In: *2015 21st Korea-Japan Joint Workshop on Frontiers of Computer Vision (FCV)*. 2015, pp. 1–4. DOI: 10.1109/FCV.2015.7103729.

[2] John Chris Kwong, Felan Garcia, Patricia Angela Abu, and Rosula Reyes. "Emotion Recognition via Facial Expression: Utilization of Numerous Feature Descriptors in Different Machine Learning Algorithms". In: Oct. 2018. DOI: 10.1109/TENCON.2018.8650192.

[3] Manas Sambare. "FER-2013 Dataset". In: *IEEE Access* 8 (2020).

[4] Alfredhhw. "Adience Dataset for age and gender classification Dataset". In: *IEEE Access* 8 (2020).

[5] Chao Qi, Min Li, Qiushi Wang, Huiquan Zhang, Jinling Xing, Zhifan Gao, and Huailing Zhang. "Facial Expressions Recognition Based on Cognition and Mapped Binary Patterns". In: *IEEE Access* 6 (2018), pp. 18795–18803. DOI: 10.1109/ACCESS.2018.2816044.

[6] Ghulam Ali, Amjad Ali, Farman Ali, Umar Draz, Fiaz Majeed, Sana Yasin, Tariq Ali, and Noman Haider. "Artificial Neural Network Based Ensemble Approach for Multicultural Facial Expressions Analysis". In: *IEEE Access* 8 (2020), pp. 134950–134963. DOI: 10.1109/ACCESS.2020.3009908.

[7] Huijuan Zhao and Ping Wang. "A Short Review of Age and Gender Recognition Based on Speech". In: *2019 IEEE 5th Intl Conference on Big Data Security on Cloud (BigDataSecurity), IEEE Intl Conference on High Performance and Smart Computing, (HPSC) and IEEE Intl Conference on Intelligent Data and Security (IDS)*. 2019, pp. 183–185. DOI: 10.1109/BigDataSecurity-HPSC-IDS.2019.00042.

[8] Ferdous Ahmed, A. S. M. Hossain Bari, and Marina L. Gavrilova. "Emotion Recognition From Body Movement". In: *IEEE Access* 8 (2020), pp. 11761–11781. DOI: 10.1109/ACCESS.2019.2963113.

[9] *What is a convolutional neural network?* URL: https://www.nvidia.com/en-us/glossary/data-science/convolutional-neural-network/.

[10] *Convolutional Neural Network*. 2019. URL: https://deepai.org/machine-learning-glossary-and-terms/convolutional-neural-network.

[11] By: IBM Cloud Education. *What are convolutional neural networks?* URL: https://www.ibm.com/cloud/learn/convolutional-neural-networks.

[12] .

[13] Youhui Tian. "Artificial Intelligence Image Recognition Method Based on Convolutional Neural Network Algorithm". In: *IEEE Access* 8 (2020), pp. 125731–125744. DOI: 10.1109/ACCESS.2020.3006097.

[14] *OpenCV Haar Cascade: Guide to opencv haar cascade*. 2021. URL: https://www.educba.com/opencv-haar-cascade/.

[15] Mahmoud Harmouch. *Haar cascade classifiers in opencv explained visually*. 2020. URL: https://medium.com/swlh/haar-cascade-classifiers-in-opencv-explained-visually-f608086fc42c.

[16] Priyanka Goel and Suneeta Agarwal. "Hybrid Approach of Haar Cascade Classifiers and Geometrical Properties of Facial Features Applied to Illumination Invariant Gender Classification System". In: *2012 International Conference on Computing Sciences*. 2012, pp. 132–136. DOI: 10.1109/ICCS.2012.40.

[17] Akshitmadan. *Emotion classification: CNN using Keras*. 2021. URL: https://www.kaggle.com/akshitmadan/emotion-classification-cnn-using-keras.

[18] Gil Levi and Tal Hassner. "Age and Gender Classification Using Convolutional Neural Networks". In: *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR) workshops*. 2015. URL: \url{https://osnathassner.github.io/talhassner/projects/cnn_agegender}.