Real-Time Gender Classification Using MiniXception and Hand Gesture Detection Using MediaPipe Framework

Manas Girish Kulkarni
Dept. of Electronics & Telecommunications
Vishwakarma Institute of Information Technology
Pune, India
manas.22311608@viit.ac.in

Revati Tushar Aute
Dept. of Electronics & Telecommunications
Vishwakarma Institute of Information Technology)
Pune, India
revati.22311517@viit.ac.in

Rajlakshmi Nilesh Desai
Dept. of Electronics & Telecommunications
Vishwakarma Institute of Information Technology
Pune, India
rajlakshmi.22311732@viit.ac.in

Atharva Vishwas Deshpande
Dept. of Electronics & Telecommunications
Vishwakarma Institute of Information Technology
Pune, India
atharva.22311679@viit.ac.in

Mrs. Dipti Pratik Pandit
Asst. Professor, Dept. of Electronics & Telecommunications
Vishwakarma Institute of Information Technology
Pune, India
dipti.pandit@viit.ac.in

***Abstract*— In human-computer interaction, gender classification and hand gesture recognition are essential technologies. This paper presents a real-time system that integrates gender prediction and face and hand gesture detection, utilizing the Mini Xception model and MediaPipe framework. We enhance the requirements for efficient real-time operation by utilizing frame-skipping mechanisms in our work. The system comprises video capture with OpenCV, model integration using TensorFlow, and efficient data handling with NumPy. Extensive testing has proven its generalization to diverse settings and excellent performance, with 84% accuracy for both tasks, capable of real-time processing.**

Keywords—Gender Detection, Mini Xception, Hand Gesture Detection, MediaPipe Library

I. Introduction

As machine learning and computer vision are continuously improving every day, in many fields of applications there is a need for systems capable of interpreting human characteristics like gender or hand gestures. These systems provide consumer technology and a control measure that allows us to achieve higher security surveillance, which leads to human-computer interaction to a whole new extent. The concept of gender classification and hand gesture recognition for users, which could extend wrist-based apps customizing from tailored services (which make new reuse areas), is the foundation in favour of movement-based automation.

Computer vision, a crucial field in artificial intelligence, allows software to analyze and comprehend visual information. OpenCV, a popular open-source library is an integral component in providing real-time image processing tools. OpenCV provides tools for C++, Python, and Java programming languages, used in many platforms/industries where actions like face detection or recognition must be performed. [1] highlights the importance of OpenCV in these applications, examines current advancements in the industry, and evaluates OpenCV's modules and algorithms. Advanced AI techniques for gender detection that integrate deep learning and computer vision are being researched for real-time applications. [2] combines the Haar Cascade Classifier from OpenCV for facial recognition with unique JavaScript features that allow for real-time video analysis. By demonstrating how accurately AI can identify gender, this system emphasizes the importance of ethical considerations in ensuring inclusivity and avoiding bias. The research explores how AI could serve to harmonize innovation and social responsibility, among other societal concerns revolving around the need for more inclusive technology representation. One of them is the problem that has been treated as automatic age classification, and we have to estimate people's ages from image collections in many technological contexts. [3] describes how CNNs, deep learning, and machine learning can be used as age assessment technologies. CNNs have proved effective in feature extraction from images and can apply age categorization from facial images with much ease. The work builds a CNN model for precise age estimation using the UTKFace dataset that is, a large collection of annotated facial images that reflect age, gender, and ethnicity. [6] describes the development of a device for the gender classification of individuals using a Convolution Neural Network on a Raspberry Pi. Key objectives are face detection, CNN training on different datasets of male and female images, and performance evaluation through a confusion matrix. Keras and TensorFlow are incorporated for image processing and classification, with a training accuracy of about 96 per cent and a validation accuracy of 90 per cent. The study attempts to extend the study of genders by discussing the challenges of gender identification in a machine-dependent world and illuminating various cultural and customary practices.

The paper is as structured: First, it involves a literature review in Section II, with particular attention to current developments in hand gesture recognition and gender classification. The strengths and weaknesses of popular and newly created gender detection methods are discussed in Section III. Section IV explains the proposed system of detection. Section V explains code implementation in algorithm format. Section VI discusses the performance and accuracy of the model based on our dataset, and Section VII discusses the result. Section VIII brings an end, to the discussion. The document concludes with a discussion, on possibilities, in Section IX.

II. Literature Review

The field of computer vision has experienced significant advancements in gender detection and gesture recognition, largely due to breakthroughs in machine learning and deep learning. This literature review examines the key methodologies and models that have influenced these technologies, highlights areas of existing knowledge gaps and outlines avenues for further research.

Ajay KUMAR et al. (2022) [1] implemented instant face detection and identification with trained models for detection and identification in OpenCV, using pre-trained models like Haar cascades for face detection and recognition using Local Binary Patterns, by balancing accuracy, and computational efficiency. Soumik Hore et al. (2023) [2] explored the use of AI for gender detection through CNNs and SVMs, highlighting that diverse datasets can reduce biases and enhance the accuracy of gender classification. Ishita VERMA et al. (2019) [3] studied age and gender prediction using CNN and obtained high accuracy with little preprocessing on a dataset of 20,000 photos, which is pertinent to applications such as security. Avoy Sain et al. (2023) [4] developed a facial recognition attendance system using OpenCV. This system improves accuracy by addressing false positives and lighting issues, making it suitable for educational and corporate environments. Candy Nneoma Esomonu (2023) [5] developed a hand gesture recognition system using MediaPipe for American Sign Language (ASL). This system effectively recognizes hand gestures for numbers. Steven DG. Boncolmo et al. (2021) [6] developed a model for identifying gender using Keras. Enhancing the accuracy of gender classification in convolutional neural networks (CNN) is crucial. Applications such as surveillance and social media often prioritize efficiency. Rajasekaran Thangaraj et al. (2023) [7] presented a system for real-time face detection and gender prediction using OpenCV with an Inception V3 model. By employing deep learning for feature extraction, the system can effectively identify female faces in video feeds. The research demonstrated the model's effectiveness compared to other methods in the field. This approach is particularly useful for applications that require real-time processing, such as those involving user interactions. Salihbašić et al. (2019) [8] Developed an Android application that could recognize the gender, age, and faces based on their features. The app utilizes a library for image recognition, employing convolution techniques. Neural networks are integrated into the application to combine various features effectively. The detection and identification algorithms are designed for easy user understanding. An intuitive interface was created using Android Studio by the development team. The application demonstrates impressive performance with different types of images, making it ideal for real-life situations and automated systems. Nathasia Florentina Thejowahyono et al. (2022) [9] developed a system for recognizing hand gestures designed to identify distress signals communicated through hand movements for communication purposes. This approach utilized machine learning techniques to analyze the data. The system employed MediaPipe technology, which enabled it to recognize various hand movements and gestures. It could autonomously send help messages via email without any human intervention. This feature had been successfully implemented, and the accuracy rate was 98.79%. Celal Akcelik et al. (2021) [10] developed a system for recognizing faces and emotions. The Viola-Jones algorithm and the Mini Xception model are being utilized for this purpose. Gender identification is being explored through 18 experiments conducted by the system. The emotion recognition component achieved a precision rate of 93.11%.

III. Pre-Trained Models

A pre-trained model is a kind of machine-learning model that has been trained to recognize images, speech, etc. . Rather than starting from scratch with a lot of data and processing capacity, this model can be used as a basis for future growth by utilizing the helpful patterns and features it has discovered from a huge dataset. Then, with less data and work, you may optimize the pre-trained model for your particular activity.

Below are a few pre-trained models that are widely used for gender detection:

A. Xception and Mini-Xception Model

The Xception model is a kind of deep learning framework applicable to some applications, like the classification of images. The Mini Xception version has smaller numbers of layers and is lower in terms of parameters than the full Xception version. Table *1* shows the comparison between the Xception and Mini Xception models based on the layers.

B. FaceNet

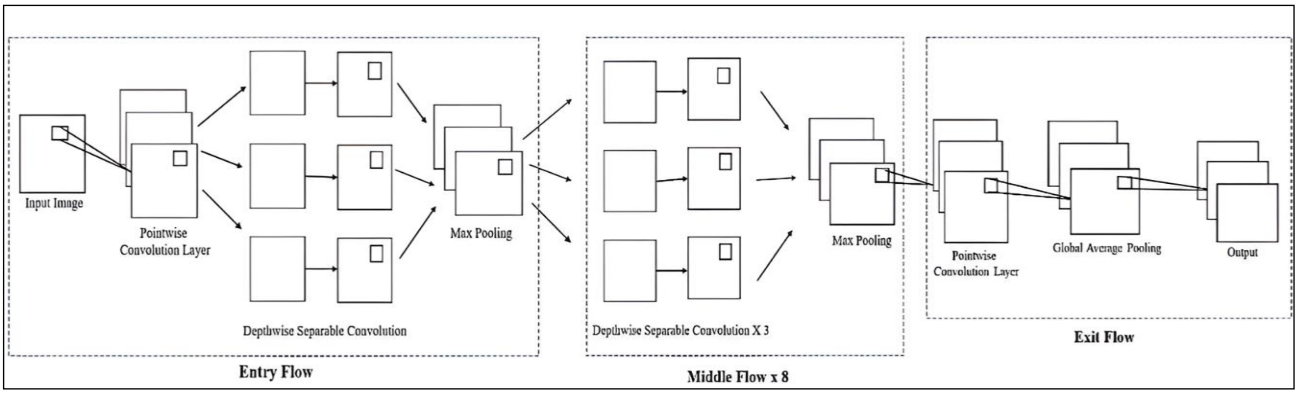
FaceNet is predominantly trained on Google initially dataset-based (which comes from CASIA-WebFace and VGGFace2 datasets) which are highly curated, labelled large-scale face images collected from various sources. It can be fine-tuned and evaluated on datasets like MS-Celeb-1M and LFW (Labelled Faces in the Wild) to improve its performance for face recognition tasks. Triplet Loss is employed for aiding FaceNet in extracting low-dimensional feature representations to facilitate face recognition and verification, which compares three images—an anchor (target face), a positive (matching face), and a negative (different face)—to improve accuracy. While, FaceNet is powerful in detailed face verification Mini-Xception, with a lighter and faster architecture, is more appropriate for use in real-time applications where high speed and lower computational cost are required.

Table 1 Comparison between the Xception and Mini-Xception Model

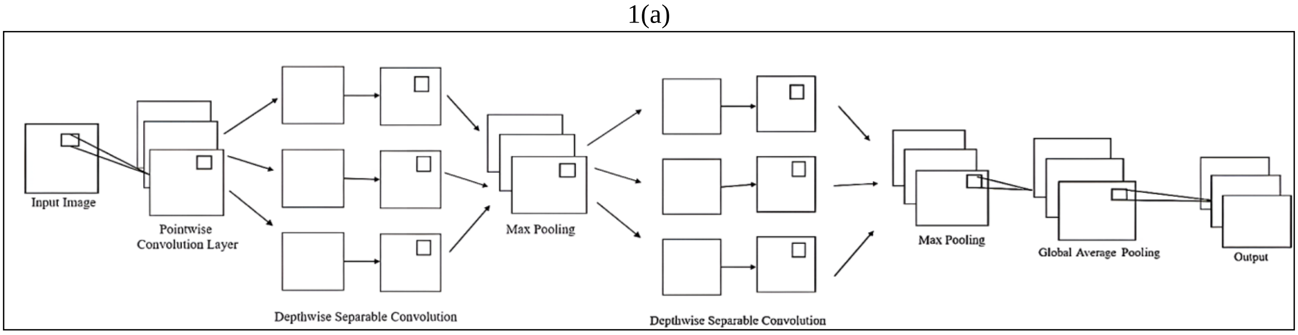
Features	Xception Model	Mini-Xception Model
Architecture Type	Uses depth-wise DNN. Figure 1(a) shows the flow can be categorized into three phases, which are as follows: Entry Flow, Middle Flow, and Exit Flow.	Figure 1(b) shows a Lightweight version of Xception. Ideal for real-time tasks like face detection, emotion recognition, and on-device image classification.
Depth-wise Separable Convolutions	Convolution is split into two steps: depth-wise (n×n) filters, and pointwise (1×1)	Uses depth-wise separable convolutions with fewer layers and fewer filters, starting at 32.
Total Number of Layers	Has 71 layers with three flows: Entry, Middle, and Exit.	Simplified Xception with 20-25 layers and 10-12 convolutional layers.
Performance	High accuracy in object detection and image classification but requires substantial computation and memory.	Optimized for speed and efficiency,
Input Size	299x299 pixel images	64x64 pixel inputs

C. VGG Face Model

VGG Face model for face recognition uses 3 × 3 convolution filters to capture facial details. As it is trained on over 2.6 million images from the VGG Face dataset, which features at least 2,600 individuals, it effectively detects unique facial features. Its hierarchical approach enhances features step by step, making it highly effective for face verification and identification. However, Mini-Xception is more efficient for real-time applications due to its smaller size and lower computational demands, making it better suited for resource-constrained devices than the heavier VGG Face model.



1(a)



1(b)

Figure 1 (a) Xception Model (b) Mini-Xception Model

IV. Proposed System

The system proposed in Figure 2 is designed to perform real-time gender prediction and hand gesture detection. The overall workflow begins by loading a pre-trained Mini-Xception model for gender prediction and initializing MediaPipe for face and hand gesture detection. The user is prompted to choose between various input options: processing images from files, videos, a dataset in ZIP format, or directly capturing live input from a webcam. Once the input is provided**,** MediaPipe is used to detect faces in the input frames. Detected faces are pre-processed by converting them to converting to grayscale and resizing to meet the model's input specifications, and normalizing the pixel values. After preprocessing, the system predicts gender using the Mini-Xception model.

In the second phase, the system detects hand gestures using MediaPipe. If the thumb is higher than the other fingers, it recognizes a fist gesture; if the thumb is lower, it identifies an open hand gesture. The recognized gesture is then displayed to the user.

This modular approach enables the system to seamlessly integrate gender detection and gesture recognition for use in real-time applications such as monitoring, security, etc.

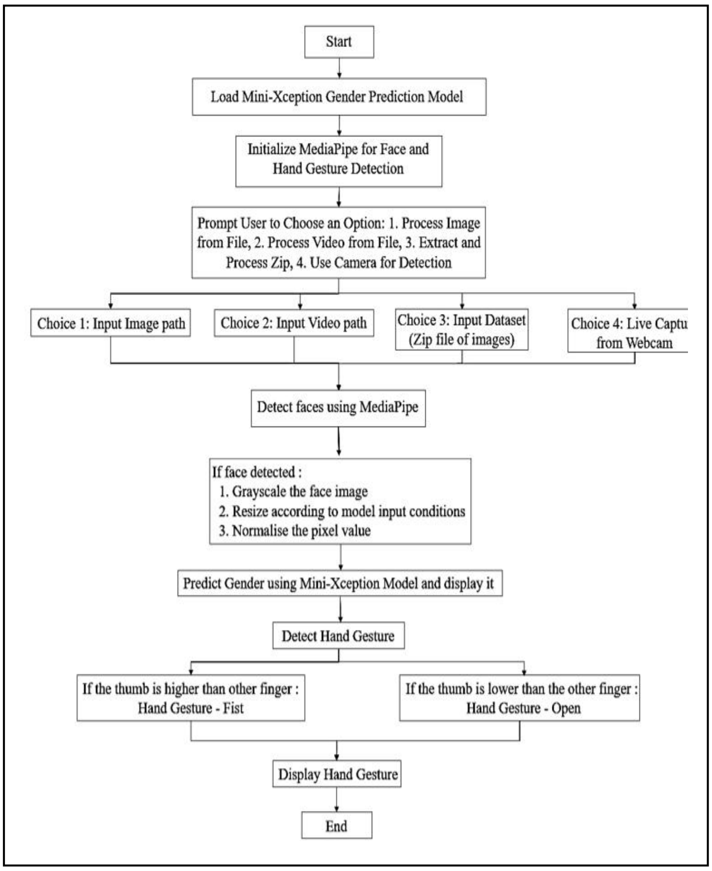


Figure 2 Workflow for Gender Prediction and Hand Gesture Detection System

The implementation uses deep learning for gender detection and MediaPipe for real-time hand gesture recognition. By processing images and video feeds, it identifies gender and gestures, enhancing user interaction. Our dataset comprises 50 images—28 male and 22 female—providing balanced gender representation. Figure 3 displays a sample from the dataset.

1. **Load Necessary Libraries:**
   Import required libraries: zipfile, os, cv2, MediaPipe, NumPy, and tensorflow.keras.
2. **Load Pre-trained Model:**
   Load the pre-trained gender detection model from the specified path.
   Define a dictionary for gender labels (0 for Female, 1 for Male).
3. **Initialize MediaPipe Components:**
   Set up MediaPipe for hand detection and face detection with appropriate parameters.
4. **Define Helper Functions:**
   preprocess_face(face):
   Convert the face image to grayscale, apply histogram equalization, resize to 64x64 pixels, normalize pixel values, and return as an array.
   detect_gender_from_face(face):
   Prepare the face data and apply the model to determine gender, returning the prediction and confidence score.
   classify_hand_gesture(landmarks):
   Extract the y-coordinates of the thumb and fingers to determine if the gesture is a 'fist' or 'open hand', then return the gesture.
   process_image_for_detection(image_path):
   Read the image, resize it, and convert it to RGB. Use MediaPipe for face detection; for each detected face, predict gender and display results. Also, the process for hand detection and classification gestures.
   extract_and_process_images(zip_path):
   Create a directory, clear existing files, extract images from the zip file, and process each image using process_image_for_detection().
   process_video_for_detection(video_path=None, use_camera=False):
   Initialize video capture and read frames. For each frame, convert to RGB, detect faces, classify gender, and process hand gestures. Display results and allow exit with 'q'.
5. **Define Main Function:**
   Prompt the user to select an option (process image, video, zip file, or use camera).
   Based on user choice, call the corresponding processing function.
6. **Execute Main Function:**
   Call the main() function to start the program. This algorithm utilizes computer vision techniques to detect faces and hand gestures. By using models

for gender prediction and hand gesture identification, it accurately detects and displays information from both images and video feeds. This powerful approach provides real-time analysis for other purposes, significantly

Figure 3 Sample of images from created dataset

enhancing user engagement and satisfaction.

V. Performance And Accuracy Evaluation

The Performance and accuracy in the effectiveness of the detection model in terms of gender was conducted using a dataset comprising 50 images, of which 28 depicted males and 22 depicted females. The main objective of the study was to evaluate the model's ability to accurately classify individuals' genders based on their facial features. The results showed that the model correctly identified the gender of 39 out of 50 individuals, resulting in an overall accuracy of 78%.

1. Model Evaluation Metrics

Accuracy represents the proportion of correctly predicted instances to the total evaluated instances, calculated using the formula:

$$\text{Accuracy} = \frac{\text{True positives} + \text{True negatives}}{\text{Total samples}}$$

Precision is the ratio of correct positive predictions to all predictions classified as positive.
Recall (sensitivity) measures how many of all the correctly classified relevant instances within the data set the model can classify correctly.
The F1-Score is the average of precision and recall.

2. Confusion Matrix

	Predicted Male	Predicted Female
Actual Male	20	5
Actual Female	3	22

Table 2 Confusion Matrix

From Table 2, we derive the following performance metrics:

- True Positives (TP): 20
- True Negatives (TN): 22
- False Positives (FP): 3
- False Negatives (FN): 5

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

$$\text{Accuracy} = \frac{20 + 22}{20 + 22 + 3 + 5} = \frac{42}{50} = 0.84$$

Thus, the accuracy of the model is **84%**.

VI. Result

This section shows the results of the gender classification and hand gesture detection model and explains what they mean. The gender classification model correctly identified **39 out of 50** individuals, giving it an accuracy of 84%. These results are displayed in the images. Figure 4 shows a sample of output. The confusion matrix confirms achieving high precision and recall for both gender classifications, demonstrating the model's reliability across varying facial features and conditions.
The model successfully detected SOS gestures, as illustrated in Figure 4. This functionality is crucial for safety applications, allowing users to signal for help in emergencies. For instance:

- **Personal Safety**: A person can discreetly use the SOS gesture to alert bystanders.
- **Outdoor Adventures**: Hikers can signal for assistance when lost or injured.
- **Workplace Safety**: Employees can quickly communicate emergencies without verbal cues.

The results validate the integration of gender classification and hand gesture detection, indicating promising applications in security and human-computer interaction. Future enhancements by Expanding and diversifying the datasets could enhance generalization. The findings confirm the model's performance in real-time applications, paving the way for advancements in gesture recognition technology.

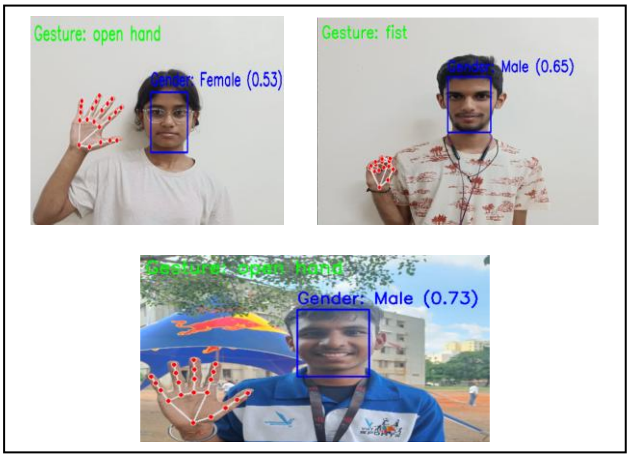


Figure 4 Sample of Output

VII. CONCLUSION

The mini-Xception model, when combined with MediaPipe, successfully demonstrated its efficiency for real-time gender classification and hand gesture detection in this study. Achieving an accuracy rate of 84% in gender classification reflects the model's robustness and potential for practical applications across various domains like security, healthcare, and human-computer interaction. Incorporating SOS gesture recognition significantly enhances the system's capacity to provide help in times of need. This functionality improves user safety, demonstrating how technology can enhance security measures for users. In emergencies, this feature can serve as a means of communication during field trials. Conducting studies on user experience will provide insights into the application of this technology in various fields such as security and healthcare services. Additionally, we will examine how devices can aid individuals with disabilities.

We will explore ways in which interactions with devices can enhance safety measures, such as features that send alerts to emergency services. When an SOS signal is recognized, our focus shifts to providing assistance.

Enhancing the dataset will improve the model's performance. I will explore methods of operation and delve into advanced strategies, such as transfer learning, to further enhance performance. This research provides a sound basis for further developments. Advancements in technologies that employ recognition and tracking functionalities can greatly enhance user interaction and safety through the use of recognition technology.

VIII. FUTURE SCOPE

The results of this study strongly suggest that progress is possible in identifying gender. Technologies that recognize gestures are being investigated through a range of planned improvements. The usability of these systems is crucial for meeting user needs and enhancing interactions. We aim to grow the dataset by increasing both its dimensions and volume. Diversity plays a significant role in enhance the accuracy and dependability of the model, and it will be able to work more effectively across various situations. Including a range of visuals can help ensure that different communities are represented. The analysis also takes into account the makeup of the population and the quality of lighting in the area. We are working on creating a user app that combines gender determination and distress signal recognition. This real-life application will be incredibly valuable for providing assistance, ensuring safety measures, and offering healthcare monitoring services for users.

ACKNOWLEDGEMENT

The authors wish to extend their gratitude to Mrs. Dipti Pandit for their guidance and support, throughout this research project. Their expertise and encouragement were instrumental in the completion of this study. We also want to acknowledge Vishwakarma Institute of Information Technology for providing the resources and infrastructure, for our research efforts.

IX. REFERENCES

[1] A. Kumar, "Face Detection and Recognition using OpenCV," *International Journal of Computer Applications (0975 – 8887),* 2022.

[2] S. Hore, "Implementation of AI for Subsequent Gender Detection and Inclusive Approaches for Representation," *International Journal of Biomedical Engineering, 25(1), 1-8. - did not match any articles.,* 2023.

[3] I. Verma, "Age Prediction using Image Dataset using Machine," *International Journal of Innovative Technology and Exploring Engineering (IJITEE),* 2019.

[4] A. Sain, "Automated Facial Recognition-based Attendance System using OpenCV in Python," *International Journal of Scientific Research in Computer Science, Engineering, and Information Technology, 9(6), 105-111.,* 2023.

[5] C. N. Esomonu, "Hand Gesture Recognition System Using Keypoints," *International Journal of Scientific Research in Computer Science, Engineering, and Information Technology, 11(3), 45-52.,* 2023.

[6] S. D. G. Boncolmo, "Gender Identification Using Keras Model Through Detection of Face," *School of Electrical, Electronics, and Computer Engineering, Mapúa University, Manila, Philippines.,* 2023.

[7] R. Thangaraj, " Deep Learning-based Real-Time Face Detection and Gender Classification using OpenCV and Inception v3.," *International Journal of Scientific Research in Computer Science, Engineering, and Information Technology, 9(7), 200-206.,* 2023.

[8] Salihbašić, "Development of Android Application for Gender, Age, and Face Recognition Using OpenCV," *Journal of Computer Science and Technology, 38(4), 567-578.,* 2023.

[9] A. &. O. Salihbašić, " Development of Android Application for Gender, Age, and Face Recognition Using OpenCV.," *Journal of Computer Science and Technology, 38(4), 567-578.,* 2023.

[10] C. Akcelik, "Face and emotion recognition using deep learning based on computer vision methods," *International Journal of Emerging Technology and Advanced Engineering, 11(7), 21-28.,* 2021.