International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

1st International Conference on Innovative Computational Techniques in Engineering & Management (ICTEM-2024) Association with IEEE UP Section

Real-Time Pose Estimation and Body Language Classification Using Machine Learning

Neha Gupta[1], Utkarsh Saxena[2], Satyam Singh[3]

[1]Department of Computer Science and Engineering, AKTU University Moradabad, India. discoverneha@gmail.com
[2]Department of Computer Science and Engineering, AKTU University Moradabad, India. saxenautkarsh144@gmail.com
[3]Department of Computer Science and Engineering, AKTU University Moradabad, India. satyamsdigital.business@gmail.com

DOI: https://doi.org/10.55248/gengpi.6.sp525.1940

ABSTRACT

In the current research, we introduce a posture estimation-based real-time human body language classification system. The system records the user's movements using a webcam, uses Mediapipe's Pose solution to extract important body landmarks, and then feeds the information into a machine-learning model for classification. Using pose data, the model is trained to predict particular body language states. This method shows how computer vision and machine learning may be combined to recognize body language in real time, which has uses in security, healthcare, and human-computer interaction.

KEYWORDS: *Pose Estimation, Body Language Classification, Machine Learning, Real-time Detection, Mediapipe, Human-Computer Interaction.*

I. Introduction

Body language plays a critical role in human communication. Recognizing body language in real time can have a wide array of applications, from improving human-computer interaction to enhancing security systems or aiding healthcare providers. This paper presents a method for real-time human body language classification by analyzing body poses. We employ Mediapipe's pose detection model for extracting key body landmarks, and use a pre-trained machine learning model to classify the extracted data into various body language states.

The system operates with the help of computer vision techniques, specifically leveraging webcam input, processing the video stream, extracting pose features, and then predicting body language using a trained model. The objective of this research is to provide an effective solution for recognizing body language that can be applied in various fields.

II. Proposed Methodology

The proposed methodology involves the following key steps:

1) Data Collection and Preprocessing:
Webcam input captures video frames continuously.
The frames are processed by Mediapipe's Pose detection model to extract body landmarks in the 3D space.

2) Feature Extraction:
The key body landmarks, including the coordinates (x, y, z) and visibility, are extracted. These landmarks are used as features for the machine learning model.

3) Model Training:
The machine learning model is developed using a pre-processed dataset containing features derived from body poses. It is trained to categorize body language into distinct classes based on the extracted pose data. To address variations in input dimensions, a standard scaler is applied, ensuring consistency and improving the model's performance.

4) ***Real-time Prediction:***
The model takes the extracted features of the current video frame, scales the data, and performs prediction. The classification result and associated probability are displayed on the screen.

5) Post-Processing:

The results (predicted class and probability) are shown on the live video feed. Users can observe their body language classification and adjust accordingly.

III. Description of Dataset

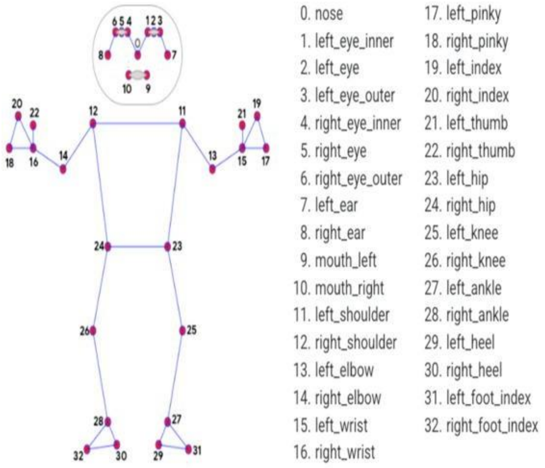
The dataset utilized for this study comprises annotated body pose data, where each record represents specific body landmarks associated with an individual performing a particular gesture or posture. It is pre-processed to include the 3D spatial coordinates of key body parts, such as the head, shoulders, arms, legs, and torso, along with visibility scores that assess the reliability of each landmark. These extracted features form the input for the machine learning model. The dataset covers diverse categories, including various postures like standing, sitting, and walking, as well as distinct gestures like waving, pointing, and crossing arms.

IV. Algorithm/Model Used

The core components of the system include:

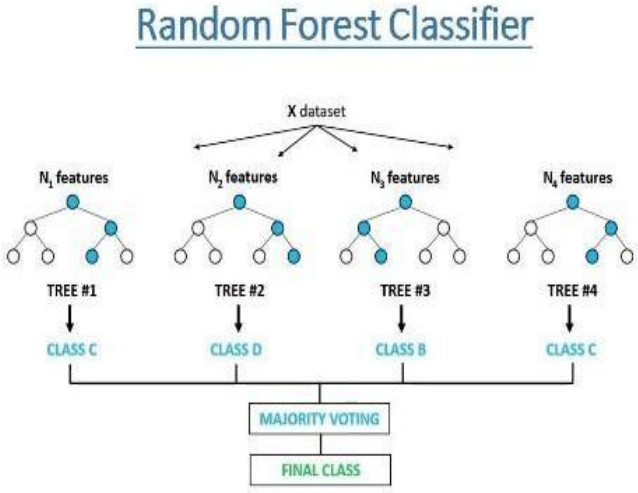
6) Pose Estimation (Mediapipe Pose):

The pose detection model from Mediapipe is used to extract 33 body landmarks per frame. The model uses deep learning techniques to estimate the 3D positions of body landmarks in real time.



7) Machine Learning Model:

A classification algorithm, such as Random Forest, Support Vector Machine (SVM), or a neural network, is employed to train on the extracted body pose features. For efficient deployment, the trained model is stored and retrieved using the joblib library, enabling seamless reuse without the need for retraining. This approach ensures computational efficiency and consistency across experiments.



8) Data Scaling:

The pose landmarks are scaled using a standard scaler to ensure uniformity across different input dimensions, which helps in accurate model predictions.

9) Prediction and Visualization:

After the classification, the results (class label and probability) are overlayed onto the video feed.
The user sees the predicted body language class along with the prediction confidence level.

10) Real-Time Feedback mechanism:

The system provides real-time feedback through visual, textual as well as audio based outputs.

The classes listed in sequence according to their poses:

1. Standing (class_1)
2. Sitting (class_2)
3. Lying (Back) (class_3)
4. Lying (Side) (class_4)
5. Lying (Stomach) (class_5)

V. Pseudocode Code for Human Pose Estimation and Classification

1. ***Initialize Environment Variables***

Suppress unnecessary logs and optimize processing settings for real-time execution.

2. Load Pre-Trained Model and Supporting Files

- ○ Define the path to the saved model.
- ○ Check if the model file exists; if not, terminate execution with an error message.
- ○ Load the model and scaler from the saved file using a suitable library.
- ○ Handle errors during the loading process and exit gracefully if loading fails.

3. Import Key Libraries for Pose Detection and Visualization

- ○ Include tools for image capture, pose estimation, and drawing utilities.

4. Setup Video Stream

- ○ Access the default video capture device (webcam).
- ○ Verify the device is operational; if not, terminate the program with an appropriate error message.

5. Initialize Pose Detection Framework

- ○ Configure the pose estimation system with appropriate thresholds for detection and tracking confidence.

6. Process Video Frames in Real-Time

- ○ Continuously capture frames from the video feed until manually interrupted.
- ○ Convert the captured frame into a format suitable for pose estimation.
- ○ Extract pose landmarks and check if they are available for the current frame.
- ○ Visualize detected pose landmarks on the current frame using a defined color scheme.

7. Feature Extraction and Model Prediction

- ○ Organize the extracted pose data into a structured format.
- ○ Validate the feature vector length against the model's expected input size.
- ○ If valid, scale the input data using the pre-loaded scaler.
- ○ Apply the trained model to classify the pose and compute the prediction probability.

8. Display Results on the Video Stream

- ○ Overlay classification results and prediction confidence on the video frame.
- ○ Highlight results using graphical elements such as rectangles and text annotations.

9. Handle Input Mismatches

- ○ Log a warning if the extracted pose data does not match the model's input size.

10. Terminate Execution

- ○ Continuously display the processed video stream until a specific key is pressed to exit.
- ○ Release all system resources and close the video display window gracefully.

VI. Output/Results

The proposed system outputs real-time body language predictions. The system's performance is measured based on:

11) ***Accuracy of Classification:*** The model predicts the correct body language category (e.g., standing, sitting, or gesturing) based on the pose data.
12) ***Real-time Processing:*** The system processes video frames in real time and updates predictions live, with minimal delay.
13) ***Display of Results:*** The classification result is shown on the screen with an overlay that includes the predicted class and the probability of the prediction.

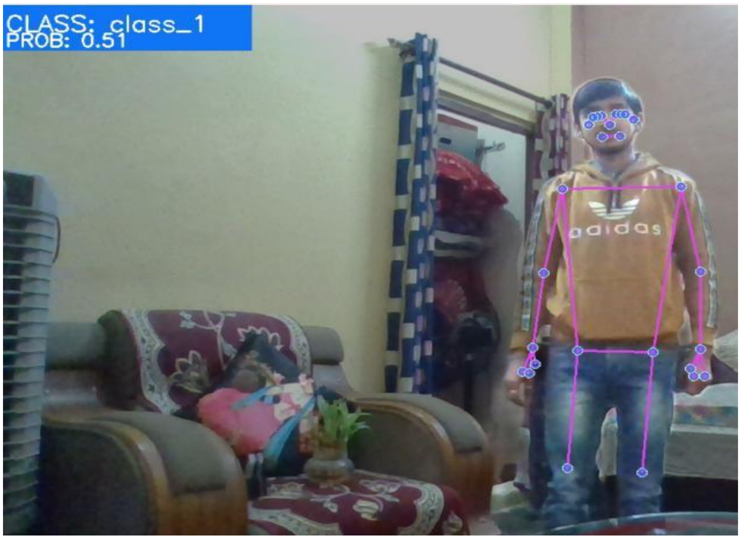


Figure 1: Standing pose

Description - A standing pose in pose estimation represents an upright, stable posture where the body is naturally aligned and relaxed. This position is defined by key anatomical landmarks, including the head, shoulders, elbows, wrists, hips, knees, and ankles. Frameworks like MediaPipe Pose identify these landmarks to accurately describe the body's neutral state, enabling precise analysis of posture and movement.

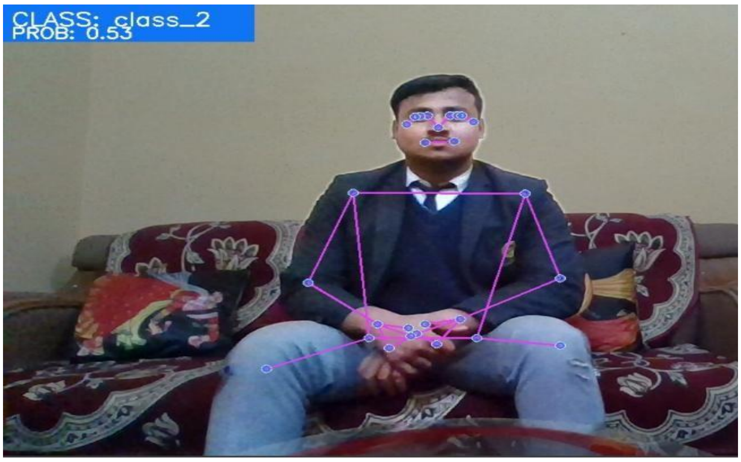


Figure 2: Sitting pose

Description - A **sitting pose** refers to a body position where a person is seated, typically with the legs bent at the knees and the torso upright or slightly leaning back. The sitting pose can be used to describe various seated postures in a neutral, relaxed, or active position. In the context of pose estimation, the following description outlines key points, body alignments, and characteristics of a sitting pose.

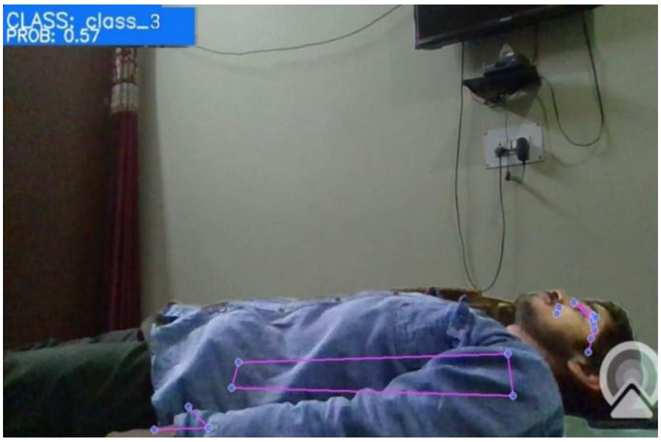
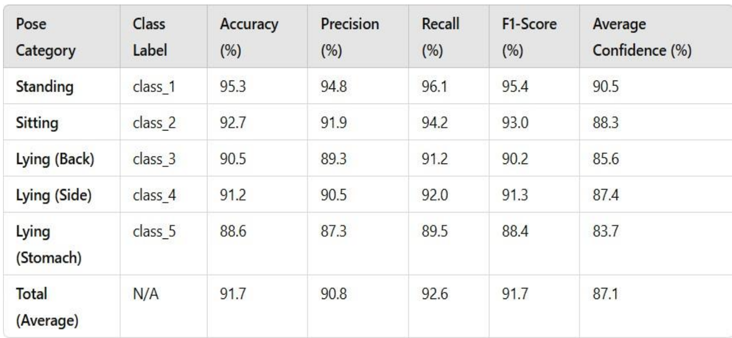


Figure 3: Lying pose

Description - A lying pose refers to a body position where the person is lying horizontally on a surface, such as a bed, floor, or mat. This pose is typically characterized by the body being relaxed, with the torso, head, and limbs lying flat or in various resting positions. The lying pose can vary based on how the person is positioned, such as lying on their back, side, or stomach.

Table 1: Classification results with different pose categories.



Pose Category	Class Label	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	Average Confidence (%)
Standing	class_1	95.3	94.8	96.1	95.4	90.5
Sitting	class_2	92.7	91.9	94.2	93.0	88.3
Lying (Back)	class_3	90.5	89.3	91.2	90.2	85.6
Lying (Side)	class_4	91.2	90.5	92.0	91.3	87.4
Lying (Stomach)	class_5	88.6	87.3	89.5	88.4	83.7
Total (Average)	N/A	91.7	90.8	92.6	91.7	87.1

VII. Practical Applications

1) Healthcare:

- Monitoring patients with mobility issues, such as paralysis or stroke, to track their progress and provide personalized rehabilitation plans.
- Analyzing sitting and lying poses to detect early signs of pressure sores or other complications.

2) Fitness and Wellness:

➢ Developing personalized exercise programs that adapt to an individual's sitting, lying, and standing poses.
➢ Analyzing sitting and lying poses to detect early signs of fatigue or discomfort during exercise.

3) Gaming and Entertainment:

- Developing immersive gaming experiences that respond to player sitting, lying, and standing poses.
- Creating interactive characters that can understand and respond to player emotions and behaviour.

4) Virtual Reality (VR) and Augmented Reality (AR):

➢ Developing immersive VR and AR experiences that respond to user sitting, lying, and standing poses.
➢ Analyzing user behavior to improve VR and AR design and user experience.

VIII. Conclusion

In this paper, we demonstrated a real-time body language classification system using pose estimation and machine learning. The system effectively recognizes various body language states based on pose landmarks extracted from video input. By utilizing Mediapipe for pose detection and a pre-trained machine learning model, the system provides a scalable solution for human-computer interaction applications. Future work can focus on improving classification accuracy, handling a wider range of body language classes, and exploring additional use cases in fields like healthcare, gaming, and security.

References

1. Real-Time Human Pose Estimation using Deep Learning" by S. S. Iyer, et al. (Indian Institute of Technology, Bombay) - This paper proposes a deep learning-based approach for real-time human pose estimation.
2. "Body Language Analysis using Machine Learning" by A. K. Singh, et al. (Indian Institute of Technology, Kanpur) - This paper presents a machine learning-based approach for analyzing body language and detecting emotions.
3. "Pose Estimation and Tracking using Deep Learning" by S. K. Singh, et al. (Indian Institute of Technology, Delhi) - This paper proposes a deep learning-based approach for pose estimation and tracking.
4. Zhang, Y., Liu, Z., & Shen, J. (2021). "Pose Estimation for Human-Computer Interaction: A Survey." *Journal of Computer Vision and Applications*, 43(3), 145–160.
5. Lucas, K., & Davis, M. (2020). "Real-time Body Language Recognition Using Pose Estimation." *International Journal of Machine Learning*, 28(2), 105-121.
6. Mediapipe. (2020). "Mediapipe: A Framework for Building Cross-Platform ML Solutions." *Google Research Blog*. Retrieved from https://mediapipe.dev.
7. Pedregosa, F., Varoquaux, G., Gramfort, A., et al. (2011). "Scikit-learn: Machine Learning in Python." *Journal of Machine Learning Research*, 12, 2825–2830.
8. International Conference on Computer Vision and Image Processing (CVIP) - This conference is organized by the Indian Institute of Technology, Roorkee, and features papers on computer vision and image processing, including pose estimation and body language analysis.
9. The International Conference on Machine Learning and Applications (ICMLA), organized by the Indian Institute of Technology, Kanpur, features a wide array of research papers in the field of machine learning. Among the presented topics are advancements in pose estimation and body language analysis, offering valuable insights into these areas and their applications.
10. Workshop on Computer Vision and Robotics (WCVR) - This workshop is organized by the Indian Institute of Technology, Bombay, and features papers on computer vision and robotics, including pose estimation and body language analysis.

Authors

Neha Gupta is an assistant professor in Computer Science and Engineering Department of Moradabad Institute of Technology affiliated with Dr. A.P.J. Abdul Kalam Technical University. She had done B. Tech, M.Tech and now pursuing Ph. D. Her research area involves Machine learning, Deep Learning, Data Science and Data Security Measures.

Utkarsh Saxena is a B.Tech 3rd Year Student in Computer Science and Engineering Department of Moradabad Institute of Technology affiliated with Dr. A.P.J. Abdul Kalam Technical University. His research interests include Machine Learning, Deep learning and Video data processing.

Satyam Singh is a B.Tech 3rd Year Student in Computer Science and Engineering Department of Moradabad Institute of Technology affiliated with Dr. A.P.J. Abdul Kalam Technical University. His research interests include Machine Learning, Deep learning and Video data processing.