IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.066 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 12, Issue 4, April 2025

DOI: 10.17148/IARJSET.2025.12405

VISIONSENSE: FACE IDENTITY, EMOTION, AGE, AND GENDER PREDICTION USING MACHINE LEARNING

Dr. Girish[1], Satyajith Manohar[2]

Professor and Head, Information Science and Engineering, NIE, Mysuru, India[1]

Student, Information Science and Engineering, NIE, Mysuru, India[2]

Abstract: "VisionSense" is an AI-powered facial recognition system that performs real-time face identity, age, gender, and emotion prediction. It uses a Convolutional Neural Network (CNN) with pre-trained weights for feature extraction. It uses a pre-trained dlib model for face alignment. It uses a pre-trained Wide Residual Network model for age and gender predictions. It uses another Convolutional Neural Network (CNN) for emotion detection. For identity classification, the system offers a choice between K-Nearest Neighbours (KNN) and Support Vector Classification (SVC). VisionSense is designed to process images, videos, and live camera feeds, automatically saving unknown faces and updating its model through retraining whenever a new customer is detected. The system provides a web-based API for seamless image and video input, allowing users to interact with it effortlessly.

Keywords: Convolutional Neural Network, Wide Residual Network, K-Nearest Neighbours, Support Vector Classification, Machine learning, Face classification, Emotion detection

I. INTRODUCTION

Facial recognition and analysis are essential in security and customer analytics. Existing systems often lack accuracy or real-time processing. VisionSense provides a reliable solution for identifying individuals, detecting emotions, and estimating age and gender with precision. Manually updating face recognition databases is inefficient. VisionSense automatically detects and stores unknown faces, which ensures that the model continuously learns new faces through periodic retraining. This allows the system to recognize new individuals without manual intervention. Real-time processing is essential for security and analytics. VisionSense processes live video feeds and images instantly. The system is user friendly as users can access all the features through a simple web-based interface.

VisionSense is an AI-powered solution that performs real-time face recognition, emotion detection, age estimation, and gender classification. It utilizes a combination of pre-trained deep learning models for feature extraction, face alignment, and classification. The system automatically detects and saves unknown faces, retrains itself dynamically, and can process both images and videos. It is designed for security, customer analytics, and human-computer interaction applications.

II. LITERATURE REVIEW

Over the years, many different image classification methods have been developed bringing various advancements in face recognition technologies. Some studies focus on boosting real-time performance using GPU acceleration like TensorRT, while others introduce lightweight models like EdgeFace that combine CNNs and Transformers for efficient edge deployment. Research also explores deep learning-based systems for real-time applications like attendance tracking using models such as dlib. Comprehensive reviews of machine learning techniques in face recognition reveal key challenges like lighting conditions and privacy issues. Additionally, newer methods for masked face recognition using image augmentation aim to improve identity preservation.

Recent advancements in face recognition have focused on enhancing speed, accuracy, and efficiency through GPU acceleration, lightweight models, and deep learning techniques. For instance, TensorRT acceleration and GPU computing have been used to significantly boost the performance of face recognition systems in mobile applications, allowing for

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.066 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 12, Issue 4, April 2025

DOI: 10.17148/IARJSET.2025.12405

faster and more efficient inference tasks [1]. Lightweight models like EdgeFace combine CNNs and Transformers in a compact architecture, achieving high accuracy while maintaining low computational costs, making them ideal for edge devices with limited resources [2]. In practical applications, such as attendance systems, real-time face detection using models like dlib and YOLO, combined with deep learning techniques, ensures fast and reliable performance in institutional settings [3]. A broader review of machine learning and deep learning methods has outlined challenges such as varying lighting conditions, camera angles, and ethical concerns, while also identifying key trends and research gaps in person detection and identification technologies [4]. Additionally, advancements in masked face recognition propose novel image augmentation techniques that improve identity preservation and recognition accuracy by focusing on occluded regions, rather than reconstructing the full face [5].

III. PROPOSED METHODOLOGY

The proposed methodology for VisionSense involves a modular, AI-powered system designed for real-time face recognition, emotion detection, age estimation, and gender classification. It integrates several deep learning models, including CNNs for feature extraction and emotion recognition, dlib for face alignment, and WideResNet for age and gender classification. The system includes an identity classification module using KNN and SVC, which identifies known individuals and flags unknown faces for storage and retraining. Through its Image and Video Predictor Modules, the system processes both static and dynamic inputs, applying predictions and overlaying labels. An Image Saving Module collects unknown face data, which is later used by the Identity Trainer Module to dynamically retrain classifiers using multithreading for seamless real-time performance. A Flask-based Web API enables user interaction by allowing uploads of images and videos and delivering labelled results. This structured, modular approach ensures high accuracy, efficiency, and adaptability, making VisionSense suitable for applications in security, analytics, and human-computer interaction.

A. System Architecture and Implementation

The project is developed using a Flask-based Web Application Framework, making HTTP requests-responses and Web-socket connections possible. Various routes are configured to implement the tasks of face identification through image files (/image), face identification through video files (/video), and manually retraining the classifier models (/manual_retrain). Web-socket is used for live camera feed.

- Image File Detection: Users can upload an image file, and if the format is valid, the system detects faces and returns a labeled image for download, otherwise, an error message is shown.
- Video File Detection: When a video file is uploaded, the system validates the format, and the format is valid the system processes it to return a labeled video, else prompts the user to re-upload a valid file.
- Live Camera Detection: Users can enable the live camera with controls for retraining and saving unknown faces, the system processes each frame in real-time, saves unknown identities, and retrains models in the background as needed.
- Automatic Model Retraining: The system automatically checks at regular intervals whether new faces have been detected and triggers the retraining of classifier models in a separate thread without interrupting real-time predictions.
- Manual Model Retraining: By clicking the manual retrain button, users can trigger the system to retrain classifier models with the latest stored face data.

B. Working of the Face Identity Predictor based on CNN, Wide Residual Network, SVC and KNN

1. **Live Camera Initialization:** The user sets parameters like automatic retraining interval and whether to save unknown faces, then starts the live camera feed for real-time face analysis.
2. **Face Detection and Alignment:** Each captured frame is passed through the face alignment module using a pre-trained dlib model, which detects and aligns facial features to standardize input.
3. **Deep Feature Extraction:** Aligned face images are processed by a pre-trained CNN that extracts facial embeddings which are numerical representations used for identity and trait classification.
4. **Identity Prediction:** The identity classifier module uses the extracted embeddings with K-Nearest Neighbors (KNN) and Support Vector Classifier (SVC) models to classify faces as known or unknown.
5. **Age, Gender, Emotion Recognition:** Simultaneously, the WideResNet model estimates age and gender, while a CNN-based emotion classifier analyzes expressions to determine emotional states.
6. **Labeling and Display:** Predictions for identity, age, gender, and emotion are overlaid on the video frame and displayed to the user in real-time through the camera interface.
7. **Retraining and Face Saving:** If new unknown faces are detected, the system saves multiple copies and retrains the classifier models automatically or manually in the background using multithreading for efficiency.

C. Figures:

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.066 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 12, Issue 4, April 2025

DOI: 10.17148/IARJSET.2025.12405

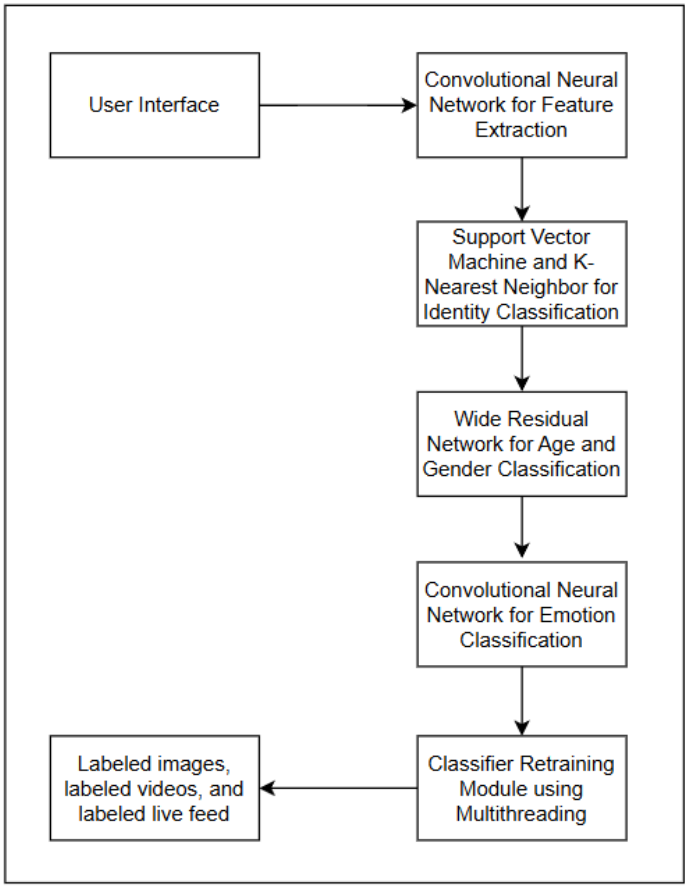

Fig. 1: Proposed System

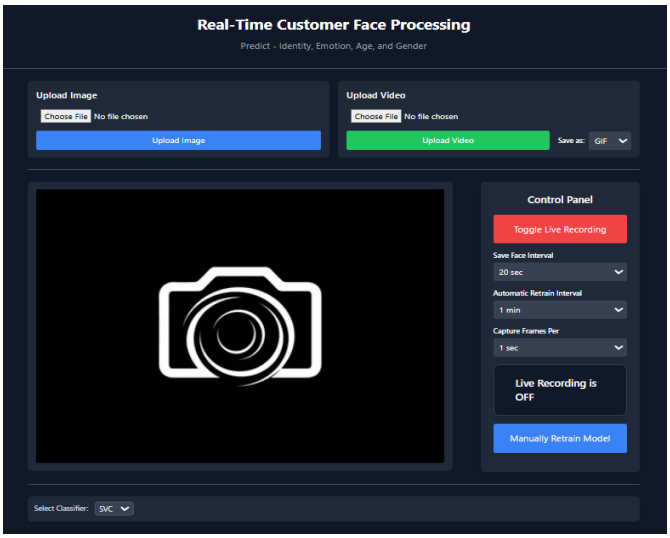

Fig. 2: User Interface

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.066 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 12, Issue 4, April 2025

DOI: 10.17148/IARJSET.2025.12405

IV. CONCLUSION

This study presents an innovative approach aimed at improving the precision and performance of face identity detection systems. The VisionSense system is an AI-powered solution designed for real-time face recognition, along with age, gender, and emotion detection, using a combination of pre-trained machine learning models. It integrates CNNs for deep feature extraction, dlib for face alignment, and WideResNet for age and gender classification, while offering SVC and KNN as options for identity prediction. The system is capable of processing static images, videos, and live camera feeds, automatically labelling detected faces and saving unknown identities to continuously expand its training dataset. Its ability to retrain classifier models either automatically or manually ensures that recognition accuracy improves over time, adapting dynamically to new data. VisionSense also leverages parallel threading to allow real-time predictions and model retraining to occur simultaneously without affecting performance. With a simple and intuitive web-based interface, the system is user-friendly and suitable for a wide range of applications such as security, customer service, and data analytics. Despite challenges like computational demands and dependency on quality training data, VisionSense delivers a reliable and scalable solution for real-time face analysis.

REFERENCES

[1] H. Li, A. Li, Y. Liu, Y. Lin, and Y. Shi, “AI Face Recognition and Processing Technology Based on GPU Computing”, JTPES, vol. 4, no. 05, pp. 9–16, May 2024, doi: 10.53469/jtpes.2024.04(05).02.

[2] A. George, C. Ecabert, H. O. Shahreza, K. Kotwal and S. Marcel, “EdgeFace: Efficient Face Recognition Model for Edge Devices,” in IEEE Transactions on Biometrics, Behavior, and Identity Science, vol. 6, no. 2, pp. 158-168, April 2024, doi: 10.1109/TBIOM.2024.3352164.

[3] K. Painuly, Y. Bisht, H. Vaidya, A. Kapruwan and R. Gupta, “Efficient Real-Time Face Recognition-Based Attendance System with Deep Learning Algorithms,” 2024 International Conference on Intelligent and Innovative Technologies in Computing, Electrical and Electronics (IITCEE), Bangalore, India, 2024, pp. 1-5, doi: 10.1109/IITCEE59897.2024.10467743.

[4] Amirgaliyev B., Mussabek M., Rakhimzhanova T., Zhumadillayeva A., “A Review of Machine Learning and Deep Learning Methods for Person Detection, Tracking and Identification, and Face Recognition with Applications,” Sensors 25, no. 5: 1410, 2025, doi: 10.3390/s25051410

[5] S. Malakar, W. Chiracharit and K. Chamnongthai, “Masked Face Recognition With Generated Occluded Part Using Image Augmentation and CNN Maintaining Face Identity,” in IEEE Access, vol. 12, pp. 126356-126375, 2024, doi: 10.1109/ACCESS.2024.3446652.