

# Multi-Human Behaviour Analysis

Dr. SivaSundaraPandian .S  
Dept. of Computer Science  
& Engineering  
ACS College of Engineering  
Bangalore, India  
[desivasundarapandian@gmail.com](mailto:desivasundarapandian@gmail.com)

Chethan.S  
Dept. of Computer Science  
& Engineering  
ACS College of Engineering  
Bangalore, India  
[chethan1309@gmail.com](mailto:chethan1309@gmail.com)

Channabasava  
Dept. of Computer Science  
& Engineering  
ACS College of Engineering  
Bangalore, India  
[bchannabasava053@gmail.com](mailto:bchannabasava053@gmail.com)

Chetan koparde  
Dept. of Computer Science  
& Engineering  
ACS College of Engineering  
Bangalore, India  
[chetankoparde2004@gmail.com](mailto:chetankoparde2004@gmail.com)

GaneshKumar Bijjal  
Dept. of Computer Science  
& Engineering  
ACS College of Engineering  
Bangalore, India  
[ganeshbijjal07@gmail.com](mailto:ganeshbijjal07@gmail.com)

## ABSTRACT

The Multi-Human Behavior Model uses AI to detect and evaluate human feelings (or emotions), physical activities (or motions), and attentiveness or focus in real time. The model does this by using Convolutional Neural Networks (CNN), MediaPipe Pose Estimation and various computer vision algorithms (software coding), which together allow for the simultaneous evaluation of multiple types of behaviour from live streams.

Facial Expressions are used to classify emotions such as happiness, sadness, anger, neutralness and surprise. Body posture and activity such as; sitting, standing, or walking are detected through the evaluation of body position using pose-landmark estimation. The model also monitors the direction of people's faces (left, right, up, down, and straight) as a measure of user attention and concentration towards others.

A web-based user interface built using Flask is also provided to users, which allows users to activate the individual detection modules and see the results in real time. The successful integration of emotion recognition, physical activity detection, and attention tracking into one intelligent model makes this model a scalable, cost-effective, and real-time solution for smart-classrooms, driver safety systems in vehicles, health monitoring systems (hospitals, patient care facilities, etc.) or for general surveillance systems.

## I.INTRODUCTION

The significance of analyzing human behavior is vital in newer intelligent systems. The traditional means of monitoring human behavior rely on a human observer, who is often slow and prone to mistakes. Through new developments in Artificial Intelligence these processes can be automated using deep learning and computer vision techniques to evaluate an individual's facial expressions, posture and attentiveness.

Convolutional Neural Networks allows classification of emotions based on facial features, and MediaPipe pose estimation can provide an efficient means to identify a person's skeletal landmarks for activity detection. By combining these two technologies together, machine systems will be able to assess psychological behaviors of an individual while at the same time assessing their physical behaviors.

The Multi-Human Behavior Analysis Model is being proposed as a real-time AI-based intelligent system that combines emotion detection, activity recognition and face focus tracking into one system, in a smart monitoring environment.

Deep learning techniques have significantly improved the performance of computer vision systems. CNN models are capable of automatically extracting hierarchical features from images, making them highly effective for facial emotion recognition tasks. In this work, CNN is used to analyze facial expressions from live video frames and classify emotions such as happiness, sadness, anger, and surprise. The use of deep learning enables accurate real-time behavior detection without the need for manual feature extraction.

## II. LITERATURE SURVEY

Recent research has focused on recognizing human behaviors such as emotion detection and activity recognition using deep learning techniques. Convolutional Neural Networks (CNN) have been widely used for facial emotion recognition due to their higher accuracy compared to traditional methods [1]. Landmark-based approaches have also been proposed for real-time emotion recognition by analyzing facial features efficiently [2].

Human pose estimation methods using MediaPipe have enabled accurate tracking of body joints in real time with low computational requirements [3]. Advanced deep learning models such as hybrid 3D-CNN and ConvLSTM have been developed for recognizing facial expressions in video sequences [4]. In addition, human behavior analysis using CNN and multimodal data has been explored to better understand human actions and emotions [6].

However, most existing systems focus on detecting a single behavioral parameter. Therefore, this study aims to develop a unified framework that integrates emotion recognition, activity detection, and attention tracking in real time.

This study will address this gap, and develop a multi-behaviour detection framework.

## III. PROBLEM STATEMENT

Understanding human behaviours is vital for safety and efficient work throughout many aspects of daily life including schools, offices, hospitals and public locations. Current systems have many deficiencies:

- The cost of continuous human supervision is both time intensive and variable.
- Existing systems monitor either emotion or physical activity individually.
- There is no way to monitor multiple behaviours in real time in an integrated manner.
- Current sensors have a limited ability to accurately record the behaviour of an individual when lighting and/or other people are in close proximity.

Hence, the requirement for a smart system which can record emotion, physical activity and attention levels in real time has become apparent.

## IV. OBJECTIVE

The aim of this multi-analytical model is to create an automated system for detecting behavior in real time by combining multiple artificial intelligence technologies into one cohesive intelligent framework. A trained convolutional neural network (CNN) model will give an accurate representation of people's emotions through face detection; MediaPipe's pose estimation model will allow for the detection of people's physical activity; and analysis of the orientation of people's faces will help to determine whether the individual is attentive.

An additional goal of this model's development will be to

integrate the above-mentioned detection features into a web-based interface built on Flask, which will enable users to both visualize the behavior as it is occurring and provide interactions that will enhance the quality of the information received through these behavior detection technologies. The model will also be developed to ensure robust performance under normal lighting conditions, while maintaining the scalability and modularity needed to support continued improvement of the performance by adding enhancements over time. Ultimately, the goal will be to build a practical and efficient model that can be used to provide smart monitoring solutions in multiple application areas.



fig : Overview of System Objectives

## V. METHODOLOGY

The proposed approach consists of implementing a pipeline for AI processing by using structured programming methods to integrate multiple AI modules together.

Face detection will first occur on each frame of live video has been captured via OpenCV (webcam). The webcam will continuously "stream" video frames, which will each be processed sequentially for behavioral analysis.

The first part of the analysis is to execute face emotion recognition. In order to do this, a face detection model will determine where in a video frame does it see an identified face. The identified face will be extracted and converted to grayscale before resizing it to fit the input format of the pre-trained CNN model being used to classify the identified emotion expressed by the face. The processed image will then be fed into the pre-trained CNN model, which will perform a probabilistic assessment of the identified emotion determined by the input image as compared to the other learned image representations stored in the model. The predicted emotion will be shown both as text and as a confidence score overlaid on the video frame.

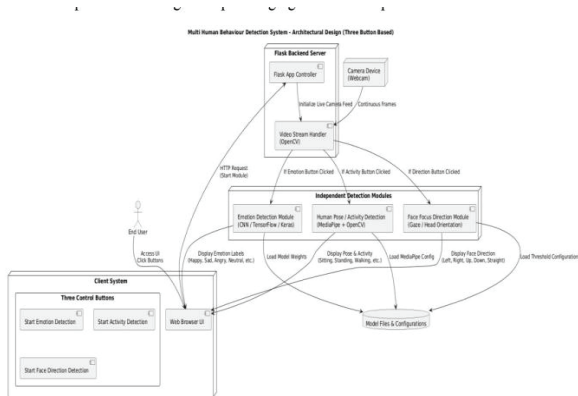


Fig : architecture diagram

While the face emotion recognition is being performed, the MediaPipe Pose framework will determine 33 points-of-reference skeletal landmark coordinates representing the foundational (i.e., elbow) and the variable (i.e., other humeral bony landmarks) locations of the joints. The joint angles and relative distance calculations for identified skeletal joints will be determined by their corresponding skeletal landmark reference coordinates. Once all the joint angles and relative distances will be calculated for an identified skeletal joint, physical activity classifications, such as sitting, standing, walking, and so forth, will occur through rule-based geometric calculations performed using the relationship between identified skeletal landmark coordinates.

Help face direction modules will analyze the nose and eye landmarks relative positions to determine left, right, up, down, or straight face orientations. This module will allow the system to be analyze attentiveness.

Finally, all detection results will be integrated and displayed in real time overlaid on the original outgoing video for visual analysis.

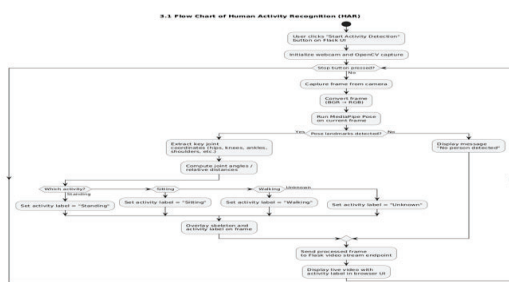


Fig 6.2 Levels of flow chart of HAR

Fig: methodology diagram

Finally, all detection results will be integrated and displayed in real time overlaid on the original outgoing video for visual analysis.

## RESULT

The effective real-time identification of human activity, facial expression and orientation performed by the system was achieved. Testing with normal ambient light up to 10 to 15 fps reveals that the frame process has very stable

performance. Emotion classification accuracy from the CNN model have been shown to be reliable for the training dataset supplied. Using MediaPipe pose estimation allows for the identification of skeletal landmarks and provides accuracy in posture classification.

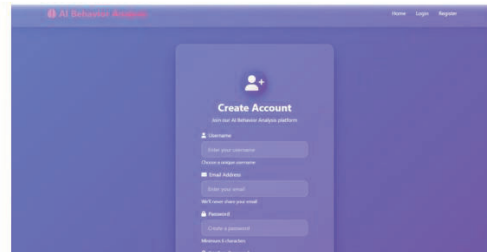


Fig : account page

The Flask Web Interface provides for error-free streaming and allows dynamic visualisation of the system's behaviour produced by the system. The functionality of the system to provide practical application in a smart classroom, surveillance monitoring, observing patients in a hospital, and monitoring the driver's level of attention makes it clear that the successful integration of multiple behavioural detection modules based on Artificial Intelligence into a connected real-time system is achievable.

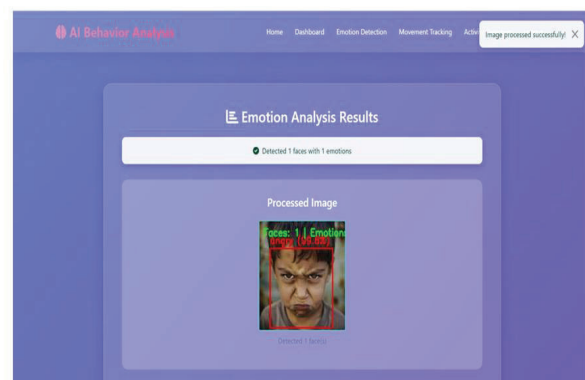


Fig : emotion analysis result

The functionality of the system to provide practical application in a smart classroom, surveillance monitoring, observing patients in a hospital, and monitoring the driver's level of attention makes it clear that the successful integration of multiple behavioural detection modules based on Artificial Intelligence into a connected real-time system is achievable.

## VII. DISCUSSION

Utilizing emotion recognition, activity recognition, and attentiveness measurement provides a complete profile of human behavior. Traditional software systems only consider individual pieces of data; whereas this multi-module system improves accuracy of monitoring and provides additional context for the parameters analyzed. This multi-faceted assessment of one person's emotional state, coupled with their physical position and direction of

focus, gives us much better insight into the behavioral characteristics of individuals.

A multitude of environmental factors — including lack of light, blockage, and camera quality — can impact system performance. Planned enhancements to the current systems include the capacity to track individuals concurrently, an upgraded deep learning architecture to improve accuracy, a cloud-based deployment, and a database allow's for long-term analytical review.

Despite the limitations of the current implementation, the Multi-Human Behavior Analysis Model illustrates the efficacy of artificial intelligence (AI) for real-time monitoring of human behavior and provides a solid foundation for future research and development.

#### REFERENCE

- [1] K. Sarvakar, *A Survey of Face Emotion Recognition Using Deep Learning Methods*.
- [2] A. Farkhod, "Development of Real-Time Landmark-Based Emotion Recognition.
- [3] J. W. Kim, Human Pose Estimation Using MediaPipe Pose and Optimization Method Based on a Humanoid Model.
- [4] R. Singh, Facial Expression Recognition in Videos Using Hybrid 3D-CNN and ConvLSTM.
- [5] R. Raj, An Improved Facial Emotion Recognition System Using Deep Learning.
- [6] A. Budhewar, Human Behaviour Analysis Using CNN and Multimodal Data.
- [7] S. Gupta, Facial Emotion Recognition Based Real-Time Learner Engagement Detection.
- [8] N. Jlidi et al., MediaPipe with GNN for Human Activity Recognition.
- [9] M. Mukhiddinov, Masked Face Emotion Recognition Based on Facial Landmarks and Low-Light Enhancement.
- [10] M. V. P. Kothari et al., A YOLO and MediaPipe-based Human Fall Detection System in Dynamic Environments.