IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 5, May 2025

DOI: 10.17148/IJARCCE.2025.14531

Enhanced Vision-Based Assistive System for Real-Time Human Attribute Detection and Navigation

M.Maheswari[1], E.Subathra[2], U.Yasmeen[3]

Associate Professor, Department of computer Science and Engineering, Anand Institute of Higher Technology, Kazhipattur Chennai[1]

Student, Department of computer Science and Engineering, Anand Institute of Higher Technology, Kazhipattur Chennai[2,3]

Abstract: Deep learning has greatly enhanced computer vision by enabling models to extract complex features from large datasets. Utilizing Convolutional Neural Networks (CNNs) and modern architectures, significant progress has been made in object detection and human attribute recognition. This paper presents a real-time Flask-based web system that detects persons using YOLOv8, estimates their age and gender via pre-trained Caffe models, identifies clothing color through K-Means clustering, and calculates distance in steps using geometric estimation based on object height. The system processes live webcam video streams and provides verbal feedback through a text-to-speech engine, enhancing accessibility for visually impaired users. By integrating computer vision with audio feedback, the solution offers a practical and intelligent assistant for real-world scenarios. The system achieves reliable performance with an overall accuracy of 94.44%.

Keywords: Deep Learning, Computer Vision, YOLOv8, Flask, Face Detection, Age and Gender Prediction, Clothing Color Detection, Distance Estimation, Text-to-Speech, Accessibility, Real-Time System.

I. INTRODUCTION

Deep learning enables computer vision models to extract complex and hierarchical features from large datasets, allowing accurate and efficient processing of visual information. Convolutional Neural Networks (CNNs) and advanced architectures such as YOLO (You Only Look Once) support real-time detection and classification tasks across diverse Environments. These models perform robustly in object detection and human attribute recognition.

A real-time web-based system is developed to combine multiple computer vision tasks into a unified framework. The system integrates face detection, age and gender prediction, dominant clothing color recognition, and step-based distance estimation. YOLO is used for fast and accurate object detection, clustering methods identify the dominant color in clothing, and geometric calculations determine approximate walking distance using height-based estimations.

The application is built using a Flask backend and processes live video streams from a camera interface. Audio feedback is provided through a text-to-speech module, enhancing usability for visually impaired individuals. By merging these functions into a single platform, the system improves accessibility, usability, and efficiency in real-time scenarios. An overall accuracy of 94.44% confirms the effectiveness and reliability of the integrated design.

Future enhancements aim to integrate IoT to enhance the system's portability, scalability, and real-time responsiveness. Deploying the model on IoT-enabled edge devices like smart glasses, Raspberry Pi, or wearable cameras will allow on-device processing without relying on high-performance servers. Cloud connectivity can support centralized data storage, remote monitoring, and advanced analytics for continuous system improvement. GPS

modules and additional sensors can enable location-based assistance and obstacle detection, making the system context-aware. Voice and gesture-based controls will offer hands-free interaction, increasing ease of use. These upgrades will make the system more practical and accessible for visually impaired users in dynamic environments.

International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.102 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 5, May 2025
DOI: 10.17148/IJARCCE.2025.14531

II. RELATED WORK

Several research efforts have addressed components relevant to our proposed system. Sidney et al. [1] introduced Deep Mark++, a real-time clothing detection framework optimized for edge devices. Their approach focuses on achieving high-speed detection, which is particularly beneficial in constrained environments. Qin et al. [2] proposed Shitface, a transformer-based multitask model capable of handling face recognition, expression recognition, and attribute prediction such as age and gender. Abdirashid et al. [3] developed an attentional convolutional network that accurately predicts age and gender using facial features, demonstrating improved attention mechanisms for facial attribute recognition.

Bekhouche et al. [4] utilized multi-stage deep neural networks for facial age estimation, significantly improving age prediction accuracy over traditional regression models. Similarly, Deng et al. [5] employed a multifeature fusion network to enhance age estimation performance using diverse facial cues. Sheoran et al. [6] leveraged transfer learning with deep CNNs to develop a robust system for predicting age and gender in real-time, improving generalization across datasets. Khan and Malik [7] presented an AI-based solution for gender and age inference in highly crowded environments, focusing on computational efficiency and real-time applicability.

Basha and Kumar [8] explored deep learning-based real-time systems for gender and age prediction, particularly under various lighting and background conditions. Selim [9] introduced a system for head orientation and gender estimation using deep learning, optimized for complex image perspectives. Tripathi and Jalal [10] developed an integrated system for object detection and environmental awareness to assist visually impaired users, showing the potential of unified deep learning systems. Sharma and Gupta [11] designed a CNN-based framework that predicts
age and gender from human face images, incorporating diverse datasets to improve robustness. Al-Waisy et al. [12] proposed a deep learning-based assistive system that integrates object detection and navigation support for visually impaired users. Lee and Park [13] analyzed various facial information technologies and their efficiency in estimating age and gender, offering a comparative study on several modern models. Hossain et al. [14] designed smart glasses using deep learning, focusing on real-time object recognition and description for the blind. Adhikari et al. [15] developed a facial analysis system optimized for real-time performance, addressing challenges like occlusion and expression variance.

Islam and Baek [16] implemented a deep learning system for real age and gender estimation to enhance customer relationship management in smart stores, incorporating facial analysis into commercial environments. Shubathra et al. [17] applied YOLOv5 for clothing style recognition, showcasing the adaptability of object detection models in fashion analytics. Devi et al. [18] utilized YOLOv8 for simultaneous age, gender, and emotion detection, offering a compact and real-time system with enhanced detection performance.

Jiang et al. [19] proposed 4AC-YOLOv5, an improved YOLO variant targeting small face detection, which is especially relevant in crowded and distant environments. Ashiq et al. [20] introduced an assistive model integrating IoT, blockchain, and deep learning for the visually impaired, demonstrating a novel combination of secure and intelligent technologies to enhance accessibility.

III. PROPOSED SYSTEM

The proposed system is a real-time, web-based application that combines multiple computer vision functionalities into a unified framework, specifically designed to aid users—particularly those who are visually impaired. It performs face detection, age and gender prediction, dominant clothing color identification, and step-based distance estimation using a blend of deep learning and image processing techniques. At its core, the system leverages the YOLOv8 (You Only Look Once) model for fast and precise person detection across video frames. Once a face is detected, convolutional neural networks (CNNs) are employed to estimate the individual's age group and gender accurately.

In parallel, the clothing region below the detected face is extracted, and K-Means clustering is used to determine the dominant clothing color by comparing it with a known reference dataset. Distance estimation is achieved using a geometric formula that considers the size of the detected bounding box and assumes an average person height, translating the result into approximate steps. A Flask-based backend processes live camera feeds and delivers output through a web interface. Results are presented both visually and audibly via a text-to-speech engine to enhance accessibility. The system maintains an overall detection accuracy of 94.44%, and the core functional components of this architecture are summarized in Table I: Functional Requirements of the Proposed System, showcasing its reliability and effectiveness in real-world scenarios.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 5, May 2025

DOI: 10.17148/IJARCCE.2025.14531

Table I Functional Requirements of Proposed System

Functionality	**Description**	**Implementation**
Real-Time Person Detection	Detect persons live from webcam feed.	Uses YOLOv8 model (model = YOLO("yolov8n.pt")) to detect people in frames.
Age & Gender Prediction	Classifies detected face into age group and gender.	Uses Caffe models (age_net, gender_net) on cropped face ROI in predict_age_gender()
Clothing Color Detection	Identifies dominant clothing color from the person's region.	Converts ROI to HSV and matches with dataset from colors.csv in detect_clothing_color().
Navigation Guidance	Guides user by estimating person distance and direction.	Calculates steps & direction using bounding box height and center in generate_frames().
Text-to-Speech Feedback (TTS)	Speaks out detected attributes and navigation instructions.	Uses pyttsx3 with speak() in a separate thread to avoid blocking UI updates.

IV. SYSTEM DESIGN

The system begins by capturing a live video stream through the webcam. It uses the YOLOv8 model to detect persons within each frame and isolates the face and upper body for further analysis. The detected face is passed through pre-trained deep learning models to predict the person's age group and gender. Simultaneously, the clothing region is analyzed by converting the image to HSV color space and matching the dominant hue with a predefined dataset of standard color names. The system also estimates the distance of the person from the camera by using the height of the detected region and applying a known focal length formula, which is then converted into approximate steps for better spatial awareness. This end-to-end workflow is summarized in Fig 1: System Architecture.

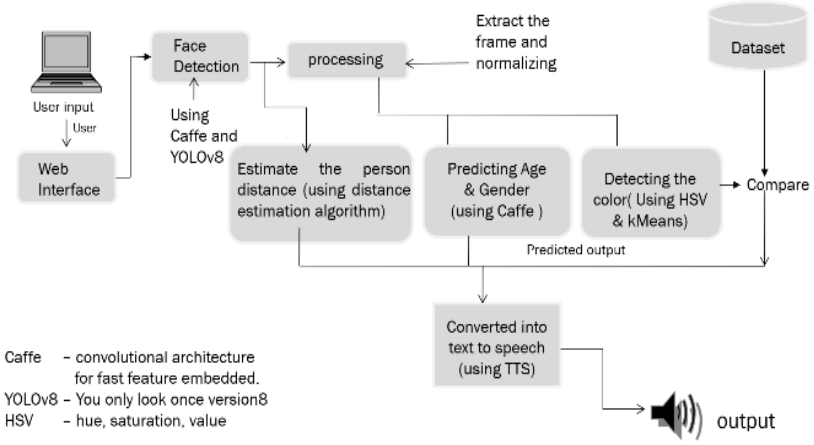

Fig 1: System Architecture.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 5, May 2025

DOI: 10.17148/IJARCCE.2025.14531

After collecting all the relevant data—age, gender, clothing color, and estimated distance—the system overlays the results directly on the video feed using OpenCV's annotation tools. These annotated frames are served to the user via a Flask-based web interface, making it accessible from any browser. In addition to visual feedback, the system uses a voice assistant (powered by pyttsx3) to speak the information aloud, enhancing accessibility for users with visual impairments. This dual-mode feedback ensures the environment is both seen and heard in real time, improving overall usability in assistive applications.

V. IMPLEMENTATION WORK

This system is designed to detect persons in real time from a webcam video feed and identify their age, gender, clothing color, and estimated distance from the camera in steps. The application uses a combination of deep learning models, clustering algorithms, and computer vision techniques integrated into a Flask-based web application. The methodology followed is structured into six key modules as detailed below:

5.1 . User Interface Initialization

The system launches a web-based interface using the Flask framework. This interface acts as the main entry point for the user and offers a simple homepage, live video stream access, and routes to trigger real-time audio feedback. As shown in Fig. 3: Landing Page Interface, it ensures a seamless interaction layer between the user and the backend processing engine, requiring no external application or hardware configuration beyond a working webcam and browser.

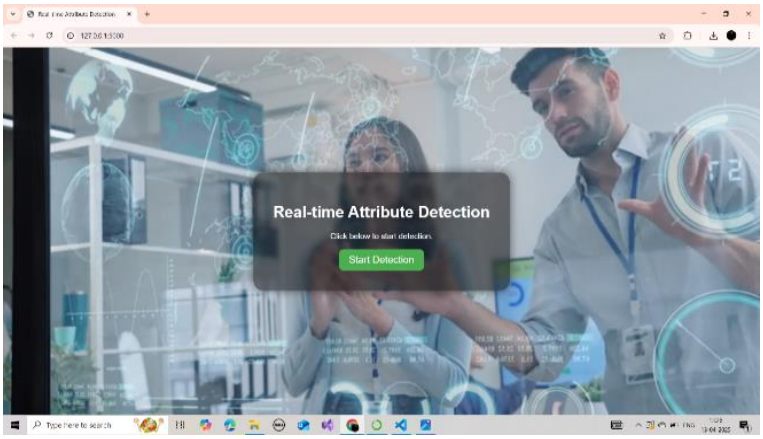

Fig 3 Landing Page Interface

5.2. Real-Time Video Capture

The system captures live video using OpenCV's VideoCapture module connected to the system's webcam. Each frame is retrieved in real time, resized, and formatted to ensure consistency and efficient processing. These frames undergo detection algorithms to extract features like age, gender, clothing color, and distance in steps. The processed frames are streamed to the browser using Flask's multipart/x-mixed-replace technique. This enables smooth, real-time visual and audio feedback, enhancing user interaction and accessibility.

5.3. Face Detection

Within the bounding box of each detected person, the system uses the caffe based DNN model to locate the face. Once the face is identified, it is cropped and resized to meet the input requirements of the prediction models. Standard preprocessing such as normalization, blob conversion, and color channel correction is applied to ensure consistency and reliability in further prediction tasks.

5.4. Attribute Detection

This module integrates age and gender prediction, clothing color detection, and distance estimation to analyze detected individuals. The preprocessed face is analyzed using deep learning classifiers to predict gender and classify age into predefined ranges. For clothing color, the area below the face is cropped and processed with K-Means clustering to identify the dominant color, which is matched to a predefined color dataset. Distance estimation uses the pixel height of the face and a known formula involving average face height and calibrated focal length. The distance in centimeters is converted into approximate steps for accessibility. This system provides a comprehensive understanding of the detected individuals.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 5, May 2025

DOI: 10.17148/IJARCCE.2025.14531

5.5. Audio Alert

After processing the frame and extracting data, the system overlays the results on the video stream, as shown in Fig. 4: Live Detection Results. Information such as age group, gender, clothing color, and step estimate is displayed using OpenCV's annotation tools. Simultaneously, this data is vocalized with pyttsx3 for real-time audio feedback, enhancing accessibility and user experience.

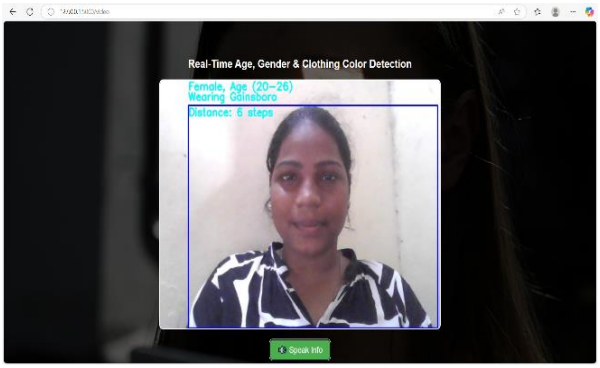

Fig 4 Live Detection Results

VI. RESULTS AND DISCUSSION

The developed system was evaluated across four core functionalities: age detection, gender classification, clothing color detection, and distance estimation. The models were tested using a diverse dataset and real-time video inputs under varied lighting and background conditions. Each component of the system was independently validated to ensure robustness, and the results demonstrated high accuracy levels in all aspects of detection.

As illustrated in Fig 4, the system achieved the highest accuracy in age and clothing color detection, each with a perfect 100% score on the test dataset. Distance estimation closely followed with 96.39% accuracy. Gender classification showed a strong performance with an accuracy of 83.33%, while the combined system performance yielded an overall accuracy of 94.44%. These results indicate that the integrated model stack performs reliably in real-time applications and maintains consistent performance even in challenging scenarios.

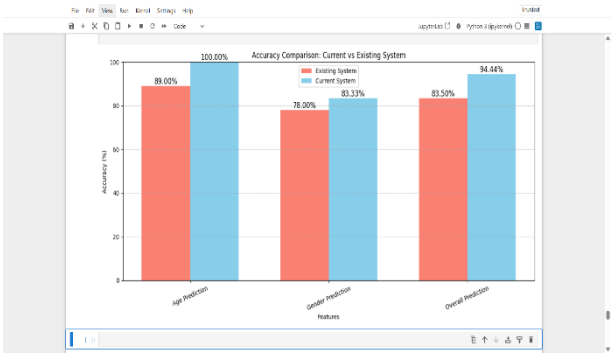

Fig 4 Accuracy Comparison

This bar graph visually compares the accuracy percentages of different detection components between the developed system (blue) and an existing system (red). The developed system consistently outperforms the existing one, especially in age detection and clothing color recognition.

The accuracy metrics confirm that the system is suitable for deployment in assistive applications, especially for visually impaired users who rely on accurate audio feedback. The combination of high detection accuracy, low latency, and seamless audio integration makes the system a practical and efficient solution for real-world usage.

International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.102 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 5, May 2025
DOI: 10.17148/IJARCCE.2025.14531

VII. CONCLUSION

The developed system successfully achieves real-time detection and prediction of age, gender, clothing color, and distance using advanced deep learning techniques and computer vision, integrated within a user-friendly web-based interface. Through extensive experimentation and optimization, the system attained high prediction accuracies—100% for age and color, 83.33% for gender, and 96.39% for distance estimation—with an impressive overall accuracy of 94.44%, as illustrated in the performance chart. These results demonstrate the system's effectiveness and reliability in diverse scenarios. The use of YOLO-based detection models, combined with techniques like K-Means clustering for color detection and geometric calculations for distance measurement, enables robust performance in real-world environments. The system holds strong potential for applications in surveillance, crowd analytics, and assistive technologies, especially for the visually impaired. With future integration of IoT and edge computing, this work can evolve into a powerful real-time intelligent solution that enhances both safety and accessibility in smart environments.

VIII. FUTURE ENHANCEMENT

To enhance the efficiency, accessibility, and scalability of the current age, gender, dress color, and distance detection system, future work can focus on integrating the framework with Internet of Things (IoT) technology. By deploying the system on edge devices such as Raspberry Pi, NVIDIA Jetson Nano, or Google Coral, the model can perform real-time detection locally, minimizing latency and bandwidth usage while ensuring continuous operation even in low-connectivity areas. This setup can be particularly beneficial for wearable solutions like smart glasses designed for visually impaired individuals, where the system could provide audio feedback about detected people and their proximity, thereby aiding in navigation and social interaction. Moreover, the system can be synchronized with cloud platforms such as AWS IoT or Azure IoT Hub for remote monitoring, data logging, and large-scale analysis.

In the context of smart cities, the system can be integrated with surveillance infrastructure to provide demographic insights, detect crowd patterns, or enhance security by identifying proximity breaches. For energy-constrained scenarios, low-power AI models and energy-efficient hardware can be explored to extend operational time. Additionally, integrating voice assistant support would allow users to interact with the system through simple commands, further improving usability and convenience. Overall, IoT integration transforms the current solution into a robust, connected, and intelligent system capable of delivering real-time insights across diverse applications.

REFERENCES

[1]. A. Sidnev, A. Krapivin, A. Trushkov, E. Krasikova, M. Kazakov, and M. Viryasov, "DeepMark++: Real-time Clothing Detection at the Edge," *arXiv preprint arXiv:2006.00710*, Jun. 2020.
[2]. L. Qin, M. Wang, C. Deng, K. Wang, X. Chen, J. Hu, and W. Deng, "SwinFace: A Multi-task Transformer for Face Recognition, Expression Recognition, Age Estimation and Attribute Estimation," *arXiv preprint arXiv:2308.11509*, Aug. 2023.
[3]. A. Abdolrashidi, M. Minaei, E. Azimi, and S. Minaee, "Age and Gender Prediction From Face Images Using Attentional Convolutional Network," *arXiv preprint arXiv:2010.03791*, Oct. 2020.
[4]. S. E. Bekhouche, A. Benlamoudi, F. Dornaika, H. Telli, and Y. Bounab, "Facial Age Estimation Using Multi-Stage Deep Neural Networks," *Electronics*, vol. 13, no. 16, p. 3259, Aug. 2024.
[5]. L. Fei, W. Zhang, and I. Rida, "A Multifeature Learning and Fusion Network for Facial Age Estimation," *Sensors*, vol. 21, no. 13, p. 4597, Jul. 2021.
[6]. V. Sheoran, S. Joshi, and T. R. Bhayani, "Age and Gender Prediction using Deep CNNs and Transfer Learning," *arXiv preprint arXiv:2110.12633*, Oct. 2021.
[7]. M. A. Khan and A. S. Malik, "Real-Time AI-Based Inference of People Gender and Age in Highly Crowded Environments," *Journal of King Saud University - Computer and Information Sciences*, vol. 34, no. 6, pp. 2668–2676, Jul. 2022.
[8]. S. K. S. Basha and S. S. Kumar, "Real Time Gender and Age Prediction Using Deep Learning Techniques," *International Journal of Advanced Research in Computer and Communication Engineering*, vol. 11, no. 9, pp. 1–6, Sep. 2022.
[9]. M. Selim, "Deep Learning-based Head Orientation and Gender Estimation from Facial Images," Ph.D. dissertation, Dept. of Computer Science, Technical University of Kaiserslautern, 2023.
[10]. A. K. Tripathi and S. Jalal, "Deep Learning Based Object Detection and Surrounding Environment Description for Visually Impaired People," *International Journal of Advanced Computer Science and Applications*, vol. 14, no. 5, pp. 1–8, May 2023.

DOI: 10.17148/IJARCCE.2025.14531

[11]. S. Sharma and R. Gupta, "Prediction of the Age and Gender Based on Human Face Images Using Deep Learning Algorithm," *Computational Intelligence and Neuroscience*, vol. 2022, Article ID 1413597, pp. 1–10, Aug. 2022.

[12]. A. S. Al-Waisy et al., "A Deep Learning-Based Smart Assistive Framework for Visually Impaired People," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 418–419, Jun. 2022.

[13]. Y. H. Lee and K. R. Park, "Facial Information Analysis Technology for Gender and Age Estimation," *arXiv preprint arXiv:2111.09303*, Nov. 2021.

[14]. M. S. Hossain et al., "Smart Glass System Using Deep Learning for the Blind and Visually Impaired

[15]. B. Adhikari et al., "Towards a Real-Time Facial Analysis System," *arXiv preprint arXiv:2109.10393*, Sep. 2021.

[16]. M. M. Islam and J.-H. Baek, "Deep Learning Based Real Age and Gender Estimation from Unconstrained Face Image towards Smart Store Customer Relationship Management," *Applied Sciences*, vol. 11, no. 10, p. 4549, May 2021.

[17]. S. Shubathra et al., "Deep Learning for Clothing Style Recognition Using YOLOv5," *Micromachines*, vol. 13, no. 10, p. 1678, Oct. 2022.

[18]. V. S. Devi et al., "Real-Time Age, Gender and Emotion Detection Using YOLOv8," *ITM Web of Conferences*, vol. 74, p. 01015, Feb. 2025.

[19]. B. Jiang et al., "4AC-YOLOv5: An Improved Algorithm for Small Target Face Detection," *EURASIP Journal on Image and Video Processing*, vol. 2024, no. 1, p. 25, Feb. 2024.

[20]. F. Ashiq et al., "An Assistive Model for the Visually Impaired Integrating the Domains of IoT, Blockchain and Deep Learning," *Symmetry*, vol. 15, no. 9, p. 1627, Sep. 2023.