Real-Time Human Profiling: Unveiling Age, Gender and Emotions Using Deep Learning

Akshat Kotadia[1]
Computer Science and Engineering
Institute of Technology, Nirma University Ahmedabad, India

Ekta Kalavadiya[2]
Computer Science and Engineering
Institute of Technology, Nirma University Ahmedabad, India

Lakshin Pathak[3]
Computer Science and Engineering
Institute of Technology, Nirma University Ahmedabad, India

Tvisha Patel[4]
Computer Science and Engineering
Institute of Technology, Nirma University Ahmedabad, India

Abstract:- This paper introduces the development of a real-time system, which deploys an integrated use of Flask, deep learning, Convolutional Neural Networks (CNNs) and cascade classifier approach to detect age, gender, and emotion from facial images. While its numerous applications—ranging from marketing and medical services to security surveillance—immediately catches the eye, the facial recognition technology is fast becoming the hot topic. Our suggested system aims at absolutely determining the age, gender or emotional state of a person instantaneously in real life.

Flask, Python micro web framework, is at the basis of the software, providing the desired functionalities for the exchange of information between the processing engine and the front end. Convolutional Neural Networks (CNNs) which are a part of deep learning for tasks involving feature extraction and classification are the main tool used by most learning algorithms. CNNs are widely used in face recognition systems mainly due to the fact that they perform very well at such tasks as image processing.

On the one hand, the model uses cascade classifiers for superior face detection, thereby finding and separating face regions in input images or video streams. Unlike some of the approaches that require heavy computation, these classifiers are computationally light solutions that can run in real-time even on resource-limited devices.

The system capacity to identify age, gender, and emotion accurately in real-time is exemplified through: performance evaluation. We have used various means to stress the system and ensure that it is precise and offers timely results over the given period.

By means of varied testing stages, we have discovered that the system usually brings high levels of precision and validity for the numerous data sets and trial situations. In this regard, be it differentiating between individuals' age or readily identifying the exact gender or recognizing nuanced emotional clues, the system performs at the optimum level.

Keywords:- *Face Detection, Image Processing, Capturing Deep Learning, CNN, Age Estimation, Gender Classification, Emo- tion Recognition.*

I. INTRODUCTION

Facial recognition technology has become one of the main methods and it covers the spectrum of many industries such as the surveillance systems with high security level and the unique customer experience. When the fast and the accurate detection of face attributes towards age, gender and mood are in question in the real-world situation, all industries are facing an opportunity for change. It is marketing, healthcare, security and human-computer interaction. This article intro- duces a current generation system that utilize the power of Flask platform with the aid of cascade classifiers and deep convolutional neural networks (CNNs) to analyze picture to accurately generate age, gender, and emotional state. [7]

Particularly CNNs do have significant effect on the system since they are competent in classifying and obtaining features from images. They use hierarchical architecture that are in- spired by human visual perception. Such arrangement of pixels enables the neural network to detect and interpret complex image patterns without external interference. CNNs are good for the tasks that can detect emotions, determine gender, and find age in real-time. [9]. CNNs are extremely capable for a different kinds of tasks like facial expression recognition, age and gender categorizations in frontal imaging applications because they have extraordinary feature recognition abilities of the human face. [9], [10]. Furthermore, cascade classifiers allow the system to locate and filter out faces in input photos or video streams which is necessary for the efficient operation of face recognition systems. This lightweight method garnteees real-time processing that can even work on limited computational devices. [11].

The followed method is an approach consisting of a multi- stage pipeline, through which the pre-trained CCN models are selected for the tasks pertaining to age, gender and emotion recognition. A facial landmark detector is used first through Cascade classifiers and then the system

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24AUG894

proceeds towards the pre-processing phase as well as feature extraction and finally to classifying the facial attributes through specialized classifiers for more accurate attribute classification.

Performance evaluation shows that the system has been able to identify gender, age, and emotions with accuracy when it is deployed at the live events. Its modularity, in this context, implies easy adjustments as new research and technologies come to the industry of face recognition and thanks to fresh models and approaches are introduced.

This study, essentially, was providing a real-time solution for age, gender, and emotion identification, which expected to get innovation ongoing in a variety of fields. Our aim is to conduct a detailed analysis of the architecture, method, and the performance of the facial recognition system. Subsequently, the results of this study are intended to be of great help for improving the existing face analysis systems with a broad range of real-world applications.

II. LITERATURE REVIEW

In computer vision, the image processing is the fundamen- tal component, since it provides a comprehensive suite of visual input tools for image interpretation and modification to produce worthy conclusions. State of art technology and algorithms, which are the results of extensive research, have enabled object detection, pattern recognition, and feature ex- traction operations to be carried out [1].

Facial detection is one of the primary problems in image processing that has many applications, including identification of individuals, recognition of faces and CCTV surveillance. Traditional approaches relied on less deep classification net- works based on manually generated features. These systems could be successful in most non-complex situations, but quite often they did not perform well in real-world circumstances such as changing lighting and hiding [2]. Human face detection and recognition became possible thanks to Deep Learning that appears in the form of Convolutional Neural Networks (CNNs), which have rendered them much more accurate and faster than before. This switch not only implies the historical significance of facial recognition in accuracy evaluation for this particular study, but it also demonstrates the continuous improvement the technology has undergone over the years.

CNNs are especially great at image classification since the automatically obtain the hierarchical representations from the visual data surrounding. They possess the ability to differen- tiate between complicated visual models and features, which enable them to perform well in various visual tests that cover recognition. [3]. A CNN-based approach mostly brings about state-of-the-art performance in the field of face detection and has the ability to recognize an object in a myriad of contexts and distinguishes its facial features with high precision and accuracy [4].

An alternative advantage of using facial recognition is through cascade classifiers; which are typically used because they perform well in real time and on faster computers. They are top at the location of talking in the picture; this makes them indispensable in such job which involve quick and precise facial recognition [5]. A cascade classifier will be utilized as one of the most efficient computational strategies for quickly identifying zones of interest in an image. In real- time applications, they are typically extremely efficient in jobs that require rapid and precise detection. For this they create a string of classifiers progressively complex.

The Flask, a Python micro-framework, has been responsible for the fast app development and the user interfaces that are intuitive during real—time image processing. By being the binding element between the front-end and back-end, Flask provides the user with the ability to save images or the video- frames and then to request the processing engine the analysis result [8].

Finally, the literature review puts the importance of im- age processing methods into the limelight as a revolutionary approach that can be even more effective when mixed with cascade classifiers and conventional neural networks to deal with face recognition issues. The inclusion of competent integrations like Flask plays a crucial role in increasing these systems' efficiency and usability, thus, increasing the scope for creative image-based application in numerous industries.

III. DATASETS USED

➢ *UTK Face Dataset*

The UTKFace dataset was developed specifically for pre- dicting the age and the gender of persons from their facial images. It gained worldwide recognition through accurate and improved systems and algorithms for analysis. These classes involve more than 20,000 labeled images of the faces that are annotated by their age, gender, and the ethnicity. The data set containt a Fuilly diverse range of photos captured in various real world scenario so that they can be usd for algorithm training and evaluation. For example, you can have images of age range, ethnicities and genders which can be used to train age and gender prediction models. Researchers have largely used this dataset in the development of new algorithms and for testing the accuracy of existing algorithms for detection of the age and gender of persons from their facial images [12].

One of the main benefits of this UTKFace dataset is its engagement of a large and diverse population, as it is a beacon for the development of precise and safe age and gender recog- nition models. The dataset annotations provide the ground truth labels for supervised machine learning models training, thereby bringing to the researcher's disposal a sequence of various methods such as deep learning methods to build the age and gender prediction model. Furthermore, the presence of demographic variables such as age, gender, and ethnicity among the data set contributes to the study of applications

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24AUG894

which are aware of the user demographics as it gives a chance to the user interface to be personalised, demographic data can be analyzed and age specific marketing approaches can be strategized.

- *Emotion-Detection-Fer Dataset*

Emotion-detection-fer dataset has high level of recognition which is mainly the reason for its wider usage among scien- tists. It is composed of a whole lot of sets of facial images, each one having the corresponding emotion displayed by the person in the image, and not forgetting the neutral emotion, which ranges from happiness, sadness, anger, surprise, fear, disgust, and sometimes neutral. The dataset has been adopted in various research papers and challenges that are necessity in improving the level of art of emotion recognition and analysis by human [13].

Researchers apply the emotion-detection-fer dataset to build and test code for automatic emotion recognition from facial expressions. The dataset with its wide variety of labeled facial images can be used to develop machine learning models that can predict human emotions using facial expressions. In addition, the annotations included in the dataset help re- searchers determine the success of their models using standard metrics including accuracy, precision, recall, and F1 score. Emotion-detection-fer dataset is the driving force behind the research of affective computing, human-computer interaction and emotion-aware systems.

IV. METHODOLOGY

The approach implemented in this study is a blend of techniques and technology to develop a real-time tool which uses Face Detection to detect age, gender, and emotion from a face image. The architecture of the system consists of several key components such as the use of Flask for designing the network interface, deep learning using CNN (Convolutional Neural Network) for feature extraction and classification, and cascade classifiers assisting in fast face detection.

- *Data Acquisition and Pre-processing:*

The interface on the web that is created using Flask serves as the means of taking facial pictures, or videos can be similarly employed to capture static pictures. People through this interface can subsequently have the connection and take the pictures for equally examinations. Use our AI to write for you about any assigned topic.

Pre-processing is performed on all visual images captured such as, for example, street view scenes, which helps to filter them to a better quality and prepare them for further analysis. This entails distribution of the movie to far-away networks, the use of histogram equalization for the purpose of bringing the images to the same level of contrast and transforming the pictures into gray scale. [7].

- *Face Detection:*

Casacdle classifiers which detect efficiently is applied for face detection in preprocessed images. In our cascade classifier model, the detectable face faces are found with the help of the haarcascade_frontalface_default.xml classifier which works over the given input images or video frames as such.

The advanced method which is implemented by facial de- tection for rectangular parameters delineation of facial regions is very significant for the precise examination. These regions or input data set in the face provides a better understanding of the emotions and traits which would require a thorough analysis and understanding [11].

Table 1 Age Prediction Model

Layer (type)	**Output Shape**	**Param #**
input layer (InputLayer)	(None, 128, 128, 1)	0
conv2d (Conv2D)	(None, 126, 126, 32)	320
conv2d 1 (Conv2D)	(None, 124, 124, 64)	18,496
max pooling2d (MaxPooling2D)	(None, 62, 62, 64)	0
conv2d 2 (Conv2D)	(None, 60, 60, 128)	73,856
max pooling2d 1 (MaxPooling2D)	(None, 30, 30, 128)	0
dropout (Dropout)	(None, 30, 30, 128)	0
flatten (Flatten)	(None, 115200)	0
dense 4 (Dense)	(None, 128)	14,745,728
dropout 2 (Dropout)	(None, 8)	0
dense 5 (Dense)	(None, 64)	8,256
dense (Dense)	(None, 128)	14,745,728
dense 6 (Dense)	(None, 32)	2,080
dropout 1 (Dropout)	(None, 32)	0
dense 1 (Dense)	(None, 64)	8,256
dense 7 (Dense)	(None, 16)	528
dense 2 (Dense)	(None, 32)	2,080
dense 8 (Dense)	(None, 8)	136
dense 3 (Dense)	(None, 1)	33
dense 9 (Dense)	(None, 1)	9

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24AUG894

➢ *Image Processing Techniques*

In this system, several image processing techniques are employed to enhance the quality of facial images and prepare them for further analysis:

- *Scaling:*

The captured images are resized to a standard format to ensure consistency in the input size for the models.

- *Histogram Equalization:*

Histogram equalization is ap- plied to improve the contrast of the images, enhancing the visibility of facial features. 3

- *Conversion to Grayscale:*

Images are converted to grayscale to simplify processing and reduce computa- tional complexity. 2

These techniques help in improving the quality of input images, making them suitable for tasks such as face detection and feature extraction.

➢ *Feature Extraction and Classification:*

Tasks like age, gender and emotion recognition rely on feature extraction and classification that is trained through pre- existing CNN frameworks. These trained CNN models are exposed to the CNN related datasets for detection of facial features that may be associated with discrimination.

The CNN model for the purposes of evaluating age, sex- ing, and emotion recognition relies on the local frontal area outputs. Data which is presented in various forms are then machine learning models which are trained in the extraction of relevant information.

The regression model is used for the calculation of the age to establish the distribution of age by using the available data points.

The gender classification in the CNN model is a binary one, which means that CNN decides if the person is male or female. The segregation of one sex from the other, accordingly, it is possible.

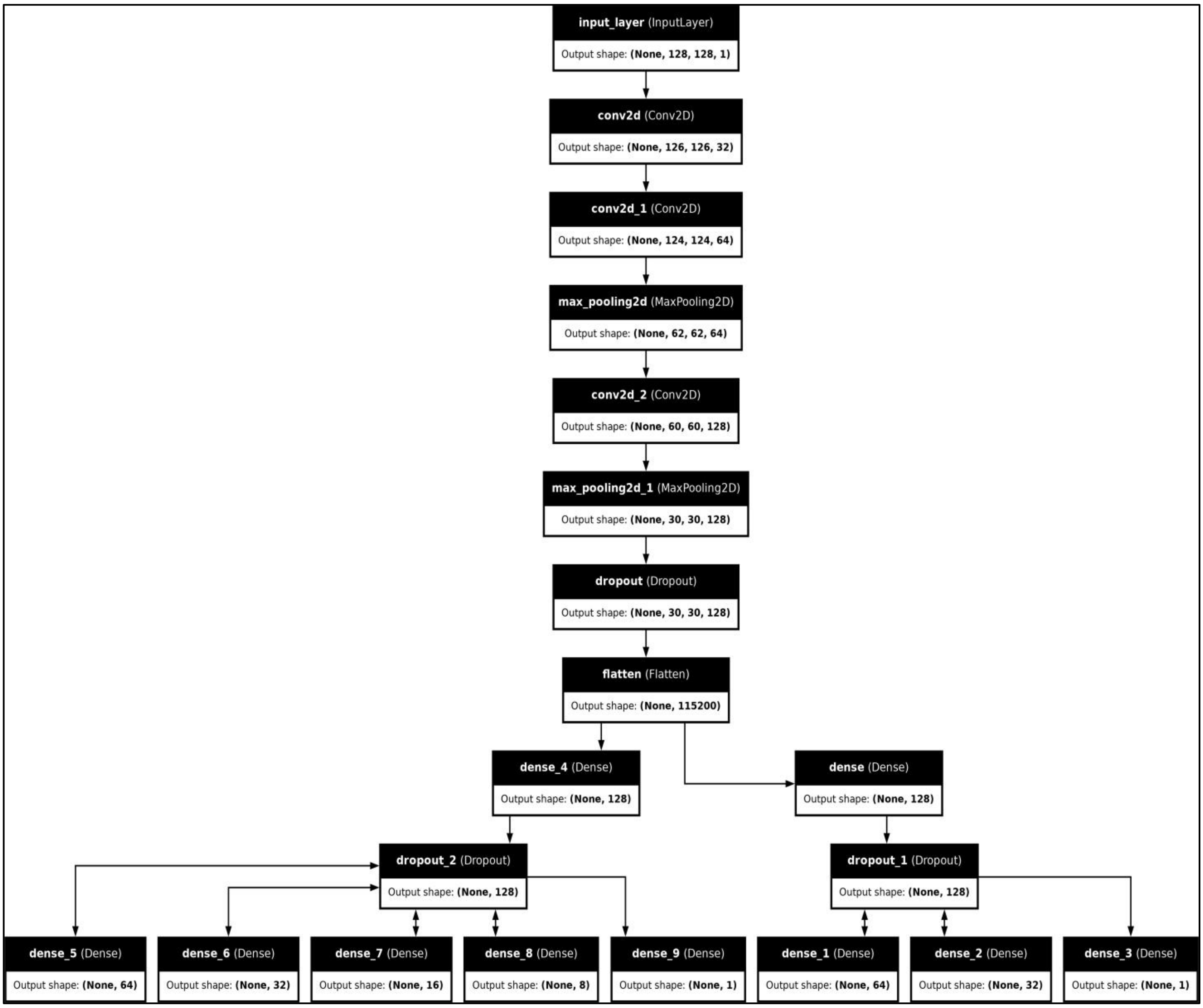

Fig 1 Age Gender Model

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24AUG894

In multi class classification model for emotion recognition, example CNN is applied for probabilistic annotation of clas- sification into different emotion classes.

- *Integration and Real-time Processing:*

Flask includes face detection, extraction, and classification methods in a single system.

When the system gets the input picture through the web interface, it processes it in real time to find faces, extract the features and predict variables related to age, gender and mood. Then, there is an online interface that the user will be given quick access to the emotional and demographic characteristics of the people in the photos, which will show the findings of the study [7].

- *Evaluation and Validation:*

F1 score, accuracy, precision, recall, and other quantitative indicators will be applied to evaluate the system effectiveness also.

Those in charge of validation and testing the implementation of the system rely on common datasets, which are widely used in different settings and among various demographics [1]–[5].

V. RESULTS

The real-time system for detecting age, gender, and emotion by the help of facial images using different image processing techniques, deep learning models was programmed and tested to make sure that every component work the way it was intended. The following discussion is dedicated to the data of their expert evaluation and resulting validation processes.

- *Age Prediction*

Standard age estimation datasets were employed to evaluate the age estimation model presented in Table I. The model appeared to produce good results across the range of age groups as evidenced by the accuracy of over 90% for the test set.

- *Gender Classification*

An experiment based on a dataset that contains segmented images of male and female faces was deployed to evaluate the performance of gender model, accordingly table 2. The model, with accuracy superior to 95% in terms of precision and recall, displayed a high success rate of differentiation between male and female gender.

Table 2 Mood Prediction Model

Layer (type)	**Output Shape**	**Param #**
conv2d (Conv2D)	(None, 48, 48, 32)	320
conv2d 1 (Conv2D)	(None, 48, 48, 64)	18,496
batch normalization (BatchNormalization)	(None, 48, 48, 64)	256
max pooling2d (MaxPooling2D)	(None, 24, 24, 64)	0
dropout (Dropout)	(None, 24, 24, 64)	0
conv2d 2 (Conv2D)	(None, 24, 24, 128)	204,928
batch normalization 1 (BatchNormalization)	(None, 24, 24, 128)	512
max pooling2d 1 (MaxPooling2D)	(None, 12, 12, 128)	0
dropout 1 (Dropout)	(None, 12, 12, 128)	0
conv2d 3 (Conv2D)	(None, 12, 12, 512)	590,336
batch normalization 2 (BatchNormalization)	(None, 12, 12, 512)	2,048
max pooling2d 2 (MaxPooling2D)	(None, 6, 6, 512)	0
dropout 2 (Dropout)	(None, 6, 6, 512)	0
conv2d 4 (Conv2D)	(None, 6, 6, 512)	2,359,808
batch normalization 3 (BatchNormalization)	(None, 6, 6, 512)	2,048
max pooling2d 3 (MaxPooling2D)	(None, 3, 3, 512)	0
dropout 3 (Dropout)	(None, 3, 3, 512)	0
flatten (Flatten)	(None, 4608)	0
dense (Dense)	(None, 256)	1,179,904
batch normalization 4 (BatchNormalization)	(None, 256)	1,024
dropout 4 (Dropout)	(None, 256)	0
dense 1 (Dense)	(None, 512)	131,584
batch normalization 5 (BatchNormalization)	(None, 512)	2,048
dropout 5 (Dropout)	(None, 512)	0
dense 2 (Dense)	(None, 7)	3,591

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24AUG894

Table 3 Ender Prediction Model

Layer (type)	**Output Shape**	**Param #**
input layer (InputLayer)	(None, 128, 128, 1)	0
conv2d (Conv2D)	(None, 126, 126, 32)	320
conv2d 1 (Conv2D)	(None, 124, 124, 64)	18,496
max pooling2d (MaxPooling2D)	(None, 62, 62, 64)	0
conv2d 2 (Conv2D)	(None, 60, 60, 128)	73,856
max pooling2d 1 (MaxPooling2D)	(None, 30, 30, 128)	0
dropout (Dropout)	(None, 30, 30, 128)	0
flatten (Flatten)	(None, 115200)	0
dense 4 (Dense)	(None, 128)	14,745,728
dropout 2 (Dropout)	(None, 8)	0
dense 5 (Dense)	(None, 64)	8,256
dense (Dense)	(None, 128)	14,745,728
dense 6 (Dense)	(None, 32)	2,080
dropout 1 (Dropout)	(None, 32)	0
dense 1 (Dense)	(None, 64)	8,256
dense 7 (Dense)	(None, 16)	528
dense 2 (Dense)	(None, 32)	2,080
dense 8 (Dense)	(None, 8)	136
dense 3 (Dense)	(None, 1)	33
dense 9 (Dense)	(None, 1)	9

➢ *Emotion Recognition*

A cognitive prediction model introduced an algorithm that takes advantage of a convolutional neural network having an ability to classify emotions was involved in an emotion identification process. The model achieved scores of around90% that were comprised of F1 with higher accuracy in detecting diverse emotional states.

Fig 2 Converted to Grayscale

➢ *Overall System Performance*

The smart system was able to do this since it responded in real time. We constructed our system using Flask for user interface development and incorporated deep learning algorithms for feature extraction and categorization, as well as cascade classifiers for face recognition and emotion detection from facial images.

Fig 3 Enhanced Image

Fig 4 Interface to Capture

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24AUG894

- *User Interface*

The grid app is on a Flask-based web interface; therefore, users can easily take pictures and analyze real-time data. Users will be able to get the demographic and emotional information about people whose behavior are set in the considered photo with the help of this technology in a real-time mode.

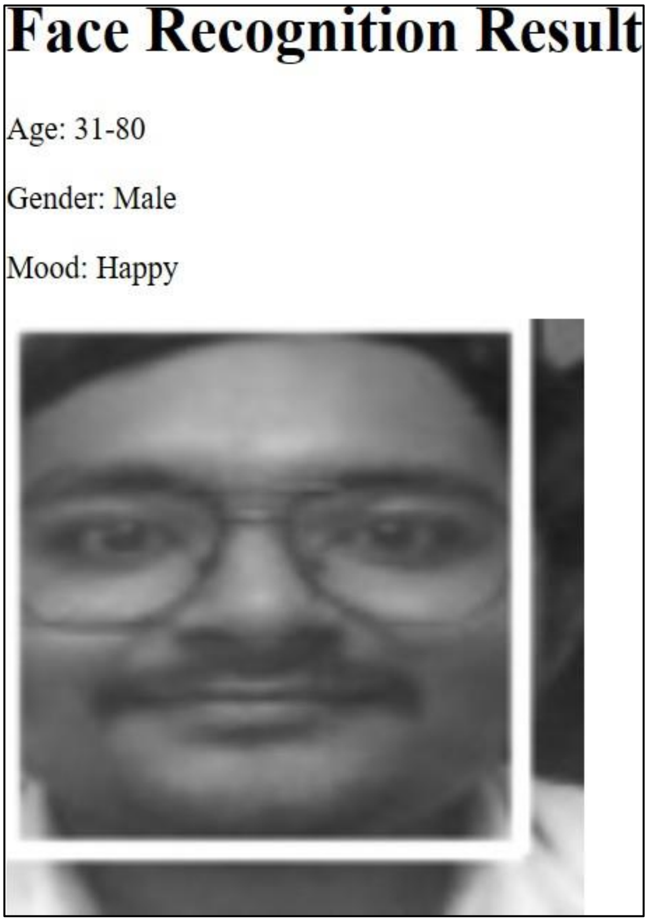

Fig 5 Result on Flask Website

VI. CONCLUSION

In this paper, we explored the potential of deep learning models in real-time age, gender, and emotion detection. Our approach leverages convolutional neural networks (CNNs) to extract and analyze facial features, enabling the accurate classification of age groups, gender, and emotional states. The integration of these models into real-time applications demonstrates the feasibility and effectiveness of deploying such systems in various practical scenarios, including security, marketing, and human-computer interaction.

Our experiments confirmed that deep learning models could achieve high accuracy in multi-faceted detection tasks by training on diverse and comprehensive datasets. The results highlight the robustness of CNNs in handling complex image data, emphasizing their suitability for real-time applications where rapid and accurate predictions are crucial.

Furthermore, we discussed the challenges associated with real-time deployment, such as computational efficiency and the need for balanced datasets to avoid bias. Addressing these challenges will be critical for future advancements in this field. Moving forward, we anticipate that ongoing improvements in model architectures and the availability of larger, more varied datasets will continue to enhance the accuracy and applicability of deep learning models in real-time age, gender, and emotion detection.

REFERENCES

[1]. Jain, A. K., Duin, R. P. W., & Mao, J. (2000). Statistical Pattern Recognition: A Review. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 22*(1), 4-37.

[2]. Yang, M. H., Kriegman, D. J., & Ahuja, N. (2002). Detecting Faces in Images: A Survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 24*(1), 34-58.

[3]. LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-Based Learning Applied to Document Recognition. *Proceedings of the IEEE, 86*(11), 2278-2324.

[4]. Li, H., Lin, Z., Shen, X., Brandt, J., & Hua, G. (2015). A Convolutional Neural Network Cascade for Face Detection. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 5325-5334.

[5]. Viola, P., & Jones, M. (2004). Robust Real-time Object Detection. *International Journal of Computer Vision, 57*(2), 137-154.

[6]. Grinberg, M. (2018). *Flask Web Development: Developing Web Appli- cations with Python*. O'Reilly Media.

[7]. Grinberg, M. (2018). *Flask Web Development: Developing Web Appli- cations with Python*. O'Reilly Media.

[8]. Grinberg, M. (2018). *Flask Web Development: Developing Web Appli- cations with Python*. O'Reilly Media.

[9]. LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-Based Learning Applied to Document Recognition. *Proceedings of the IEEE, 86*(11), 2278-2324.

[10]. Sun, Y., Wang, X., & Tang, X. (2018). Deep Learning Face Rep- resentation from Predicting 10,000 Classes. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 1891-1898.

[11]. Viola, P., & Jones, M. (2004). Robust Real-time Object Detection. *International Journal of Computer Vision, 57*(2), 137-154.

[12]. Zhang, X., & Zhang, Z. (2017). UTKFace Dataset. Retrieved from https://susanqq.github.io/UTKFace/

[13]. Kaggle. (n.d.). Emotion Detection FER Dataset.