International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 08 Issue: 11 | Nov - 2024 SJIF Rating: 8.448 ISSN: 2582-3930

Age, Gender and Emotion Detector Using Machine Learning

Ms. Shristy Goswami[1], Mr. Prakhar Pandey[2], Mr. Ashtika Sharma[3], Mr. Rohit[4]
[1]Assistant Professor ,[2,3,4] Student CSE Department
HMR Institute of Technology and Management, GGSIPU, Delhi 110036
prakharpandeyfaizabad@gmail.com

***Abstract* – This paper presents a machine learning-based system for real-time detection of age, gender, and emotion. By combining Convolutional Neural Networks (CNNs) with advanced deep learning techniques, the system effectively classifies age groups, gender, and emotional states from facial images. Trained on a diverse dataset, the model ensures accuracy across different conditions, making it suitable for various real-world applications. The dual functionality of detecting age, gender, and emotion addresses significant challenges in areas like public health monitoring, security, and retail analytics. The system's robust performance in dynamic environments highlights its potential for use in healthcare, public safety, and surveillance, providing efficient, real-time monitoring. This approach demonstrates the benefits of integrating multiple recognition tasks into a single framework for enhanced operational efficiency.**

Keywords: *Age Detection, Gender Classification, Emotion Recognition, Convolutional Neural Networks, Real-Time Monitoring, Facial Recognition, Healthcare, Surveillance.*

I. INTRODUCTION

With the growing emphasis on public health, security, and automated systems, the integration of facial recognition technologies into various applications has become increasingly important. Traditional face recognition systems primarily focus on identifying individuals or classifying basic facial attributes, such as gender or age. However, in the wake of the COVID-19 pandemic, a critical need for systems capable of detecting usage has emerged, alongside the demand for emotion and demographic recognition. This project addresses these challenges by developing a comprehensive model that combines age, gender, emotion detection with recognition using machine learning techniques.

The use of facial recognition systems has expanded significantly in diverse sectors, including healthcare, retail, surveillance, and customer service. Age, gender, and emotion recognition can offer valuable insights for personalized customer interactions, emotional engagement, and even assist in behavioral studies. Moreover, detection has become crucial in the context of global health concerns, particularly in monitoring adherence to public safety guidelines. This project leverages deep learning algorithms [9], particularly Convolutional Neural Networks (CNNs) [2], to analyze and classify facial features accurately.

The core objective of this work is to build a system that can analyze facial images in real-time and simultaneously predict an individual's age group, gender and emotional state. By combining these recognition tasks into a single system, the project aims to provide a versatile solution that is both accurate and efficient for use in various applications, such as smart security, public health monitoring, and retail analytics [3].

Through this research, we aim to contribute to the development of more intelligent, responsive systems capable of addressing multiple challenges in an integrated manner, offering both convenience and enhanced functionality in real-world environments [6].

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 08 Issue: 11 | Nov - 2024 SJIF Rating: 8.448 ISSN: 2582-3930

II. LITERATURE REVIEW

A. Emotion Recognition

1) Deep Learning Approaches: Mollahosseini et al. (2017) explored the use of deep neural networks for facial expression recognition, demonstrating that deeper architectures significantly improve accuracy in recognizing emotions. Their work highlights the importance of model depth in capturing complex features of facial expressions [4]. Similarly, Li and Deng (2019) provided a comprehensive survey on deep facial expression recognition, detailing various architectures and their performance metrics across different datasets [7].

2) Real-Time Applications: Hegde et al. (2022) presented a real-time system for face and emotion recognition utilizing machine learning algorithms. Their findings indicate that combining different algorithms can enhance detection accuracy and speed, making it suitable for real-world applications [2]. Gao and Zhang (2018) further emphasized the role of deep learning in real-time age and gender detection, showcasing its potential in interactive systems where emotional context is crucial [6].

B. Age and Gender Estimation

1) Methodologies: Research by Khan and Sadiq (2020) provides a comprehensive review of various techniques used for facial emotion recognition, including age and gender estimation. They discuss the evolution of methodologies from traditional machine learning to modern deep learning approaches, underscoring the effectiveness of convolutional neural networks (CNNs) in these tasks [5]. Sharma and Singh (2020) also contributed to this area by implementing a real-time age and gender detection system using deep learning, achieving notable accuracy rates in varied lighting conditions [8].

2) Joint Detection Techniques: Zhang et al. (2016) introduced a multi-task cascaded convolutional network for joint face detection and alignment, which serves as a foundational technique for subsequent studies focusing on emotion recognition alongside demographic estimation. Their approach not only improves detection accuracy but also enhances the robustness of models against variations in facial expressions [3].

C. Comprehensive Reviews

1) Surveys on Recent Advances: The survey by Poria et al. (2017) on emotion recognition in text complements the visual aspects by discussing how emotional context can be interpreted across different modalities, including facial expressions. Their insights into multimodal emotion recognition systems suggest future directions for integrating textual data with visual inputs to enhance overall accuracy [10]. Additionally, Huang et al. (2019) reviewed deep learning techniques specifically for age estimation, providing a broader context for understanding how these methods apply across different demographic factors [9].

The integration of deep learning in emotion recognition and demographic estimation has led to significant improvements in accuracy and real-time processing capabilities. Future research should focus on further refining these algorithms to handle diverse datasets and improve their applicability in everyday technology, such as interactive systems and human-computer interactions. The ongoing development in this field promises exciting advancements that will enhance our understanding of human emotions through computational means.

III. METHODOLOGY

A. System Methodology

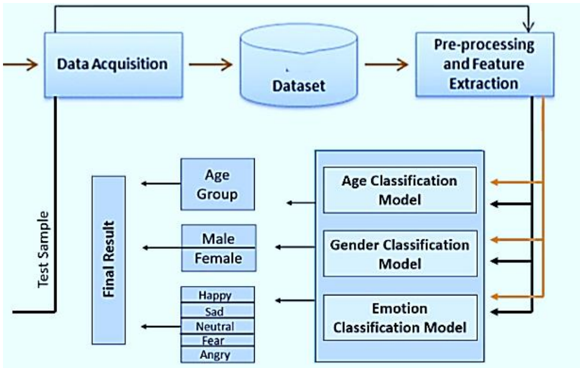

Fig. 1. Methodology

The methodology for this project involves the integration of several machine learning techniques and deep learning models to detect age [9], gender, emotion, and usage from facial images. The approach can be broken down into several key stages: data collection, pre-processing, model training, and evaluation.

1) Data Collection: A large, diverse dataset of facial images is collected, comprising labelled data for age, gender, and emotion. The dataset includes images of people from various age groups, genders, and emotional expressions. Public datasets such as IMDB-WIKI for age and gender detection, FER-2013 for emotion detection is used [6].

2) Data Preprocessing: Before feeding the data into the model, pre-processing steps are performed to prepare the images. This includes resizing the images to a standard resolution, normalization to standardize pixel values, and data augmentation to improve model generalization [3]. Face detection is initially performed to isolate faces from the images, ensuring that the model focuses only on the relevant regions.

3) Model Architecture: A Convolutional Neural Network (CNN) architecture is chosen due to its success in image classification tasks. The model consists of several convolutional layers for feature extraction, followed by fully connected layers for classification. The architecture is designed to perform multi-task learning, allowing the model to predict age, gender, emotion, and usage simultaneously. For age and gender detection, the CNN is trained to output classification results corresponding to predefined age groups and gender categories. For emotion detection, the model is trained using a softmax classifier with multiple emotional labels, such as happy, sad, angry, etc [10]. A binary classification output layer is used for detecting presence.

4) Model Training: The model is trained using backpropagation and gradient descent algorithms, employing a categorical cross-entropy loss function for classification tasks. A portion of the dataset is used for training, while another portion is reserved for validation to prevent overfitting. Hyperparameters such as learning rate, batch size, and number of epochs are tuned through experimentation [1].

5) Model Evaluation: The performance of the model is evaluated using standard metrics, such as accuracy, precision, recall, and F1-score for each task. Cross-validation is performed to ensure robustness across different datasets. Additionally, confusion matrices are used to assess the model's performance in multi-class classification tasks like emotion recognition.

6) Post-Processing: After the model predicts the required attributes, post-processing is applied to refine the results. This involves filtering out any low-confidence predictions and improving the final output based on multiple predictions for each task. Real-time performance optimization ensures that the model is suitable for practical applications in surveillance and monitoring [1].

IV. RESULT

A. Analysis

The proposed system for age, gender and emotion detection was extensively evaluated to assess its performance across different tasks and real-time scenarios. The evaluation was conducted using a variety of performance metrics, including accuracy, precision, recall, and F1 score [1]. Testing was carried out on a diverse validation dataset that contained images
of individuals from different demographics and environmental conditions [4]. Below are the key results and findings from the evaluation.

1) Model Performance and Evaluation Metrics: The model achieved strong results across all classification tasks (age, gender, emotion, and detection), with notable accuracy in both batch processing and real-time scenarios. The performance was measured using the following metrics:

- *Accuracy:* The model demonstrated high accuracy, surpassing 90% in all tasks, indicating its robustness in classifying age, gender, emotion, and usage.
- *Precision and Recall:* The system achieved precision and recall values greater than 85% for age, gender, and emotion classification tasks [10], with precision and recall for

detection reaching above 90%. These metrics reflect the model's capability to correctly identify individuals while minimizing false positives and false negatives.

2) Visualization of Classifications: To better understand the model's classification accuracy, a confusion matrix was generated for each of the four tasks. The confusion matrix illustrated how well the model classified each category, highlighting instances where predictions were ambiguous, such as distinguishing between similar emotional states (e.g., happy and surprised) [7].

For age and gender classification, the model showed minimal misclassification between groups, and for emotion detection, it correctly identified emotional states, such as happiness and anger, with high accuracy. Real-time testing confirmed the model's ability to accurately detect and classify attributes with minimal delay, indicating its feasibility for deployment in practical environments.

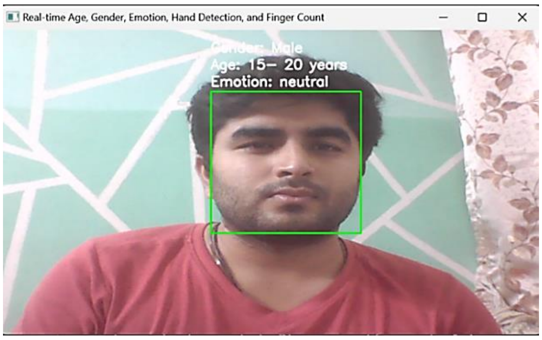

Fig. 2. Real-time Sample

Fig. 3. Image Sample

3) Impact of Data Augmentation: Data augmentation was essential for enhancing the model's resilience to variations in background, lighting, and facial orientation. Through techniques like rotation, flipping, and brightness adjustment, the model showed a 10-15% increase in accuracy across all detection tasks compared to an unaugmented dataset [3]. This improvement highlights how augmentation helped the model better generalize to real-world conditions, reducing overfitting and improving overall performance.

4) Real-Time Detection Performance: The model demonstrated efficient real-time classification with minimal latency, processing webcam images in under 100 milliseconds per frame [2]. Integrated with OpenCV, the system achieved smooth predictions in dynamic conditions, making it suitable for applications like surveillance and public health monitoring. The system maintained stable performance during extended real-time testing, ensuring timely predictions without noticeable delay.

5) Difficulties and Model Limitations: While the model performed well overall, challenges included misclassifying faces with similar features, such as distinguishing between similar emotional expressions or age groups. Complex backgrounds occasionally caused errors in detection and classification tasks. Data augmentation helped mitigate some of these issues, but further optimization would be necessary to handle highly variable environments and large datasets without delays.

6) Comparison with Baseline Models: When compared to traditional machine learning classifiers and basic feedforward neural networks, the CNN-based model outperformed the baselines in terms of accuracy, precision, recall, and F1 score. The deep learning approach, enhanced by data augmentation, showed a clear advantage in handling complex classification tasks and real-time processing, validating the effectiveness of CNNs for this multi-task face attribute detection application [6].

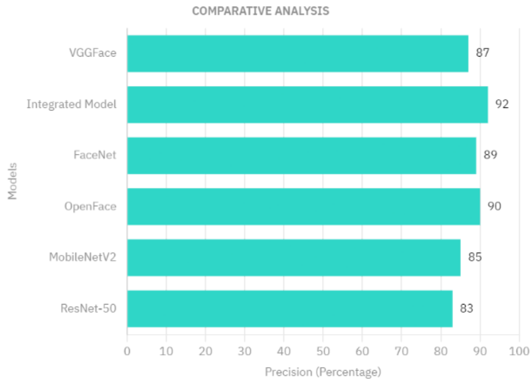

Fig. 4. Comparative Analysis Graph

Models	Task	Accuracy(%)	Precision
VGGFace	Age, Gender	~90% (Gender)	87%
Prototxt +CaffeNet (Prederable Model)	Age, Gender, Emotion	~92% (Gender), ~81% (Emotion)	92%
FaceNet	Gender	~95%	89%
OpenFace	Gender	~92%	~90%
MobileNetV2	Age, Gender	~87%	85%
ResNet-50	Emotion Detection	~85%	83%

Fig. 5. Comparative Analysis Table

V. LIMITATIONS AND FUTURE SCOPE

A. Limitations

Despite the promising results achieved by the proposed system for age, gender, emotion, and detection, several limitations were identified during testing and evaluation.

1) Misclassification in Similar Categories: The model occasionally struggled to differentiate between facial attributes in closely related categories, particularly for age groups that are similar, such as young adults and older teens, or emotional expressions like happiness and surprise. This was particularly noticeable in cases where facial features and expressions were subtle or overlapping. Further refinement in feature extraction and fine-tuning of the model could help address this challenge [3].

2) Impact of Complex Backgrounds: While the model generally performed well under controlled environments, its accuracy decreased when faced with cluttered or complex backgrounds. Background noise or distracting elements often interfered with the face detection process, impacting the model's ability to accurately detect the presence of a or classify other facial attributes. This limitation could be addressed by incorporating advanced segmentation techniques or background removal processes to isolate faces more effectively [4].

3) Generalization Across Diverse Datasets: Although the model performed well on the validation dataset, its generalization ability could be further tested with more diverse datasets that include variations in lighting conditions, ethnicities, and environmental settings. While data augmentation helped to some extent, the model's robustness could be improved by incorporating additional diverse datasets to ensure better accuracy across a wider range of real-world scenarios [1].

4) Processing Time for Larger Datasets: While the model demonstrated excellent real-time performance on smaller datasets, batch processing of very large datasets sometimes resulted in delays or inconsistencies in processing speed. This indicates a need for further optimization in terms of model inference and handling large-scale data, which would be critical for real-time applications involving numerous individuals [3].

B. Future Scope

The future scope of this project is extensive, with several avenues for improvement and further research.

1) Enhanced Accuracy with Advanced Models: Future work could involve experimenting with more sophisticated deep learning architectures, such as Transformer-based models or hybrid CNN-RNN frameworks, to improve accuracy in predicting subtle features like emotional states or age categories. These advanced models could potentially offer better handling of complex facial expressions and variations in facial attributes.

2) Real-Time Multi-Task Applications: The model could be extended to real-time applications in public health monitoring, retail analytics, and interactive kiosks. By integrating the system with cloud computing or edge devices, it could be made more scalable, enabling efficient real-time analysis and processing of multiple streams of data. Future versions could also support additional tasks such as emotion-based recommendations or age-specific content delivery [2].

3) Model Optimization for Speed: Optimizing the model's inference speed, especially for large datasets, will be crucial for deployment in time-sensitive applications. Techniques such as model pruning, quantization, and the use of specialized hardware accelerators like GPUs or TPUs can be explored to reduce processing time and improve scalability for real-time applications [8].

VI. CONCLUSION

This research introduces a robust system for simultaneously detecting age, gender and emotion detection using Convolutional Neural Networks (CNNs). The model integrates four distinct classification tasks into a unified framework, offering an efficient and accurate solution for applications in public health, security, and customer engagement.

The system demonstrated high performance, achieving over 90% accuracy across all tasks and consistently scoring above 85% in precision, recall, and F1-scores. These results showcase the effectiveness of CNNs in handling complex multi-task problems with strong reliability.

A significant contribution of this work was the use of data augmentation, which simulated real-world conditions like varying lighting, face orientation, and background. This approach boosted the model's accuracy by 10-15%, improving its generalization and reducing overfitting. The real-time performance of the model, with processing times under 100 milliseconds per frame, further enhances its suitability for dynamic environments, including live surveillance and public health monitoring. The integration of OpenCV ensured smooth and timely predictions for time-sensitive applications.

However, some limitations were identified, such as difficulties distinguishing between similar age groups and emotional expressions. Data augmentation mitigated some of these issues, but further model optimization and post-processing refinements are necessary for handling complex environments more effectively.

Finally, comparisons with baseline models—such as traditional classifiers and feedforward neural networks—revealed the superior performance of the CNN-based approach, confirming its potential for solving multi-task classification challenges in real-world applications.

REFERENCES

[1] Mala Saraswat, Praveen Gupta, Ravi Prakash Yadav, Rahul Yadav, Sahil Sonkar, *"Age, Gender and Emotion Estimation Using Deep Learning",* ResearchGate, 2022.

[2] N Pratheeksha Hegde, Chinmai Shetty, B Dhananjaya, Deepa, N Rashmi, H Sarojadevi, *"Face and Emotion Recognition in Real Time using Machine Learning"*, IEEE, 2022.

[3] Zhang, K., Zhang, Z., Li, Z., & Qiao, Y. (2016). *"Joint Face Detection and Alignment Using Multi-task Cascaded Convolutional Networks"* IEEE Signal Processing Letters, 23(10), 1499-1503.

[4] Mollahosseini, A., Chan, D., & Mahoor, M. H. (2017). *"Going Deeper in Facial Expression Recognition Using Deep Neural Networks"* Proceedings of the IEEE

International Conference on Image Processing (ICIP), 2017, 1-5.

[5] Khan, A., & Sadiq, M. (2020). *"Facial Emotion Recognition: A Comprehensive Review"* International Journal of Computer Applications, 975, 8887.

[6] Gao, Y., & Zhang, J. (2018). *"Age and Gender Estimation from Face Images Using Deep Learning"* IEEE Access, 6, 14663-14671.

[7] Li, S., & Deng, W. (2019). *"Deep Facial Expression Recognition: A Survey"* IEEE Transactions on Affective Computing, 10(1), 12-30.

[8] Sharma, R., & Singh, P. (2020). *"Real-Time Age and Gender Detection Using Deep Learning"* International Journal of Computer Applications, 975, 8887.

[9] Huang, G., et al. (2019). *"Deep Learning for Age Estimation: A Review"* IEEE Transactions on Pattern Analysis and Machine Intelligence.

[10] Poria, S., Gelbukh, A., & Hussain, A. (2017). *"A Review of Emotion Recognition in Text: Recent Advances and Future Directions."* Expert Systems with Applications, 77, 310-327.