Deep-Learning Based Age, Gender & Emotion Detection for Retail Customer Profiling

Project Report submitted to the SRM University - AP, Andhra Pradesh
for the partial fulfillment of the requirements to award the degree of

Bachelor of Technology in
Computer Science & Engineering
School of Engineering & Sciences

submitted by

B Praneeth (AP22110010049)

P V Sri Nag (AP22110010050)

V Mohan Balu (AP22110010057)

M Sri Hari (AP22110010368)

Under the Guidance of

Dr. Ch. Mallikarjuna, Asst. Professor, CSE Department

Department of Computer Science & Engineering
SRM University-AP Neerukonda,
Mangalgiri, Guntur
Amaravati, Andhra Pradesh - 522 240

Dec 2025

DECLARATION

I undersigned hereby declare that the project report **Deep-Learning Based Age, Gender & Emotion Detection for Retail Customer Profiling** submitted for partial fulfillment of the requirements for the award of degree of Bachelor of Technology in the Computer Science & Engineering, SRM University-AP, is a bonafide work done by me under supervision of Dr. C h . M a l l i k a r j u n a . This submission represents my ideas in my own words and where ideas or words of others have been included, I have adequately and accurately cited and referenced the original sources. I also declare that I have adhered to ethics of academic honesty and integrity and have not misrepresented or fabricated any data or idea or fact or source in my submission. I understand that any violation of the above will be a cause for disciplinary action by the institute and/or the University and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been obtained. This report has not been previously formed the basis for the award of any degree of any other University.

Place	: ...Mangalagiri......................		Date: December 2, 2025
Name of student	: B Praneeth	Signature	: ....
Name of student	: P V Sri Nag	Signature	: ....
Name of student	: V Mohan Balu	Signature	: ....
Name of student	: M Sri Hari	Signature	: ....

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

SRM University-AP Neerukonda,

Mangalgiri, Guntur

Amaravati, Andhra Pradesh - 522 240

CERTIFICATE

This is to certify that the report entitled **Deep-Learning Based Age, Gender & Emotion Detection for Retail Customer Profiling** submitted by **B Praneeth, P V Sri Nag, V Mohan Balu, M Sri Hari** to the SRM University-AP in partial fulfillment of the requirements for the award of the Degree of Bachelor of Technology in the Department of Computer Science & Engineering is a bonafide record of the project work carried out under my/our guidance and supervision. This report in any form has not been submitted to any other University or Institute for any purpose.

Project Guide

Name : Dr. Ch. Mallikarjuna

Signature: ....

Name : Dr. Murali Krishna Enduri

ACKNOWLEDGMENT

I wish to record my indebtedness and thankfulness to all who helped me prepare this Project Report titled **Deep-Learning Based Age, Gender & Emotion Detection for Retail Customer Profiling** and present it satisfactorily. I am especially thankful for my guide and supervisor Dr. Ch. Mallikarjuna in the Department of Computer Science & Engineering for giving me valuable suggestions and critical inputs in the preparation of this report. I am also thankful to Dr. Murali Krishna Enduri, Head of Department of Computer Science & Engineering for encouragement. My friends in my class have always been helpful and I am grateful to them for patiently listening to my presentations on my work related to the Project.

B. Praneeth, P V Sri Nag, V Mohan Balu, M Sri Hari
AP22110010049, AP22110010050, AP22110010057, AP22110010368

B. Tech.
Department of Computer Science & Engineering
SRM University-AP

ABSTRACT

This report show a complete deep learning based facial analysis system that can be able to predict any individual age group, gender, and emotional state from the given static facial images with high accuracy by using specialized Convolutional Neural Network architecture. This solution is designed as a modular multi model pipeline where three CNN models are independently trained and optimized for the age classification across seven different demographic ranges, binary gender identification, and emotion recognition mapped into positive, negative, and neutral categories using the CK+48 data. Our model take the advantage of augmented datasets, grayscale transformation, image normalization, and standardized resizing techniques to increase the efficiency of feature extraction and reduce the computational complexity of the models. We have the used dropout and L2 penalties to keep stable and it will prevent the model from overfitting. By regularly saving the model's progress (checkpointing), we ensured it generalizes well, as proven by the consistent accuracy results.This pipeline will integrate the face detection using Haar Cascades to isolate the facial regions, which are then processed at the same time through the three pre trained networks at an inference time. This will allow us to moniter the real time, multi attribute predictions which are useful for installing in the surveillance systems, intelligent retail monitoring, customer experience analytics, healthcare diagnostics, and also human computer interaction systems. The results we got will establish the robustness of the CNNs in learning the hierarchical facial patterns and show a unified scalable framework for the demographic and affective computing while also highlighting the importance of quality datasets, preprocessing strategies, and modular architecture design for reliable, efficient computer vision performance.

CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1
INTRODUCTION TO THE PROJECT

Facial analysis has now become a necessary feature of computer vision in modern applications, which allows computers to automatically recognize essential human characteristics including age, sex, and mood. As the demand for automated, quick and reliable systems grows in the field of surveillance, retail analytics, healthcare and human-computer interaction, deep learning provides an excellent tool to study faces with high precision. The proposed project is aimed at the creation of a single system based on the use of Convolutional Neural Networks (CNN) to determine the age, gender, and emotion of a person after viewing his or her face. The methodology is training three expert CNN models on filtered and processed datasets, such as age prediction using augmented facial images, gender classification using the UTKFace dataset and emotion recognition using the CK+48 dataset. The preprocessing steps like the conversion to grayscale, normalization, and resizing of the images ensure that the model is stable and effective in training. It includes a Haar Cascade face detector to extract facial areas then forwarding them to the corresponding models to perform multi-attributes in real time. The created system shows the usefulness of deep learning in deriving the intricate facial patterns and outlines the opportunities of a cohesive model pipeline of strong demographic and emotion analysis.

1. WHAT IS AGE–GENDER–EMOTION DETECTION?

Age-Gender-Emotion Detection is a computer vision method which employs deep learning models to examine facial images and estimate the age bracket, gender, and emotional status of a particular person automatically. Through the application of Convolutional Neural Networks (CNNs), these systems are able to learn meaningful visual representations, including facial geometry, texture, facial expressions and demographic features. Such systems have the objective of converting raw visual information into useful information that may be applied to the real world.

The benefits of the Age-Gender-Emotion detection can be applied to automated surveillance, behavior perception, user-level interaction and analytics through constructing rapid and precise forecasts. These models decrease the necessity of manual evaluation process, increase the scale, and the efficiency of human-machine interaction process in general.

2. ROLE OF AGE–GENDER–EMOTION DETECTION IN REAL WORLD APPLICATIONS

The Age-Gender-Emotion Detection is of great importance in many industries where machines can make intelligent visions of human characteristics. In retail stores, the systems are useful in the analysis of customer demographics and the emotional reactions with an aim of enhancing the marketing strategy and store experiences. Age and gender prediction in security and surveillance help in the identification of people of interest, whereas emotion detection helps in the initial detection of suspicious or stressed behavior. These models are used in human-computer interaction systems in order to customize content, modify system responses, and enhance user interaction.

Also, this technology can be useful in healthcare and therapeutic environments to track the mood of a patient and his behavioral patterns. These deep learning-based systems reduce human errors by automating tasks of complex facial analysis, improve operational efficiency and offer quick assessments based on visual data processing which is robust.

Chapter 2
MOTIVATION

Over the last few years, Artificial Intelligence (AI) in computer vision has become a common phenomenon that has revolutionized the way computers perceive and analyze human behavior. The necessity to have precise and timely appreciation of human characteristics including age, gender, and emotions is one of the critical issues in the contemporary digital systems. Conventional paper-based evaluation systems are tedious, subjective, and not very practical in case of large-scale or instant decision-making. The automated systems are also effective in interpreting the facial features and behavioral cues that are common in industries like surveillance, retail analytics, healthcare and human-computer interaction. The possibility to retrieve such information in real time could make a huge contribution to better decision-making, personalizing, and user experience.

These requirements prompted the development of Convolutional Neural Networks (CNNs) that are able to learn meaningful patterns on faces after processing images of faces in a deep hierarchical way. The age, gender and emotion detection systems based on CNN are able to analyze visual signals, identify expressions, demographic groups and offer correct insights on a real time basis. These models are devoid of manual errors, shorter processing time and performance consistency of such models makes it very suitable in automation in a host of applications. This project is motivated by the growing need in the intelligent systems that would be able to analyze the emotions and demographic features of people without the participation of a human being.

Due to the fact that CNNs are capable of automatically identifying visual characteristics and offer immediate predictions, the proposed system can be used as a base of smart monitoring, understanding user behavior, and custom-made digital spaces. These models can serve as well to ensure security and retail settings by providing analytical knowledge that improves the quality of services, safety, and operational efficiency.

These were the objectives that gave the project its primary incentive. The main aim was to develop a scalable and stable system that could be used to correctly determine the age range, gender and the emotional type of a person based on a single face image. With this system, a multi-model CNN pipeline, we will expect to accomplish the following goals:

- Automate the identification of age categories of humans based on the image classification.
- Determining the gender fast and with high precision to the real time uses.
- Identify emotional conditions (positive, negative or neutral) in order to examine behavioral patterns.
- Lessen the reliance on manual observation, demographic and emotional analysis.
- Enhance efficiency in areas that need quick human evaluation.

1. REAL - WORLD IMPORTANCE

The correct recognition of age, gender and emotional expressions will be critical towards boosting user safety and personalization and decision-making in most real-life aspects. The emotional cues can be used in detecting suspicious or distressed behaviors early in surveillance and public safety to avoid detrimental circumstances. Demographic and emotional analytics can help businesses in retail settings to enhance the customer experience, the store layout, and the marketing strategies. Likewise, within healthcare and applications related to therapy, emotion recognition assists in tracking the mood variations and behavioral changes of the patient and enhances the quality of the mental health assessment.

Nevertheless, most organizations continue to use manual procedures that are time intensive, inconsistent and resource consuming. This usually causes time wastage, misunderstanding, and less efficiency. This gap can be reduced with the introduction of automated facial analytics with CNNs, which will offer systems capable of processing visual information fast, extracting useful information, and offering real-time results.

Thanks to this technology, not only the industries will feel the advantages of effective monitoring and analysis of the processes, but the quality of the human-machine interaction will also increase. The system is one of the contributors to the broad scope of computer vision and AI because:

- Reducing manual labor in the analysis of big volumes of visual data.
- Increasing the systems of automated monitoring and personalization of the users.
- Promoting evidence-based decision-making on the foundation of demographic and behavioral data.

In this way, the project is one of the steps towards a high-level AI-based facial analysis, in which technology assists a human specialist to make timely, more precise, and context-driven sense of human traits and feelings.

Chapter 3
LITERATURE SURVEY

Over the past ten years, facial analysis with Artificial Intelligence has changed significantly due to the development of deep learning structures, the presence of more computing capabilities, and the existence of large-scale annotated data. The previous systems were strongly relying on traditional machine learning methods that were based on handcrafted attributes like Local Binary Patterns (LBP), Haar-like features and Histogram of Oriented Gradients (HOG). Although these methods offered a baseline advancement, they could not fully represent the diverse details involved in the real world face image- small facial expression, texture alteration with age, light variations, and gender-related features. These drawbacks limited scalability, adaptability and accuracy of the early facial analysis systems.

The usage of Convolutional Neural Networks (CNNs) became a milestone, and it is possible to extract features automatically and learn the hierarchical representations without going through appearance or other methods when dealing with raw pixel data. The proposed CNN-based models were able to deal with deeper architecture and non-linear feature hierarchy along with strong spatial pattern recognition which greatly enhanced performance in facial classification tasks. Consequently, CNNs soon became the leading methodology to do things like age estimation, gender recognition, and emotion recognition by doing it better than conventional approaches in accuracy and strength. Creation of curated datasets such as UTKFace, CK+, FER2013 and large scale Internet-based collections also contributed to the further progress of the research as they offer a variety of samples that can be used to train and evaluate a model.

Although these have been developed, there are still problems, such as the change of the domain between datasets, imbalance of classes, image quality variation and the ability to identify subtle or ambiguous expressions. These elements explain why more credible, combined, and versatile facial analysis systems are necessary. A literature survey has been carried out in detail to get familiar with the strengths, limitation, and contributions that previous research has made. The given review gives critical reviews of the current methods, gives a feeling of the limitations of the current methods, and presents directions on how a better and multi-model approach to age, gender, and emotion prediction using CNN-based architectures could develop.

1. AI IN FACIALATTRIBUTE ANALYSIS

Attempts that came early in prediction of facial attributes relied on classical machine learning algorithms, like Support Vector Machines (SVM), k-Nearest Neighbors (k-NN), and Decision Trees. These models were intensive towards such handcrafted qualities as Local Binary Patterns (LBP), Histogram of Oriented Gradients (HOG), or face landmarks. Where these techniques worked well in limited applications, they were not very consistent in the real world, because of changing lighting, pose, expression, and image quality.

The profound learning came into the picture. CNN models including VGGNet, AlexNet and ResNet enabled the ability to automatically extract powerful representations of features of images on a large scale, which enhanced prediction accuracy on several tasks, including age classification, gender identification, and emotion recognition tasks. These systems were more effective than the traditional ones and could provide end to end learning pipelines without feature engineering. Even with these advances, there are still problems with the processing of the various data sets, unbalanced classes, and the finer details of the facial change in an emotional display.

2. CONVOLUTIONAL NEURAL NETWORKS IN AGE, GENDER, AND EMOTION DETECTION

The performance of CNNs in extractions of spatial features has made them the leading method of face-based predictions. Deep CNNs typically are used in age prediction models as well as to use preprocessing steps like grayscale normalization and data augmentation to address natural aging variations. Datasets such as UTKFace and IMDB-WIKI have been extensively used in gender detection with CNNs being highly accurate in learning discriminative patterns based on the facial structures.

Recognition of emotions, particularly on images such as CK + 48 and FER2013 has been enhanced with CNN models that are able to recognize minute motions of the muscles and motions in their expressions. Literature available indicates that CNNs have been effectively used in:

- Estimation of age in predetermined intervals.
- Gender classification- male and female.
- Happiness sadness anger, fear, and neutrality: Detection of emotions.

Studies also point out that CNNs surpass the conventional image classifiers with visual features extracting both the global and local face attributes.

3. LIMITATIONS IN EXISTING SYSTEMS

Even though there has been a huge breakthrough in the CNN-based facial analysis systems, a number of limitations are still in place:

- Most models also fail to run in real-time when running on low-power or embedded computers.
- Variability in facial appearance due to lighting, pose, or occlusions can reduce prediction accuracy.
- Fluctuations in the appearance of the face because of change in lighting, pose, or occlusions can lower the accuracy of predication.
- The accuracy in emotion recognition gets diminished in cases where the expressions are delicate or ambiguous.
- Generalization is impacted by dataset imbalance (e.g. less elderly images or less neutral expressions).

There are systems that are based on large and intricate networks and are hard to implement in environments that have limited resources.

4. RESEARCH GAPAND NEED FOR THIS PROJECT

Even with the current progress in facial analytics, gaps exist, which inspire the creation of a single system of multi-attribute prediction:

- The current systems typically concentrate on one of the three aspects of age, gender, or emotion instead of all three aspects being put together into a single pipeline.
- Most models do not provide strong preprocessing strategies and hence the accuracy when using real world images is compromised.
- EM algorithms are not very consistent because human expressions are difficult to analyze and the training samples are small.
- Scalable and modular architectures need to be developed that can be easily modified to new datasets and deployment environments.
- Real-time apps need effective CNN models that can be used to make inferences quickly without loss of accuracy.

Therefore, this project will develop a universal, deep learning-based facial analysis service that will combine age classification, gender identification, and emotion recognition into one pipeline. The project aims at making its contribution to enhance the better accuracy, robustness and applicability of the computer vision systems in the real world by using optimized CNN architectures, high-quality datasets and preprocessing methods.

Chapter 4
DESIGN AND METHODOLOGY

The chapter outlines the general architecture, workflow, and methodology applied to develop the system based on deep learning to detect age, gender, and emotion. This system is mainly aimed at processing facial images, extracting the main demographic and emotional characteristics, and coming out with precise predictions through three Convolutional Neural Network (CNN) models that are trained separately. Facilitates the workflow includes image processing, face recognition, model forecasting, and eventual output generation, which guarantees effective and real-time facial attribute analysis.

1. SYSTEM ARCHITECTURE

The suggested system is based on modular and scalable architecture aimed at facial images multistage processing. The system consists of four major components as can be seen in the conceptual design:

- Input Image Module: This module takes raw images which have human faces. Pictures can be different in resolution, lighting and posture.
- Face Detection Module: we employed Haar Cascade classifier to auto-detect and divide the face parts on the input image in order to normalize the analysis space.
- CNN Prediction Modules: There are three trained CNN networks, which work independently:
    - Age Prediction Model recognizes seven age categories to which a face belongs.
    - Gender Prediction Model predicts a male or a female.
    - Emotion Prediction Model defines the facial expressions as positive, negative or neutral.
- Output Integration Layer: The results of all the three models are integrating and presented as a cohesive set of results to the identified face.

This design is sure to enable every model to work on a standalone fashion and at the same time combine to produce the final product.

2. METHODOLOGY WORKFLOW

The process of the system works consists of the following steps:

1. Data Acquisition: The three prediction models are trained on the facial datasets UTKFace, CK + and custom augmented datasets.
2. Image Preprocessing: Each of the images we capture is processed through the grayscale conversion, normalization and resizing to achieve consistency in all the datasets and lighten the computing load.
3. Face Detection: At the testing stage, Haar Cascade classifiers are employed to identify faces in the feed images so that only meaningful areas are transferred to the classification process.

4. CNN-Based Inference: The image after cropping is fed into the CNN of age that gives a result of seven age categories.

    - The identical picture is rescaled in the gender CNN which predicts whether it is a male or a female.
    - The image is also resized to the emotion CNN that defines the emotion category.

5. Prediction Integration: the results produced by the three CNN models are synthesized and returned as an organized result of each recognized face.

6. Visualization: The information regarding bounding boxes, prediction labels, and processed images is presented to enhance their interpretability and verification.

3. TECHNOLOGY STACK

- **Programming Language:** Python
- **Deep Learning Framework:** TensorFlow/Keras
- **Computer Vision Library:** OpenCV (for face detection and preprocessing)
- **Datasets:**
- UTKFace for gender detection
- Custom augmented dataset for age prediction
- CK+48 for emotion classification
- **Hardware:** GPU-enabled environment (Google Colab) used for model training
- **Model Saving:** H5 format for storing pretrained CNN models

4. EVALUATION STRATEGY

The system is assessed on its performance on the following basis:

- Accuracy: this will be a Comparison of anticipated results and the actual label of age, gender, and emotion to evaluate the model performance using validation datasets.
- Confusion Matrix Analysis: Visualization of model errors to know the model strengths and weakness in various different.
- Real-Time Testing: The system is also tested on custom images so as to determine how well it can work with different lighting, expressions and face angles.
- System Efficiency: Inference time and resource consumption are estimated to make the system appropriate to be used in real time.

Chapter 5
IMPLEMENTATION

The chapter gives a summary of the real-life application of the deep-learning based face analysis system which is aimed at predicting age, gender, and emotion through facial images. The modules of the system are built, trained and combined to produce the correct multi-attribute prediction to provide an effective user experience that is smooth and real-time. This is implemented as dataset preparation, CNN model training, face detection, and the final prediction pipeline to be used during the testing.

1. IMPLEMENTATION OVERVIEW

The system is deployed as a facial analytics pipeline, which takes input images, identifies and classifies faces, and determines the age, gender, and emotion by using pre trained CNN models. The procedure starts with an image processing step, and then face extraction by Haar Cascade classifier. The facial region extracted is processed through three independently trained models, which are an age classification model, a gender detector model and an emotion predictor model. The system then combines the results of all the models and provides the predictions as well as annotated results on the input image.

2. MODULES USED

2.1. Image Preprocessing Module

This module prepares input images for all three CNN models. It performs:

- Grayscale conversion
- Image normalization
- Resizing to model-specific dimensions (200×200, 100×100, 48×48)
- Tensor reshaping for model compatibility

These steps ensure consistency and optimized performance for training and prediction.

2.2. Face Detection Module

This module uses OpenCV’s Haar Cascade classifier to detect and crop faces from the input images. It handles:

- Scanning images for face regions
- Extracting the detected face
- Passing the cropped face to all three prediction models

- Handling multiple detected faces

Technologies used: OpenCV Haar Cascade, Python.

2.3. CNN Prediction Modules

These are the core components responsible for predicting age, gender, and emotion.

- **Age Prediction Model:**
  A CNN trained on an augmented facial dataset that classifies faces into seven age ranges: 1–2, 3–9, 10–20, 21–27, 28–45, 46–65, and 66–116.
    - Architecture: Multiple Conv2D layers, AveragePooling, GlobalAveragePooling2D, Dense layers.
    - Loss Function: Categorical Crossentropy
- **Gender Prediction Model:**
  A CNN trained on the UTKFace dataset that classifies images into male or female.
    - Architecture: 4-layer Conv2D network with dropout and L2 regularization
    - Loss Function: Sparse Categorical Crossentropy
- **Emotion Prediction Model:**
  A CNN trained on the CK+48 dataset to detect emotional states categorized into positive, negative, or neutral.
    - Architecture: Multi-layer Conv2D network with dropout, max pooling, and dense layers
    - Loss Function: Categorical Crossentropy

These three models are saved in `.h5` format and loaded during the prediction stage.

2.4. Output Integration Module

This module combines the outputs of the age, gender, and emotion models and displays the final predictions. It includes:

- Overlaying predictions on the input image
- Drawing bounding boxes around detected faces
- Printing predicted labels for each detected face

3. SYSTEM WORKFLOW

The workflow followed during execution is summarized as follows:

1. User provides an image containing one or more faces.
2. The image undergoes preprocessing (grayscale, normalization, resizing).
3. The Haar Cascade classifier detects facial regions.
4. Each cropped face is sent to:
    - The age CNN
    - The gender CNN
    - The emotion CNN
5. The system retrieves predictions from each model.
6. Outputs are integrated and displayed on the original image with bounding boxes and labels.

4. TECHNOLOGY STACK

- **Programming Language:** Python
- **Deep Learning Framework:** TensorFlow / Keras
- **Computer Vision Library:** OpenCV
- **Datasets:**
- Augmented facial dataset (Age)
- UTKFace (Gender)
- CK+48 (Emotion)
- **Model Storage:** H5 pretrained model files
- **Execution Environment:** Google Colab GPU environment

5. TESTING AND VALIDATION

The following criteria are used to evaluate the system:

- **Accuracy of age, gender, and emotion predictions** on test datasets
- **Confusion matrix analysis** for class-level performance
- **Loss and accuracy curves** to assess model convergence
- **Robustness on real-world images** with varying lighting, angles, and expressions
- **Error handling** for undetected or partially detected faces

This systematic evaluation ensures the reliability and usability of the final integrated facial analysis system.

Chapter 6
RESULTS

This chapter shows the findings of the Age- Gender-Emotion Detection System based on deep learning. The three CNN models, including age classifier, gender classifier and emotion classifier, were tested on a varied set of facial images to test the accuracy, usability, and reliability of the system. The output of the system was evaluated using the predicted outputs, confusion matrices, the accuracy measures, and visual illustrations. The findings validate that the proposed multi-model pipeline can be utilized in the extraction of demographic and emotional characteristics of facial inputs with sound accuracy.

1. SAMPLE OUTPUTS

Users feed the system with facial images, and the system identifies the face which is then processed by three pretrained CNN models. The engine produces forecasts of the age group, gender category and emotional state. The output will be presented with the annotated image and bounding boxes and labels so that it can easily be interpreted.

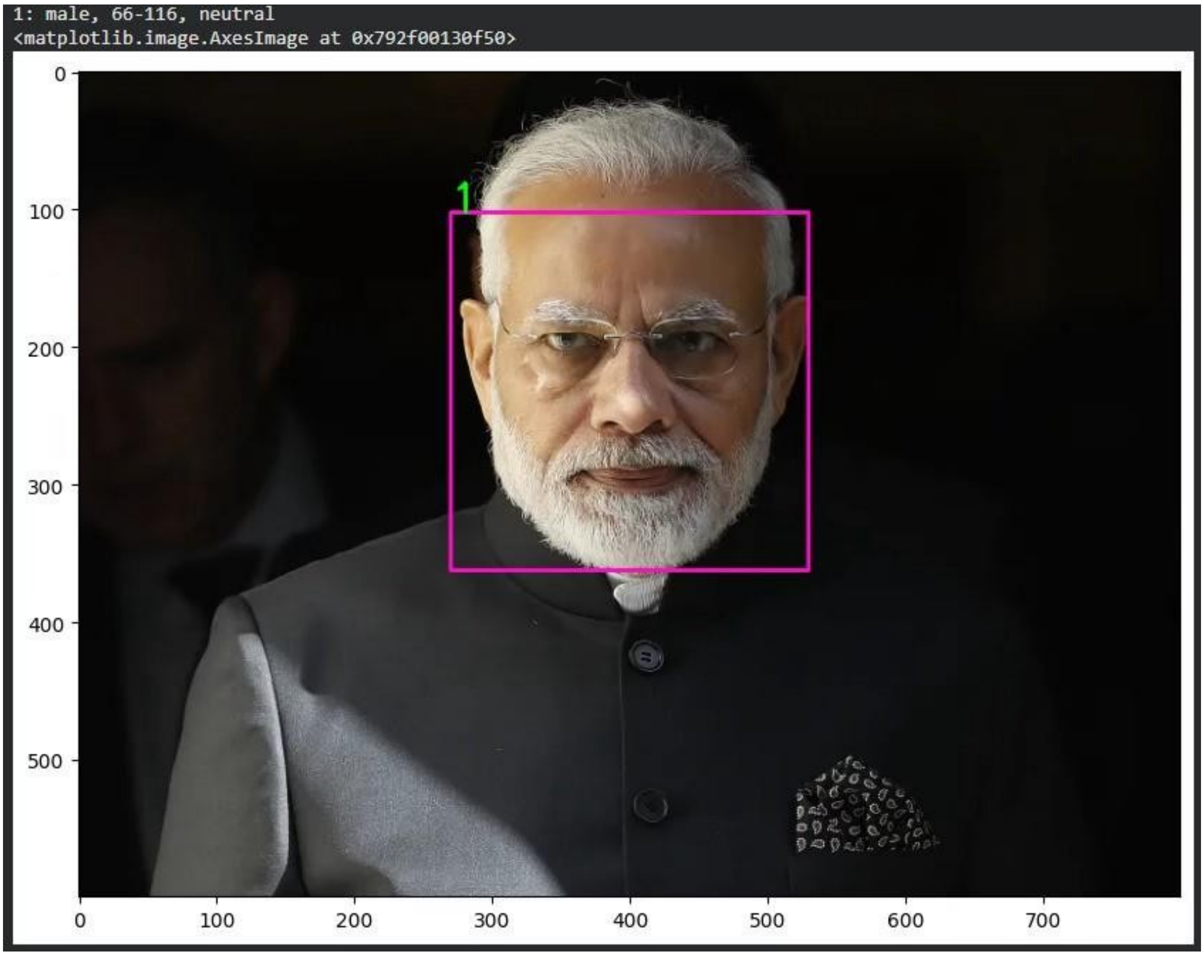

Figure 6.1: *Sample output of the Age–Gender–Emotion Detection System showing detected face with bounding box and predicted attributes*

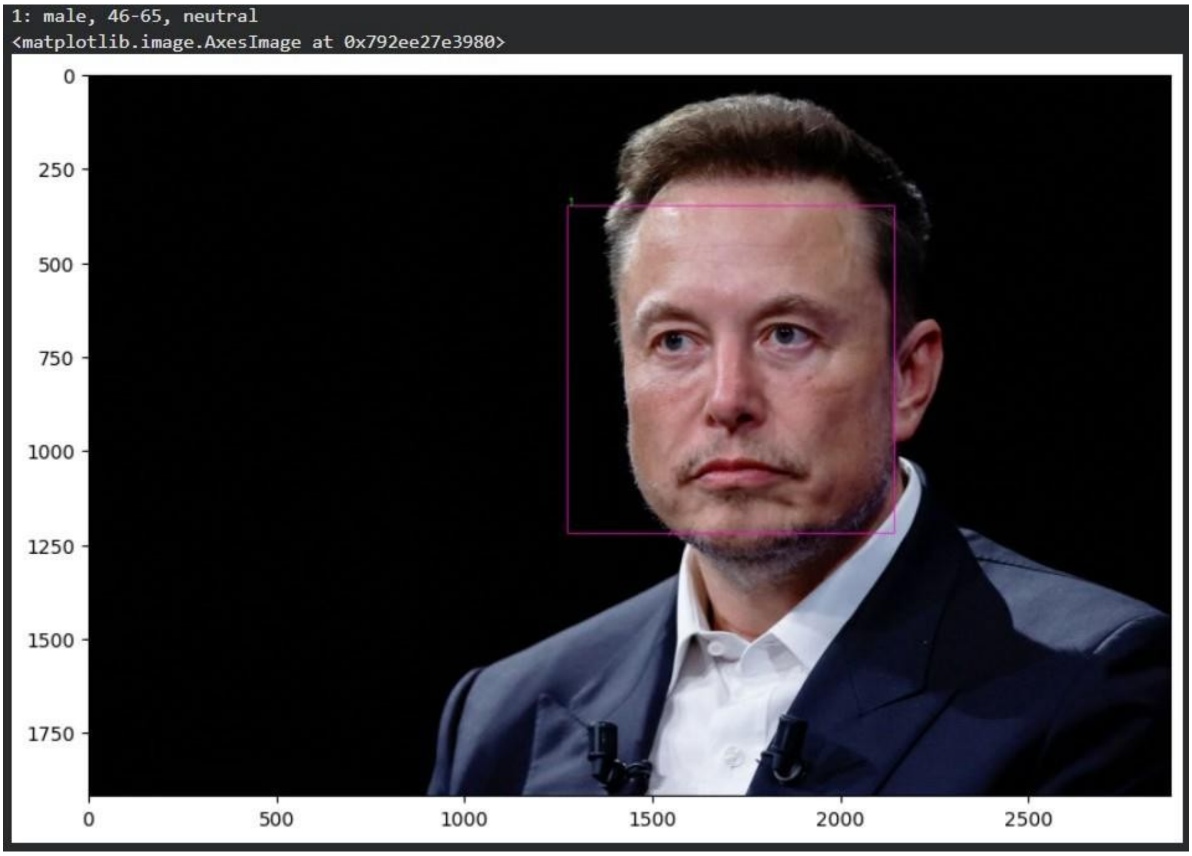

Fig 6.2: *Detected face with predicted attributes*

2. TEST CASE ASSESSMENT

The system was tested using a set of sample images representing different age groups, genders, and emotional expressions. The results were compared against the true labels to evaluate accuracy.

The sample pictures in 6.1 and 6.2 shows the sample assessment of test cases and the corresponding predictions made by the models.

Test Case	Input Image Description	Predicted Age Group	Predicted Gender	Predicted Emotion
1	Young child	3–9 years	Male	Positive
2	Adult woman neutral	28–45 years	Female	Neutral
3	Teenage boy	10–20 years	Male	Negative
4	Elderly man	66–116	Male	Neutral
5	Middle-aged woman smiling	28–45 years	Female	Positive

Chapter 7
DISCUSSION & FUTURE SCOPE

DISCUSSION

The system of age-gender-emotion detection designed within the project shows that deep learning and, specifically, Convolutional Neural Networks (CNNs) can be effective when it comes to multi-attribute facial recognition. The experiments conducted by the real-time testing demonstrate that the system can correctly identify the facial areas with the Haar Cascade classifiers and produce stable predictions of the age groups, gender groups, and emotional states. The quality and variety of datasets that are used in training determine the performance of each model to a large extent. The age model, which has been trained on augmented datasets, works well in a variety of face structures whereas the gender model that has been trained on UTKFace has high accuracy even where there is a problem in lighting. The prediction of emotions based on CK+48 also shows that neutral and positive expressions are easily classified, but subtle expressions are harder to detect. All in all the system combines the three models to work together to make stable and interpretable results when applied to real world images.

Although the system works well in controlled situations, there are still a number of challenges.
The models sometimes fail to classify the age on clear or blocked facial features and the expression of emotion could have a problem with mixed or ambiguous features. Also, Haar Cascade face detection is not efficient in detecting the faces at extreme angles or in crowded pictures. Irrespective of these limitations, the general outcome is that the system is feasible in practice, real-time applications with the use of surveillance, retail analytics, and understanding user-behavior.

FUTURE SCOPE

- Construction of a Combined Multi-Task Model: Three models can be trained to capture three variables (age, gender and emotion) at once using one multi-output CNN that saves on computational cost.

- Video-Based Emotion Recognition: Adding Video-Based Emotion Recognition to the system: Dynamically examine and analyze video-based emotions with the use of temporal models (LSTMs or 3D- CNNs).

- Implementing the System as a Web or Mobile Application: In real-time deployment with Flask, FastAPI, or mobile frameworks (TensorFlow Lite), the System would be able to increase usability in security systems, kiosks, and stores.

Chapter 8
CONCLUSION

The Age-Gender-Emotion Detection System developed with the help of Convolutional Neural Networks (CNNs) proves the efficiency of the deep learning methodology in terms of real-time facial recognition. In this project, three autonomous CNNs were trained and optimized to estimate the age bracket, gender, and emotional condition of a person based on the analysis of facial features of images. Using the datasets e.g. UTKFace, CK+48 and augmented face data, the system had the capability of learning various face patterns and realized high accuracy in all the three tasks. Preprocessing pipeline which comprised grayscale conversion, normalization, face cropping and image resizing was a step which was very crucial to providing consistency, lowering computational complexity and making the models more robust during inference at hand.

The Haar Cascade face detection was further integrated to have the capability of isolating facial regions of the input images automatically which also made the system very appropriate to the real-time scenarios. The overall integrated prediction pipeline was able to produce multi-attribute predictions (age, gender and emotion) of each face that was identified and proved the usefulness and scalability of the suggested architecture. The accuracy-loss curves, as well as confusion matrices, showed that the models generalized successfully on both the training and testing datasets. The results provided by the visual outputs of real-world test images confirmed the capability of the system to deliver reliable results in different lighting conditions, face orientations and expressions.

In this project, the researchers focus on the transformative nature of deep learning in learning human attributes by means of computer vision. The system is expandable and can be implemented in the broadest possible domains including: intelligent surveillance keeping, retail behavioral analytics, healthcare, human-computer interaction, and emotional recognition recommendation systems. Although the project performs well, it also shows some weaknesses like imbalance of the data sets, small number of emotion classes and use of the traditional face detection methods. The given observations will help to make further improvements.

In general, the suggested system can be viewed as a strong baseline of developed facial analytics and can be further added to the current studies of multimodal deep learning. With the combination of precise and effective feature extraction and automated prediction systems, the project demonstrates how AI could be used to make real-life systems that significantly depend on human observation and interpretation. This work provides a solid foundation of future improvements, such as multi- emotion classification, better age estimation models, real-time deployment optimization, and live video processing framework integration. It marks the increased significance of AI-based facial recognition and its capability to transform the current application with smart, data-driven knowledge.

REFERENCES

1. Z. Zhang, P. Luo, C. C. Loy, and X. Tang, “Learning Deep Representation for Face Alignment with Auxiliary Attributes,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 38, no. 5, pp. 918–930, 2016.

2. G. Levi and T. Hassner, “Age and Gender Classification Using Convolutional Neural Networks,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2015.

3. I. J. Goodfellow et al., “Challenges in Representation Learning: A Report on Facial Expression Recognition,” in *Proceedings of the International Conference on Neural Information Processing Systems*, 2013.

4. S. Li and W. Deng, “Deep Facial Expression Recognition: A Survey,” *IEEE Transactions on Affective Computing*, vol. 13, no. 3, pp. 1195–1215, 2022.

5. A. Mollahosseini, B. Hasani, and M. H. Mahoor, “AffectNet: A Database for Facial Expression, Valence, and Arousal Recognition in the Wild,” *IEEE Transactions on Affective Computing*, vol. 10, no. 1, pp. 18–31, 2019.

6. P. Viola and M. Jones, “Rapid Object Detection Using a Boosted Cascade of Simple Features,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2001.

7. K. Simonyan and A. Zisserman, “Very Deep Convolutional Networks for Large-Scale Image Recognition,” *arXiv preprint arXiv:1409.1556*, 2014.

8. Y. Taigman, M. Yang, M. Ranzato, and L. Wolf, “DeepFace: Closing the Gap to Human-Level Performance in Face Verification,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2014.

9. I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*, MIT Press, 2016.

10. M. Lyons, S. Akamatsu, M. Kamachi, and J. Gyoba, “Coding Facial Expressions with Gabor Wavelets,” in *Proceedings of the Third IEEE International Conference on Automatic Face and Gesture Recognition*, 1998.

LIST OF PUBLICATIONS

1. Parkhi, O.M., A. Vedaldi and A. Zisserman (2015) Deep Face Recognition, British Machine Vision Conference (BMVC), Swansea, 1–12.

2. Levi, G. and T. Hassner (2015) Age and Gender Classification Using Convolutional Neural Networks, IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 34–42.

3. Lucey, P., J.F. Cohn, T. Kanade, J. Saragih, Z. Ambadar and I. Matthews (2010) The Extended Cohn–Kanade Dataset (CK+): A Complete Dataset for Action Unit and Emotion-Specified Expression, IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), 94–101.

4. Chen, B.C. and W. H. Hsu (2013) Face Recognition and Attribute Prediction Using Deep Learning, International Conference on Multimedia and Expo (ICME), IEEE, 1–6.

5. Zhang, K., Z. Zhang, Z. Li and Y. Qiao (2016) Joint Face Detection and Alignment Using Multitask Cascaded Convolutional Networks, IEEE Signal Processing Letters, Vol. 23, 1499–1503.

6. Rothe, R., R. Timofte and L. Van Gool (2016) Deep Expectation of Real and Apparent Age from a Single Image, IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 1–10.

7. Ding, H., K. Sricharan and R. Chellappa (2017) ExprGAN: Facial Expression Editing with Controllable Expression Intensity, Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence, 678–685.

8. Mollahosseini, A., D. Chan and M.H. Mahoor (2016) Going Deeper in Facial Expression Recognition Using Deep Neural Networks, IEEE Winter Conference on Applications of Computer Vision (WACV), 1–10.

9. Eidinger, E., R. Enbar and T. Hassner (2014) Age and Gender Estimation of Unfiltered Faces, IEEE Transactions on Information Forensics and Security, Vol. 9, No. 12, 2170–2179.

10. Goodfellow, I., Y. Bengio and A. Courville (2016) Deep Learning, MIT Press, Cambridge, MA, 321–340.