Novel Deep Learning Techniques to Design the Model and Predict Facial Expression, Gender, and Age Recognition

N. Sujata Gupta[1(✉)], Saroja Kumar Rout[1], Viyyapu Lokeshwari Vinya[3], Koti Tejasvi[2], and Bhargavi Rani[3]

[1] Department of Information Technology, Vardhaman College of Engineering (Autonomous), Hyderabad, India
gsuji29@gmail.com
[2] Department of Computer Science and Engineering, Vardhaman College of Engineering (Autonomous), Hyderabad, India
[3] Department of Information Technology, Sarvajanik College of Engineering, Surat, Gujarat, India

Abstract. For computer and human 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 behavior, detecting mental diseases, and creating synthetic human expressions are only a few of the applications of automatic human facial recognition . Salespeople can employ age, gender, and emotional state prediction to help them better understand their consumers. Convolutional Neural Network one of the Deep Learning techniques is utilized to design the model and predict emotion, age, and gender, using the Haar-Cascade frontal face algorithm to detect the face. This model can predict from video in real-time. The 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. The process of detecting face, pre-processing, feature extraction, and the prediction of expression, gender, and age is carried out in steps.

Keywords: Convolution Neural Network · Haar-Cascade Classifier · Facial expression · Emotion

1 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 processing image. The rate of progress has been exponential, and it continues to rise. The facial expression of a person shows the person's mood, state of mind, thinking, 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 anger, disgust, fear, happiness, neutral, sadness, and surprise [1]. The

S. Nandan Mohanty et al. (Eds.): ICISML 2022, LNICST 471, pp. 345–354, 2023.
https://doi.org/10.1007/978-3-031-35081-8_29

activation of several sets of facial muscles expresses our facial emotions. These seemingly minor, but complex, signals in our expressions often convey a great deal about our mental state. Age and gender classification can be used in many areas in biometrics, security, video surveillance, communication between humans and computers, and in forensic.

The goal of this project is to develop an Automatic Facial Expression, age, and gender Recognition System through a web application that can recognize and categorize 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].

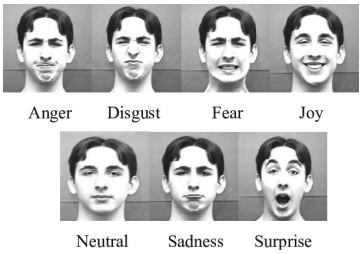

Fig. 1. Seven basic facial expressions.

1.1 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 human and computer interaction, 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, feature extraction from the face, 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 is either "male" or "female," and the expected age is either (0–2), (4–6), (8–12), (15–20), (25–32), (38–43), (48–53), or (60–100). In the final softmax activation function, there are eight nodes for age and two for gender. When we consider various features like lighting, makeup, and facial expressions, determining the precise age of a person instead of ranging it would be difficult. So, we have used a classification task instead of a regression task.

1.1.1 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 Audience benchmark age and gender dataset [4] for age and gender recognition to build CNN for facial expressions, age, and gender classification.

1.1.2 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 in that label is limited. It's exceedingly difficult to determine a precise age from a single photo due to factors like makeup, lighting, impediments, and facial expressions. According to our observations, the accuracy of detecting age and gender can be improved by increasing the random brightness and fluctuations in RGB. Each photo contains two labels in both training and test data: gender and age. Some of the images of the audience data collection, lack gender or age descriptors.

The rest of the paper is furnished into 3 sections. Section 2 represents the related research and the techniques used by the authors. The methodology of the proposed system is described in Sect. 3. Simulation and results have been depicted in Sect. 4. Section 5 Describes the conclusion and future direction of the research work.

2 Literature Review

This research paper proposed an approach to recognizing expressions from faces using binary patterns and cognition. They observed that the eyes and mouth are the main parts to classify an expression. The procedure begins by extracting facial contours using the LBP operator. They used a 3D model to divide the face into six sub-regions. They used the mapped LBP approach to recognize the expression from the sub-parts they divided. They have used a support-vector machine and a softmax activation function for two types of models. Finally, they compared the facial expression dataset Cohn-Kanade (CK +) with the test dataset in which they considered ten members. They observed that their model will remove the image's problematic elements. The expression model outperforms the traditional emotional model by a wide margin [5].

This project is to determine expression by building an artificial neural network based on a different culture. Variations in the appearance of the face, the structure of the face, and facial emotion representation due to cultural differences are the main problems with the facial emotion detection system. Because of these differences, multicultural face expression analysis is required. Several computational strategies are presented in this paper to address these changes and achieve excellent expression recognition accuracy [6].

For intercultural facial expression analysis, they presented an artificial neural network-based ensemble classifier. By merging facial images from the Japanese female facial expression database, the Taiwanese facial expression image database, and the Radboud faces database, a multi-culture facial expression dataset is constructed. Members of the multicultural dataset are Japanese, Taiwanese, Caucasians, and Moroccans, who represent four ethnic groupings. Local binary patterns, uniform local binary patterns, and principal component analysis are utilized to express facial features Figs. 1, 2, 3.

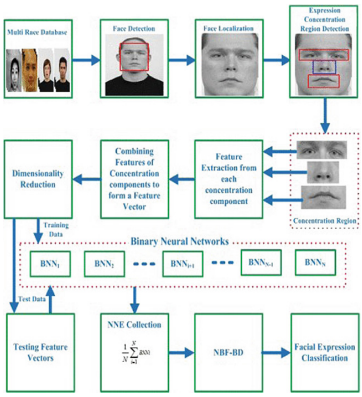

Fig. 2. Framework for Multi-Culture Facial Expression Recognition

They have developed a model that recognizes the voice and classifies it into age and gender. Many factors affect the process of automatic speech recognition like the speaker's weight, height, and speech based on a person's mood. There is also a requirement for a very large database that consists of a large number of speakers of different ages and genders. Since there will also be a problem of noise when trying to record the audio of the speaker, high-quality microphones and filters are also required. The result varies when different speakers are used to record the same statement from the same speaker.

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 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].

This project is to detect the emotions of body movements. They have used these body movement characteristics and developed a two-feature selection framework. They have considered only five basic features namely anger, happiness, sadness, fear, and neutrality. The first layer is the combination of Multivariate Analysis of Variance (MANOVA) and analysis of Variance (ANOVA) to remove unnecessary features for emotion detection. In the second layer, they used a binary chromosome-based genetic algorithm to pick the relevant features that improve in giving the correct label of emotion. Some of the movements they have considered are walking, sitting, and action-independent instances.

Based on experiments of various body movements, this model is outperformed in terms of emotion recognition rate. The accuracy of the action walking is 90%, 96% for sitting, and for other action-independent scenarios, the accuracy is 86.66 [8].

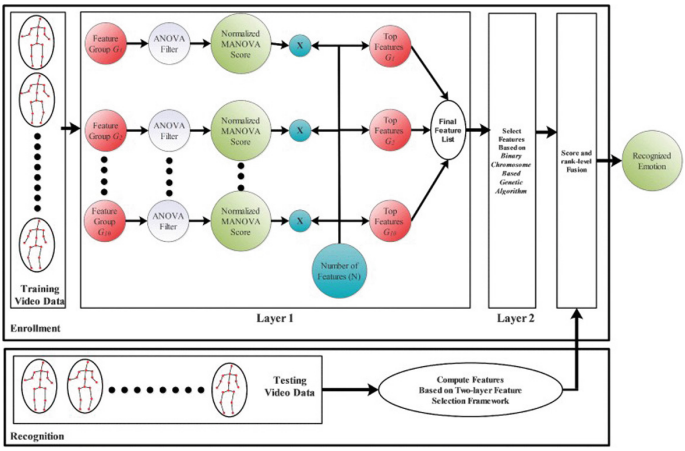

Fig. 3. Proposed method for emotion recognition from body movements.

3 Methodology

3.1 Convolutional Neural Network

A Deep Learning method called Convolutional Neural Network (CNN) takes the image as input. This method allocates the weights to objects of the image, which helps to distinguish one object from the other [9–11]. When compared to other approaches, pre-processing for this approach is less. After the training, the model extracts the features from the images by applying the filters and uses those features for its learning.

Convolution architecture has one input layer, one or more hidden layers, and one output layer. For each layer, there would be CONV, applying activation functions like RELU, and Softmax, applying to pools like max pooling, and mean pooling. There might or might not be parameters for each layer. Each layer takes the input from the previous layer's output. Each layer performs a differential function to give the output to the next layer [12].

3.2 Haar Cascade Classifier

Now, Facial recognition is present in every place namely security cameras, and sensors on iPhone X. There are many different human faces and there are many differences between humans to humans and many also similarities. How does facial recognition work to classify all faces?

Haar Cascade classifiers are the classifiers that detect the face of humans in real-time and can differentiate between humans and non-humans. Haar classifier is a machine learning algorithm that accepts input as an image or video and recognizes objects. Haar Cascade's model got trained on many positive images that consist of objects that the classifier wants to identify and also negative images that consist of objects that the classifiers don't want to recognize. Haar features are used to find how a given point is part of an object. To get a strong prediction, a group of weak learners is used with

boosting algorithms. These algorithms are executed on multiple sections of an input image or video using cascade classifiers [13, 14]. Open CV can be used to implement the Haar Cascade model [15].

3.3 Model

Convolutional Neural Network is used to build a model. It is one of the approaches in Deep learning which takes input as an image and extract features from it. When CNN processes the image, it acts as the human brain.

A convolutional neural network architecture consists of an input layer, one or more hidden layers, and an output layer. All the layers are stacked one by one linearly. The architecture is built by using sequential() in Keras and layers are added one by one by including filters, activation function, and pooling.

3.3.1 Input Layer

The images that come to the input layer must be of fixed size. So, pre-processing of images must be done before the images are fed into the input layer. The image's geometry must be fixed to the constant size. A computer vision package OpenCV is used to detect the face from the image. Haar cascade frontal face algorithm already has pre-trained filters of the face and employs AdaBoost to locate and get the face quickly from the image.

3.3.2 Convolutional Layers

In each convolution layer, a NumPy array is taken as input which includes parameters like kernel size, and several filters. In total, we have used 4 convolutional layers for expression classification and 3 convolutional layers for forage and gender classification. Convolution produces feature maps that show how pixel values are improved. **Pooling** is one of the convolutional neural network techniques that is used to reduce the dimension. This technique is used so that only the features that mainly affect the label for the classification. Using more and more convolutional layers increases processing time, this layer helps in reducing it. We have used the max-pooling layer of dimension (2,2) and reduced the image size by two.

3.3.3 Dense Layers

Dense layers take a huge number of characteristics from input and use them to get the trainable weights and layers. Forward propagation is used to train those weights, and then back-propagation to resolve errors. Parameters like learning rate, batch size, etc. can be used to improve the training pace. Dropout is a way that helps to avoid over-fitting the model that we build. It randomly picks some subset of nodes and assigns weights as zero to them.

3.3.4 Output Layer

In the output layer, instead of using the sigmoid activation function which is used for two classes as labels, we have used the softmax activation function which is used when

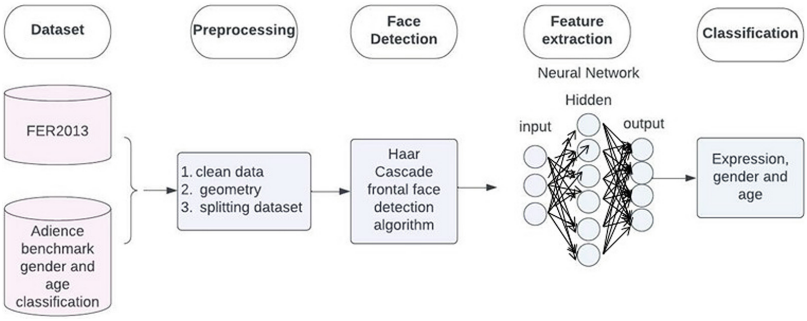

Fig. 4. Architecture

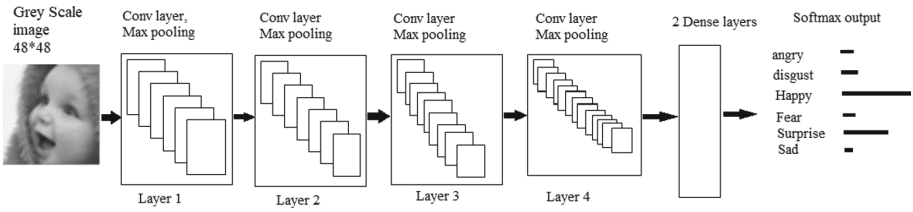

Fig. 5. Working model of the expression

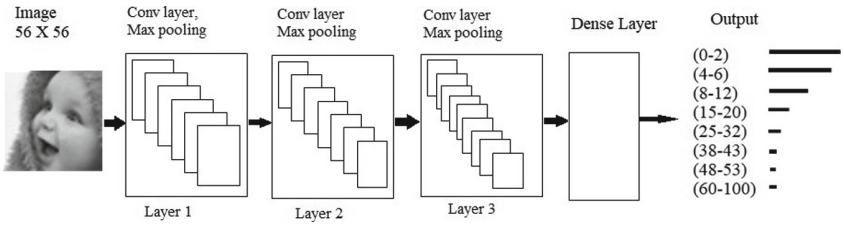

Fig. 6. Working model of the age

there are more than two classes as labels. After softmax activation is applied, the model gives the probability for each class Fig. 5. As a result, the model can show the probability of each class is the label for the given input. Finally, the label which gives the highest probability is taken and displayed to the user.

3.4 Applications

A web program that detects a person's expression, gender, and age can be utilized on a shopping website to gather product reviews from people of a specific age and gender. So that likely products of a certain age and gender can be separated (Fig. 4).

This software application may be used to store a person's age, gender, and expression in a single database, which can then be utilized for a variety of applications.

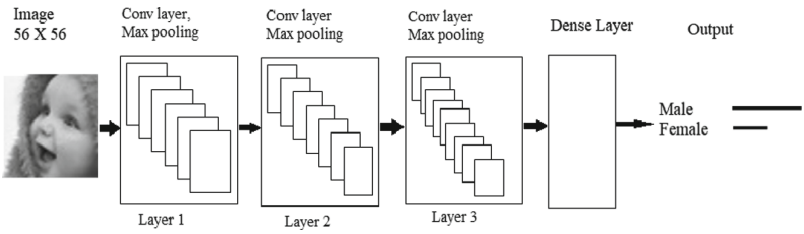

Fig. 7. Working model of the gender

Facial expression, gender, and age recognition are also useful in determining which course a person should take depending on his or her mental state, as well as his or her age and gender (Fig. 6).

This application can also be used to determine human behavior in a given situation. It aids in determining how a person reacts to a given situation based on age and gender (Fig. 7).

4 Results

Our model downsizes the training and testing images into 48 × 48 for expression and 56 × 56 for age and gender and then processes. 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 given graphs for accuracy and loss of the training dataset and validation dataset.

Below are the steps used to lively detect the image through a web camera and classify the detected face into one of the seven basic expressions [16], one of the two ages, and one of the eight groups of ages [17].

- Download the haar cascade frontal face default.xml and use the location where that is downloaded and store it in a variable (face classifier).
- Use the location where the .h5 file was created after running the model and stored in the variable (classifier).

3.Use the OpenCV method VideoCapture(0) by cv2.VideoCapture, 0 is to capture from the 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 probability.
6. Put the text on the rectangle drawn

We have used flask and designed a website whose root page shows the button to capture the face, clicking on it will navigate to another route that on the web camera and lively detects the human face and gives the label for expression, gender, and age Figs. 8 and 9.

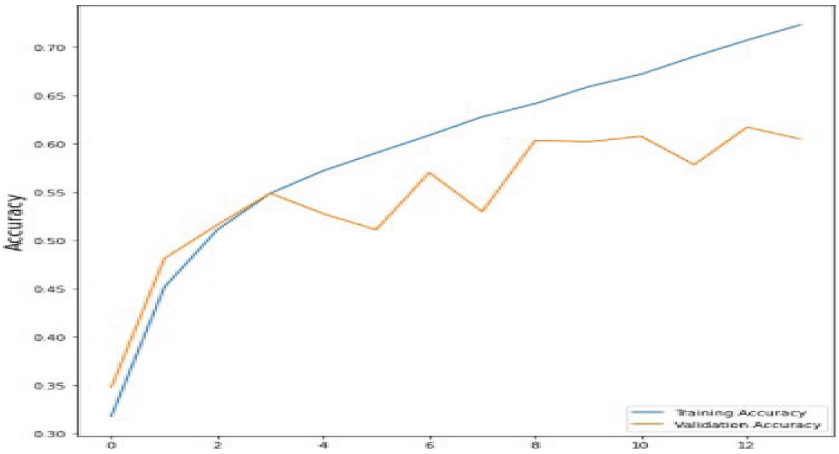

Fig. 8. Accuracy of the model

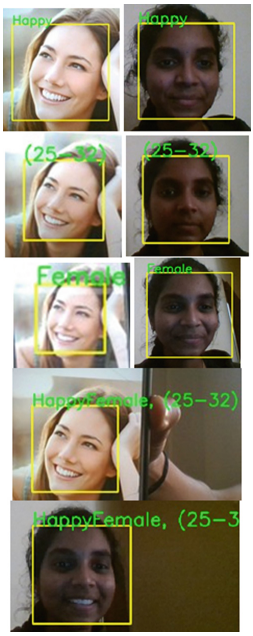

Fig. 9. Output of expression, age, and gender

5 Conclusion and Futurework

When a model incorrectly predicts an emotion, gender, or age, the right label is often the second closest emotion. The qualities of the human face are related to geometrical structures that are rebuilt as the recognition system's basic matching template, which are significant to varied expressions. 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 the arrangement of thick layers, the percentage of dropouts in dense layers, and the count of convolutional layers can be increased. To improve the model's accuracy, we'd like to add new databases to the system. This project may be expanded to accept the input file as an image and classify it into one of the seven expressions, one of the two genders, and one of the eight age groups.

References

1. Jung, H.: January. Development of deep learning-based facial expression recognition system. In: 2015 21st Korea-Japan Joint Workshop on Frontiers of Computer Vision (FCV), pp. 1–4. IEEE (2015)
2. Kwong, J.C.T., Garcia, F.C.C., Abu, P.A.R., Reyes, R.S.: Emotion recognition via facial expression: utilization of numerous feature descriptors in different machine learning algorithms. In: TENCON 2018–2018 IEEE Region 10 Conference, pp. 2045–2049. IEEE (2018)
3. Sambar, M.: FER-2013Dataset. IEEE Access 8 (2020)
4. Kumar, S., Singh, S., Kumar, J., Prasad, K.M.V.V.: Age and gender classification using SegNet based architecture and machine learning. Multimed. Tools Appl. **81**, 4228542308 (2022). https://doi.org/10.1007/s11042-021-11499-3
5. Qi, C., et al.: Facial expressions recognition based on cognition and mapped binary patterns. IEEE Access **6**, 18795–18803 (2018)
6. Ali, G., et al.: Artificial neural network based ensemble approach for multicultural facial expressions analysis. IEEE Access **8**, 134950–134963 (2020)
7. Zhao, H., Wang, P.: A short review of age and gender recognition based on speech. In: 2019 IEEE 5th International Conference on Big Data Security on Cloud (BigDataSecurity), IEEE International Conference on High Performance and Smart Computing, (HPSC) and IEEE International Conference on Intelligent Data and Security (IDS), pp. 183–185. IEEE (2019)
8. Ahmed, F., Bari, A.H., Gavrilova, M.L.: Emotion recognition from body movement. IEEE Access **8**, 11761–11781 (2019)
9. Gu, J., et al.: Recent advances in convolutional neural networks. Pattern Recogn. **77**, 354–377 (2018)
10. Albawi, S., Mohammed, T.A., Al-Zawi, S.: Understanding of a convolutional neural network. In: 2017 international conference on engineering and technology (ICET), pp. 1–6. IEEE (2017)
11. Tian, Y.: Artificial intelligence image recognition method based on convolutional neural network algorithm. IEEE Access **8**, 125731–125744 (2020)
12. Howse, J., Joshi, P., Beyeler, M.: OpenCV: Computer Vision Projects with Python. Packt Publishing Ltd. (2016)
13. Shilkrot, R., Escrivá, D.M.: Mastering OpenCV 4: a comprehensive guide to building computer vision and image processing applications with C++. Packt Publishing Ltd. (2018)
14. Hung, J., et al.: Keras R-CNN: library for cell detection in biological images using deep neural networks. BMC Bioinformat. **21**(1), 1–7 (2020)
15. Levi, G., Hassner, T.: Age and gender classification using convolutional neural networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, pp. 34–42 (2015)