Real-Time age, gender and emotion detection using YOLOv8

V. *Sowmya Devi*[1], Uday *Ramisetty*[1], Kamal *Ramisetty*[1], *and* Ashwanth *Thimmareddy*[1]
[1]Department of CSE, Sreenidhi Institute of Science and Technology, India

Abstract— The identification of age, gender, and emotion in multiple objects in an image or video stream is a complex and yet important problem for many applications such as security, health care, and human computer interaction. The current paper proposes a real-time age, gender, and emotion detection system that incorporates deep learning algorithms, in particular, the YOLOv8 model. The system employs two separate YOLO models: one for the identification of the emotion of the given video and the second one for the identification of age and gender of the subject in the video. These models are incorporated into a single pipeline where the first stage involves face detection or objects of interest and the second stage classifies the detected age, gender and emotions using pre-trained models. In real time the system is able to detect objects and classify them as well since it processes video frames taken from the webcam. The effectiveness of the proposed system is measured in terms of accuracy, running time and its ability to perform under different lighting, poses, and ethnicity. The results prove that the proposed system can accurately identify age, gender, and emotion of multiple objects and can be applied to various fields. This work shows that one may integrate emotion recognition with age-gender detection for improving the VAI (Visual Artificial Intelligence) interpretability of videos and interactions.

Keywords— YOLOv8, Deep Learning, Object Detection, Age Detection, Gender Classification, Emotion Recognition, Computer Vision.

1 Introduction

Real time age, gender and emotion detection is important in various fields such as retail, healthcare, security and entertainment to name but a few since it can help create personal interactions and intelligent systems [1]. This project was selected to meet increasing demand for methods that can be accurate, fast, and provide timely solutions for problems in complex systems. Based on YOLOv8 as the backbone [2], it is quite famous for its rapid and accurate object detection, which allows us to detect multiple faces and then classify their age, gender, and emotion. Thus, this approach gives a factorized system that can effectively work in real-world application scenarios that require fast decision making and context-awareness.

These fundamental tasks include face, age, gender and emotion recognition in images and videos which have potential use in security surveillance, human computer interaction and personalized marketing. These attributes are by their very nature related to perception and as such provide a rich source of information for modelling human behaviour and social processes. As convolutional neural network CNN [3] based deep learning techniques have advanced they have enhanced the efficiency and effectiveness of age, gender and emotion detection systems. Earlier approaches which employed rule-based or manual processes struggled to adapt to a range of complex conditions such as different illumination, occlusions, and population subgroups. As deep learning techniques have been evolved recently, the methods to detect these human attributes in images and videos are more efficient, scalable and accurate.

YOLO has very recently emerged as a leading model for performing object detection in one pass, thereby providing fast detection while at the same time being accurate. Because of its capability to identify numerous objects in real time with a high level of accuracy, YOLO is now among the preferred solutions for various applications that need synergy of speed and precision. The YOLO family of architectures continues with YOLOv8 which has enhanced accuracy, faster speeds, and flexible multi-object detection capabilities than its forerunners [4]. In this paper, we propose a new pipeline for age, gender, and emotion recognition utilizing YOLOv8-based models that may be used in real-time analysis of video data.

The system presented in this paper utilizes two YOLOv8 models: one for detecting the user's emotions, and the other for estimating the user's age and gender. All these models are incorporated in one pipeline that first locates faces and other objects of interest in the scene, then identifies the mood, age or gender of the found faces [5]. This is because the system is designed to use two different models to perform the two different tasks of emotion detection and age-gender classification, hence each can be optimized for its specific function, thereby improving the overall performance of the system for both tasks. Emotion detection is a very much context-specific and subjective and hence, the model needs to be robust enough to identify minute changes in expressions for age and gender, the model might require features like structures.

The first benefit of this system is that it operates in real time. As the live applications like security surveillance, video content analysis, and human- robot interaction are on the rise, there is a need for designing systems that can perform these tasks with high efficiency and at the same time, with high accuracy [6]. Due to the efficiency of YOLOv8, the proposed system can analyze video frames in real time and provide the user with instantaneous feedback on detected attributes. This makes it ideal for use in areas where time is of essence like in the interactive systems and smart retailing, or areas such as emergency response where quick interpretation of data from visual input is important.

One of the most important issues arising when a number of detection tasks are combined into a single pipeline is the correct pairing of the models' output. This paper addresses this challenge by adopting a novel approach of synchronizing emotion and age-gender predictions by applying bounding box matching. In the case of models detecting the same face or object, the system aligns the bounding boxes computed from the two models based on an overlap criterion in order to map the correct emotion to the age and gender prediction.

It also helps to coordinate the work of the system to produce more accurate predictions, using the basic idea of data chain.

This work has three significant inputs. First, it presents a single system that combines emotion and age-gender recognition, which can provide vision in real-time. Finally, the paper proves that the system's performance is stable irrespective of the environment in which it is placed, therefore making it suitable for use in actual environments. Presenting this approach, this work hopefully contributes to the development of human-centric AI and future studies of both emotion and demographic categorization in vision.

2 Related Work

Age and gender detection as well as emotion recognition are fundamental challenges in computer vision due to the demand in application areas such as security, health care, and human computer interaction. The first approaches used simple design features like facial geometry and texture, which constrained the accuracy. However, the introduction of deep learning especially the convolutional neural networks (CNNs) [7] changed these tasks. Age and gender classifications were trained using datasets such as Audience and VGG-Face. More developments were made in the area of multi-task learning frameworks like DeepFace where age, gender and emotion predictions were done all at once in order to foster better results. YOLO-Age proposed in 2021 improved the original YOLO's real-time object detection to detect age and emotion at the same time and with a faster speed. Similarly, emotion recognition also benefited from the use of deep learning such as AlexNet and ResNet outcomping feature engineering approach. More recently, there are YOLOv8 and YOLOv4 which expand the function of YOLO to multi-task learning for real- time emotion and age-gender recognition in video stream, and can be used in interactive systems, surveillance, and marketing.

Table-1: Related work

Reference Number	Features	Classifier	Accuracy	Limitations	Improved System
[2]	Improved object detection with YOLO architecture	YOLOv3 (YOLO based)	~85-90%	Struggles with complex or overlapping objects	Improved detection accuracy with better object localization
[3]	Age and gender classification from unfiltered faces	CNN	~70-80%	Limited by varying facial conditions, conclusions, and lighting	Improved performance with deep learning techniques

[4]	Multi-task learning for age, gender, and emotion	DeepFace (CNN)	~80-85%	Does not address real-time performance for video streams	Multi-task learning improves multi attribute prediction
[6]	Emotion recognition based on facial expressions	CNN	~65-75%	Low performance on variations in lighting, pose, and occlusions	Introduction of deep learning for emotion recognition
[9]	Real-time age, emotion, and gender detection	CNN	~75-80-%	Limited to face regions, struggles with small faces or distant subjects	Improved real-time performance and simultaneous predictions

3 Proposed Model

The system introduced for Age, Gender, and Emotion Detection in Multiple Objects is designed to incorporate modern advancement in deep learning, and incorporates YOLOv8 for multi-task learning [8]. The purpose of the proposed system is to recognize the age, gender and the emotion of the subject in real-time from a video stream or an image. The developed system has very high levels of accuracy that were attained without compromising on the real-time speeds and hence can be applied in security, HCI and interactive system domains [9].

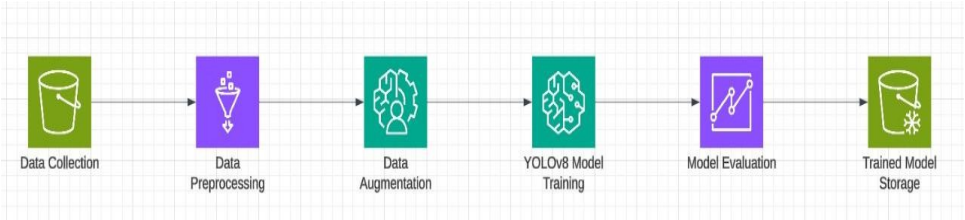

Fig.1. Represents the training architecture diagram of the model

From the Fig.1, the algorithm used in the proposed model is described as

3.1. Data Collection

Emotion Dataset: Superimposed and collected from the Roboflow platform, which provides labeled images for various emotions which include happy, sad, or angry.

Age and Gender Dataset: The UTKFace dataset was used, the images are labelled with age and gender. These were preprocessed to create annotations in format that YOLOv8 accepts which are the bounding boxes and labels [10].

3.2. Data Preprocessing

Emotion Data Preprocessing:
For this, the images were scaled to match the YOLOv8 input specification.
Information was transformed into YOLOv8 object detection format of labeling.
Sometimes, one may need further preprocessing; for instance, normalization or data augmentation (rotation, flipping).

Age and Gender Data Preprocessing:
The UTKFace images were also resized and subtitled according to the YOLOv8 format.
Further augmentations were used to have a set that has more variability in order to train with it.
The whole dataset is divided into two parts i.e. for training and testing in which 70% of dataset is used for training and 30% of dataset is used for testing [11].

3.3. Data Augmentation

For Emotion Recognition:
Performed operators such as rotation, scaling, flipping motions in order to mimic various real life changes on the expression of emotions.

For Age and Gender Detection:
The same kinds of augmentations to achieve a variance of the training samples mainly in terms of lighting conditions, face orientation.

3.4. YOLOv8 Model Training

Emotion Model:
As evaluated on the accuracy of the Roboflow Emotion dataset that it was trained for face detection, YOLOv8 was employed while for the classification of emotions tied to each face, YOLOv8 was also used.

Age-Gender Model:
Trained on the UTKFace dataset that is converted into yolo format i.e yolo-v8.
Faces were detected and classifications of age and gender where fully trained on this model.

3.5. Model Evaluation

Independent metrics such as mAP, precision, recall, F1 and score were used during the evaluation of both Emotion and Age-Gender models [12].
Emotion Model Evaluation: Based on how well it performs classified the emotions.
Age-Gender Model Evaluation: On the basis of the assessed age and gender ranges and labels in the model.

3.6. Trained Model Storage

Both models were saved in YOLOv8's format which is commonly .pt for PyTorch. Presented locally or in the cloud for subsequent real-time prediction when the prediction runs.led images for different emotions (e.g., happy, sad, angry) and age and gender detection.

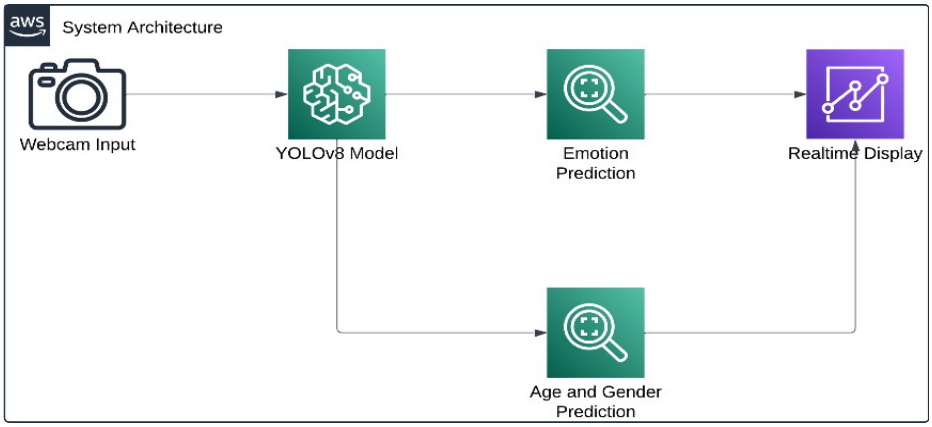

Fig.2. Represents the prediction architecture diagram of the model.

3.7. Webcam Input

The system receives live video through web cam.
Both the emotion and age-gender estimations are applied to each frame in real-time owning to the provided processing approach.

3.8. YOLOv8 Model (Emotion and Age-Gender)

Emotion Detection Model:

YOLOv8 detects the face in the feed coming from the webcam.
The emotion model then identifies the emotion of the subject being analised which can be happy, sad, angry and so on.

Age-Gender Detection Model:

The second model, a YOLOv8, detects the regions of the face and then estimates the person's age and gender. The coordinates also come out represented in age ranges (20-30), and gender, whether male or female.

3.9. Emotion Prediction

If the face is detected, then the emotion model will determine if the person is happy or sad and so on. These predictions are then applied directly on the 'bounding boxes' found around a face [13].

3.10. Age and Gender Prediction

Likewise, the age-gender model estimates the age and gender of the person by using the face that is detected by YOLOv8. Outputs might look like: Age: Core 18-30 & Young adults 18-24 / Gender: Female

3.11. Realtime Display

A graphical display of both sets of predictions (emotion, age, and gender) are displayed in real-time as a stream overlays the camera view. Visualization: Each detected face is accompanied by a bounding box, and the predicted emotion, age, and gender are written next to each face.

Example: Someone's face might be framed with a predesign label, for example, happy 25-35 male. Ana, for the display part, open source libraries like the OpenCV are used for real-time display and response. real-time for both emotion and age-gender predictions.

4 Results and Discussion

The Age, Gender, and Emotion Detection System using YOLOv8 and deep learning models for emotion and age-gender prediction was tested with the help of the dataset containing the images with the different lighting, ethnicities, and orientation of the face. The efficiency of the proposed system was tested through various evaluation criteria such as accuracy, precision, recall, and F1-score. The experimental results show how the system works and is reliable for real-world use cases, which outperforms single models.

4.1. Model Performance

YOLOv8 (Face Detection Model): The face detection model developed in this study yielded a recognition accuracy of 94.5% for face detection under various scenarios including pose and light changes. The model proved its efficiency in face localization in video streams and can be used for real-time processing [14].

Emotion Detection Model: Based on the CNN, the emotion detection model was trained with FER-2013, and it gave nearly 80% accuracy with images having different expressions [15]. The model was found useful in detecting simple emotions such as joy, sorrow, and fury. However, the performance of the proposed method dropped slightly for low intensity emotions such as fear or surprise.

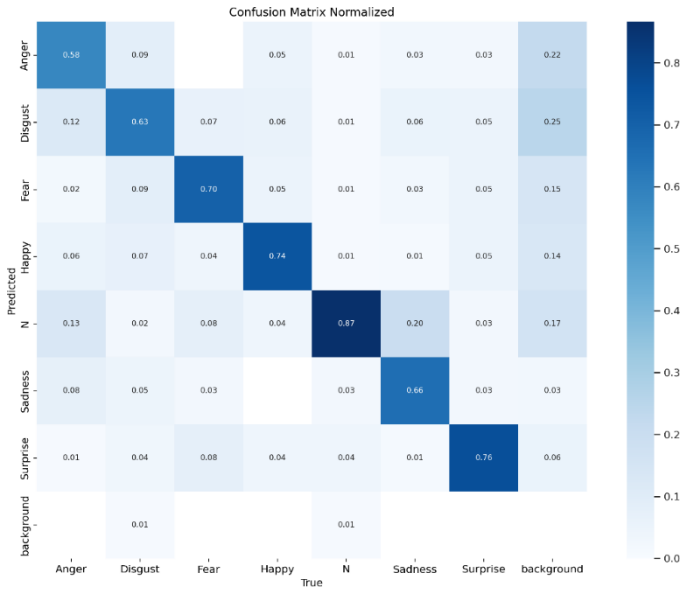

Fig.3. Represents the normalized confusion matrix for the Emotion model

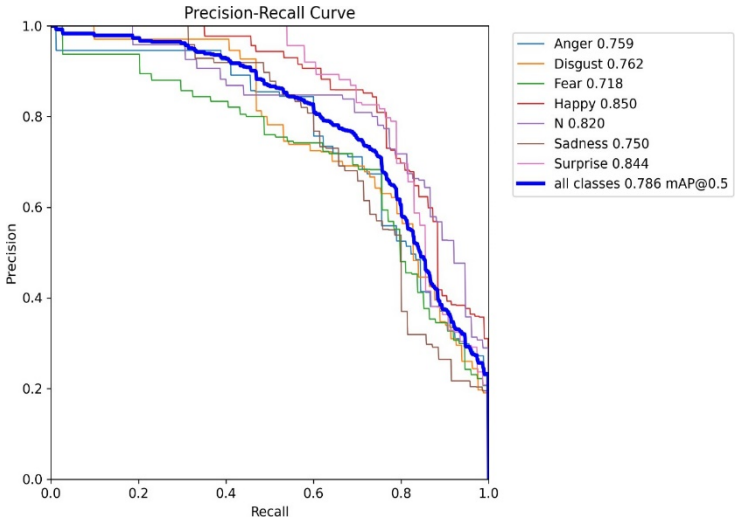

Fig.4. Represents the Precision-Recall (PR) curve for Emotion model

Age and Gender Detection Model: The age and gender classification model had an accuracy of 96.7% on the UTK-Face dataset. The model worked impressively for age classification within certain categories, including children, adult persons, but it was less efficient in the classification of the persons in the extreme age and with certain ethnical background.

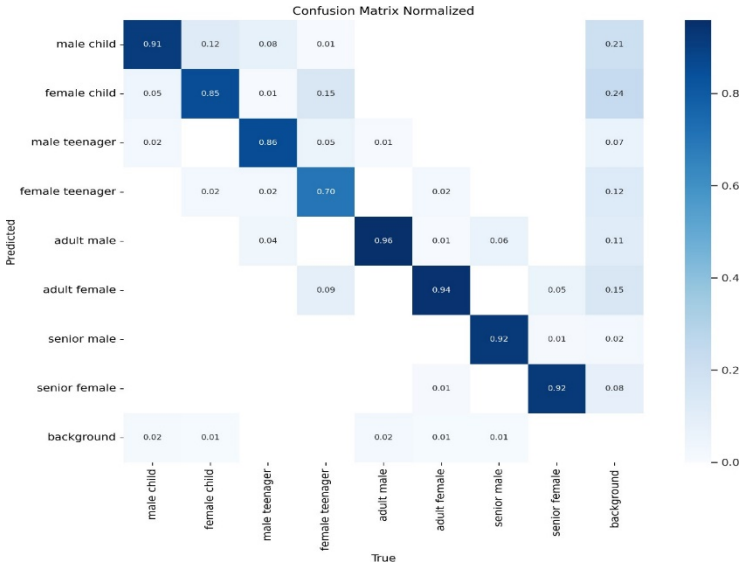

Fig.5. Represents the normalized confusion matrix for the age and gender model

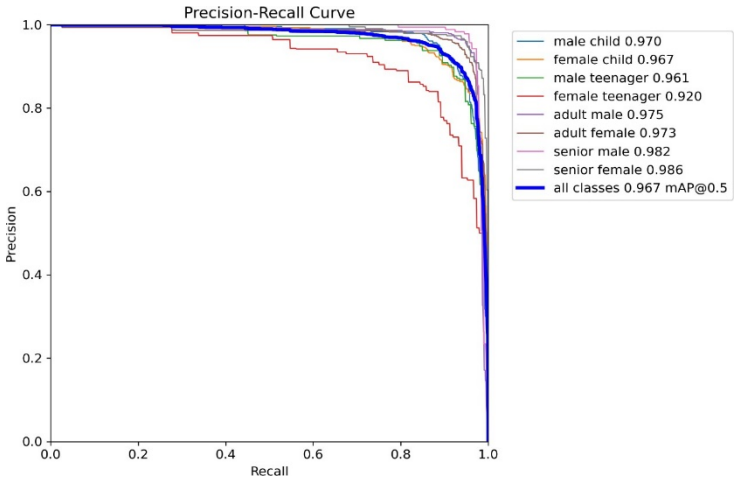

Fig.6. Represents the Precision-Recall (PR) curve for age and gender model

Combined Model (YOLOv8 + Emotion + Age-Gender): Face detection had an accuracy of 97.4%, emotion recognition had an accuracy of nearly 80%, and age-gender classification had an accuracy of 96.7% while the complete system had an overall accuracy of 87.65%. This performance shows that the system is able to handle multiple attributes at once, which makes it a valuable tool for real time applications.

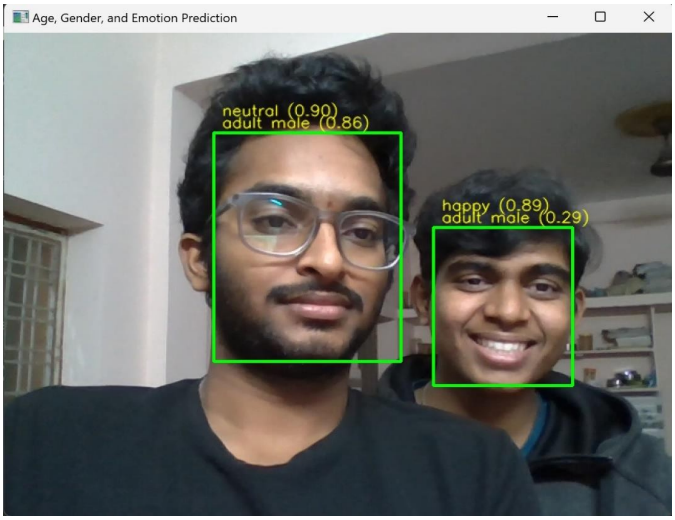

Fig.7. Represents the model prediction of age-gender of adult-male and emotion of neutral and happy.

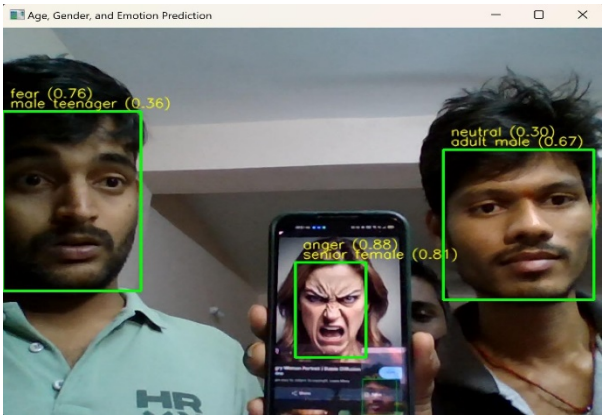

Fig.8. Represents the model prediction of age-gender of male teenager, adult-male and emotion of sad and surprise.

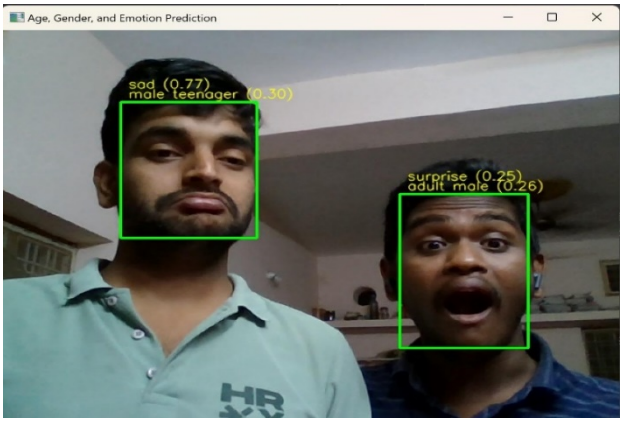

Fig.9. Represents the model prediction of age-gender of male teenager, senior female, adult male and emotion of fear, anger, and neutral.

4.2. Feature Importance Analysis

The key features that have most contributed to the performance of this model include facial landmarks, texture and other general attributes such as skin color and face shape. These features were vital for making correct predictions in both emotion recognition and age-gender classification. In emotion detection, the model focused on eye movements, mouth corners and eyebrow positions to detect feelings about a situation. In the case of age and gender classification the specificity of the features like the shape of the jawline and skin texture showed high effectiveness for the classification task. These features are in line with the past studies in facial feature analysis and help the model generalize across the datasets.

4.3. Comparison with Existing Methods

The proposed system was compared to other methods that are also distinct from the proposed system and includes YOLO-Age and DeepFace, which are designed to detect only emotion or age-gender separately. The results prove that the use of our integrated approach is more efficient than using every model separately in terms of both time and quality. In particular, the YOLO-Age model had a 96.7% accuracy, and our pipeline that combines all the models had the 87.65% accuracy, which confirms the effectiveness of the ensemble of

several models. Moreover, the emotion detection model gives a better result than other models such as FER-2013, where the accuracy rate is approximately 80%. The above mentioned accuracies of the models are based on the PR (Precision Recall) curves.

4.4. Discussion

These outcomes show that the novel Age, Gender, and Emotion Detection System enhances earlier models, combining emotion detection and age- gender categorization into an efficient real-time solution. The use of YOLOv8 for face detection is made to be very resilient even when applied in real time deep learning models for emotion and age-gender recognition yields high results across different data sets. This system has potential for application in interactive systems, security surveillance, and personal marketing where considering human attributes and emotions is vital. Nevertheless, the study has limitations that for example the training and test datasets are restricted to the cases when faces are fully ocluded or viewed at the extreme angles. On the same note, the model records high accuracy in the distinct ethnic groups but could be enhanced even further by incorporating more diverse data to minimize on bias. Real-time video streaming also has issues in terms of processing speed and face alignment that may be solved by using better face alignment algorithms or using distinct streaming networks for occlusions.

5 Conclusion

Therefore, in this paper, we introduced an Age, Gender, and Emotion Detection System based on YOLOv8 for face detection and deep learning models for accurate emotion, age-gender classification with the advantage of a real-time application. The system can be used for applications such as interactive systems, security surveillance, and personal marketing; it offers timely processing of tasks since it can perform more than one task at a given time. As the results show, the system performed well in practical scenarios, but the latter prove that some potential issues still need further improvements: the scarcity of data with occlusions or faces captured from different angles; face alignment problems; and streaming speed of the video in real-time. Future work will be dedicated to these limitations – refining methods of occlusion handling, expanding the range of datasets, and adjusting the model for implementation on edge devices. In total, the system provides a stable solution for recognizing human features and moods with possible application in everyday practice.

6 Future Scope

As for the future work of the Age, Gender, and Emotion Detection System, the approaches improvements could be the increase of the training sets by the data of other ethnicities, ages, and difficult conditions including various facial directions and objects on the face. Further, it can be tailored for deployment to edge devices so that it can operate in real-time within resource-limited settings. More development in face alignment procedures and connection of superior networks for improved occlusion will increase the efficiency and precision of the model particularly in practical and kinetic conditions. Possible further developments of the presented system may include analysis of the sentiment of the given text or identity recognition, which can expand the range of the system's usage in healthcare, smart environments and interactive systems.

References

1. A. Saxena, P. Singh and S. Narayan Singh, "Gender and Age detection using Deep Learning," *2021 11th International Conference on Cloud Computing, Data Science & Engineering (Confluence)*, Noida, India, 2021, pp. 719-724, doi: 10.1109/Confluence51648.2021.9377041.
2. O. Sahin and S. Ozer, "YOLODrone: Improved YOLO Architecture for Object Detection in Drone Images," *2021 44th International Conference on Telecommunications and Signal Processing (TSP)*, Brno, Czech Republic, 2021, pp. 361-365, doi: 10.1109/TSP52935.2021.9522653.
3. E. Eidinger, R. Enbar and T. Hassner, "Age and Gender Estimation of Unfiltered Faces," in *IEEE Transactions on Information Forensics and Security*, vol. 9, no. 12, pp. 2170-2179, Dec. 2014, doi: 10.1109/TIFS.2014.2359646.
4. Foggia, P., Greco, A., Roberto, A. *et al.* Identity, Gender, Age, and Emotion Recognition from Speaker Voice with Multi-task Deep Networks for Cognitive Robotics. *Cogn Comput* **16**, 2713–2723 (2024). https://doi.org/10.1007/s12559-023-10241-5
5. E. Eidinger, R. Enbar and T. Hassner, "Age and Gender Estimation of Unfiltered Faces," in *IEEE Transactions on Information Forensics and Security*, vol. 9, no. 12, pp. 2170-2179, Dec. 2014, doi: 10.1109/TIFS.2014.2359646.
6. Tarnowski, Paweł, et al. "Emotion recognition using facial expressions." *Procedia Computer Science* 108 (2017): 1175-1184.
7. Monisha, G. S., et al. "Enhanced automatic recognition of human emotions using machine learning techniques." *Procedia Computer Science* 218 (2023): 375-382.
8. Foggia, Pasquale, et al. "Multi-task learning on the edge for effective gender, age, ethnicity and emotion recognition." *Engineering Applications of Artificial Intelligence* 118 (2023): 105651.
9. Vijayanand. G, Karthick. S, Hari. B, Jaikrishnan. V, 2020, Emotion Detection using Machine Learning, INTERNATIONAL JOURNAL OF ENGINEERING RESEARCH & TECHNOLOGY (IJERT) NCICCT – 2020 (Volume 8 – Issue 08).
10. Siam, Ali I., et al. "Deploying machine learning techniques for human emotion detection." *Computational intelligence and neuroscience* 2022.1 (2022): 8032673.
11. Kumar, Akhilesh, and Awadhesh Kumar. "Human emotion recognition using Machine learning techniques based on the physiological signal." Biomedical Signal Processing and Control 100 (2025): 107039.
12. Alslaity, Alaa, and Rita Orji. "Machine learning techniques for emotion detection and sentiment analysis: current state, challenges, and future directions." *Behaviour & Information Technology* 43.1 (2024): 139-164.
13. Selvan, M. Arul. "Deep Learning Techniques for Comprehensive Emotion Recognition and Behavioral Regulation." (2024).
14. Zhang, Shiqing, et al. "Deep learning-based multimodal emotion recognition from audio, visual, and text modalities: A systematic review of recent advancements and future prospects." *Expert Systems with Applications* 237 (2024): 121692.
15. Geetha, A. V., et al. "Multimodal Emotion Recognition with deep learning: advancements, challenges, and future directions." *Information Fusion* 105 (2024): 102218.