Journal of Artificial Intelligence and Engineering Applications

Website: https://ioinformatic.org/

15th October 2025. Vol. 5. No. 1; e-ISSN: 2808-4519

Detection of Student Focus Level using Body Posture Analysis with Blazepose

Adrian Michael Vincensius Purba[1]*, Hermawan Syahputra[2], Mansur As[3], Said Iskandar Al Idrus[4], Debi Yandra Niska [5]

[1,2,3,4,5]*Ilmu Komputer, Universitas Negeri Medan*
Sinagaromlah@gmail.com[1*]

Abstract

University student engagement monitoring is crucial for effective learning outcomes, yet traditional methods rely on subjective educator observations and time-consuming questionnaires. This research develops and evaluates a real-time university student focus monitoring system based on body posture analysis using the BlazePose pose estimation model with a rule-based classification approach. The system detects 33 body keypoints from video input using MediaPipe Pose, utilizing 7 specific keypoints including nose, left and right shoulders, left and right hips, and left and right knees to calculate head and body angles for classifying postures into three categories: Focused, Less Focused, and Not Focused. Testing was conducted on 10 video datasets with various sitting posture scenarios. The system successfully operates in real-time with a frame rate of 25-30 FPS and achieves an overall average accuracy of 81.05%, exceeding the minimum performance target of 80%. However, consistency remains challenging with only 60% of test videos reaching the target threshold. System performance proves excellent under ideal conditions such as adequate lighting and contrasting backgrounds, but decreases significantly under challenging conditions including poor lighting, noise, and ambiguous postures. The system successfully demonstrates proof of concept, yet requires further optimization to enhance robustness against environmental variations for practical classroom implementation.

Keywords: *BlazePose; Computer Vision; MediaPipe; Pose Estimation; Student Engagement*

1. Introduction

The development of information and communication technology (ICT) has brought significant transformations across various sectors, including education. In the educational domain, this technological progress enables the implementation of innovative approaches to enhance teaching quality and learning outcomes. One critical aspect receiving increasing attention is student engagement during classroom learning. Engagement is a fundamental element that greatly influences learning effectiveness, as students who are actively engaged tend to achieve better academic performance and deeper understanding of learning materials [1].

However, accurately measuring student engagement remains a challenge. Conventional methods—such as direct teacher observation or student questionnaires—are often subjective, time-consuming, and unsuitable for continuous monitoring. These methods are limited by personal biases and cannot be sustained throughout an entire lesson. Consequently, there is a growing need for objective and automated systems that can continuously and reliably assess student engagement, providing real-time data to inform instructional decisions [2].

One promising technological solution is the use of computer vision, particularly through pose estimation. Pose estimation is an image processing technique that identifies and tracks key points on the human body (e.g., head, shoulders, hips) from digital images or video. This technique enables non-invasive, real-time posture monitoring. A notable example is BlazePose, developed by Bazarevsky et al. [3][4], which can detect 33 body landmarks with high accuracy and low computational requirements—ideal for educational settings and classroom applications.

Several studies have confirmed a strong relationship between body posture and student engagement. Systematic reviews have shown that computer vision can effectively analyze classroom behaviors by recognizing visual features such as facial expressions, head orientation, and body movements [5]. Moreover, researchers have developed real-time applications for student behavior monitoring, demonstrating that body posture and other visual cues are reliable non-verbal indicators of focus and engagement [6].

Importantly, pose-estimation-based monitoring can be used in both in-person and remote learning environments. It respects students' privacy, operates without wearable devices, and offers real-time insights into focus or distraction levels. By integrating BlazePose with the MediaPipe framework [7], educators and researchers can build adaptive, data-driven learning environments that dynamically respond to students' physical behaviors and support improved educational outcomes.

2. Research Method

This research consists of six main processes to develop a student activity monitoring system using BlazePose technology (as illustrated in Figure 1):

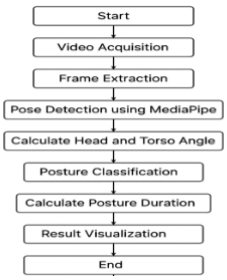

Fig. 1: Research workflows

2.1. Data Collection and Preprocessing

The initial stage of this research focused on collecting video data consisting of various student postures in classroom environments. Data collection was conducted in a controlled laboratory setting at the Mathematics Department, Universitas Negeri Medan.

2.1.1. Data Collection Setup

a. Location: Mathematics Laboratory, Universitas Negeri Medan
b. Recording Equipment: 1080p digital camera mounted on tripod
c. Perspective: Side view (left/right) for optimal landmark visibility
d. Subjects: Students in various sitting postures
e. Total Dataset: 10 videos with varying durations (60-80 seconds each)

2.1.2. Posture Categories

The system classifies student postures into six main categories based on head and body positions:

Table 1: Posture variations and focus status mapping

No.	Posture Description	Focus Status Category
1	Upright posture with upright head	Focused
2	Leaning posture with upright head	Focused
3	Upright posture with head looking down	Less Focused
4	Leaning posture with head looking down	Less Focused
5	Slouching posture with upright head	Not Focused
6	Slouching posture with head looking down	Not Focused

2.1.3. Data Preprocessing

The preprocessing stage involves several critical steps:
a. Frame Extraction: Videos segmented into individual frames for analysis
b. Resize: Frame dimensions optimized for computational efficiency
c. Color Normalization: BGR to RGB conversion for MediaPipe compatibility
d. Quality Control: Removal of frames with poor lighting or occlusion

Final Dataset Specifications:
a. Quantity: 10 selected videos with varying durations
b. Resolution: 720p to 1080p
c. Format: MP4 with 30 fps frame rate
d. Lighting Conditions: Varying from normal to low-light conditions
e. Subjects: Various body postures in sitting positions

2.2. MediaPipe BlazePose Implementation

This research utilizes MediaPipe Pose, a pose detection library developed by Google, to detect 33 human body landmarks in real-time. MediaPipe was chosen for its high accuracy in pose detection even under varying lighting conditions and efficient performance on devices with limited computational resources.

2.2.1. Model Initialization

The MediaPipe Pose model is initialized with the following parameters for video processing optimization:

```
mp_pose = mp.solutions.pose
pose = mp_pose.Pose(
    static_image_mode=False,
    min_detection_confidence=0.5,
    min_tracking_confidence=0.5
)
```

2.2.2. Key Landmark Extraction

The system uses 7 primary landmarks for posture analysis:

Table 2: MediaPipe landmark indices and functions

Landmark Name	Indonesian	BlazePose Index	Function
Nose	Hidung	0	Head direction reference
Left Shoulder	Bahu Kiri	11	Upper body reference (left)
Right Shoulder	Bahu Kanan	12	Upper body reference (right)
Left Hip	Pinggul Kiri	23	Middle body reference (left)
Right Hip	Pinggul Kanan	24	Middle body reference (right)
Left Knee	Lutut Kiri	25	Lower body reference (left)
Right Knee	Lutut Kanan	26	Lower body reference (right)

2.2.3. Side Selection Algorithm

The system automatically selects the body side (left or right) with the highest visibility value from landmarks to ensure analysis accuracy.

2.3. Angle Calculation Method

2.3.1. Mathematical Foundation: Dot Product Method

Posture analysis in computer vision systems requires mathematical approaches that can accurately measure angles between body segments. This research uses the dot product method as the basis for angle calculation due to its precision and computational efficiency.

The basic formula used in this research is:

$$\theta = \arccos((u \cdot v)/(|u| \cdot |v|)) \quad (1)$$

Where:

a. u is the vector from the center point (shoulder) to the first point (nose)
b. v is the vector from the center point (shoulder) to the third point (hip)
c. θ is the angle being calculated
d. $u \cdot v$ represents the dot product of both vectors
e. $|u|$ and $|v|$ are the magnitudes of respective vectors

2.3.2. Head Angle Calculation

Head angle is used to determine whether the head is in an upright or looking-down position. This angle is formed by three landmark points:

a. Point P1: Nose
b. Point P2: Shoulder (as center/vertex)
c. Point P3: Hip

Calculation Process

.1. Vector Formation: Two vectors are formed from the center point Shoulder (P2):

   a. Vector u from Shoulder to Nose: $u = P_1 - P_2$
   b. Vector v from Shoulder to Hip: $v = P_3 - P_2$

.2. Formula Application: Head Angle (θ_head) is calculated using the dot product formula

2.4. Rule-Based Classification System

After angles are calculated, the system uses a rule-based classification approach with predeterminedthresholds.

2.4.1. Classification Rules

Head Position Classification:

a. If head angle ≥ 125°: Head "Upright"
b. If head angle < 125°: Head "Looking Down"

Body Posture Classification:

a. If body angle ≥ 90°: Body "Upright"
b. If 70° ≤ body angle < 90°: Body "Leaning"
c. If body angle < 70°: Body "Slouching"

2.4.2. Final Status Mapping

The combination of head and body classifications is then mapped to final focus status:

Table 3: Focus status classification matrix

Head Position	Body Posture	Focus Status
Upright	Upright	Focused
Upright	Leaning	Focused
Looking Down	Upright	Less Focused
Looking Down	Leaning	Less Focused
Upright	Slouching	Not Focused
Looking Down	Slouching	Not Focused

2.5. System Architecture Design

2.5.1. System Flowchart

The system workflow consists of the following sequential processes:

1. Video Input: Real-time video stream or recorded video file
2. Frame Processing: Individual frame extraction and preprocessing
3. Landmark Detection: MediaPipe pose estimation for key body points
4. Angle Calculation: Mathematical computation of head and body angles
5. Classification: Rule-based categorization into focus status
6. Output Generation: Visual feedback and status reporting

2.5.2. Real-time Performance Requirements

a. Frame Rate: Target 25-30 FPS for real-time processing
b. Latency: < 33ms per frame for responsive feedback
c. Computational Efficiency: Optimized for standard hardware configurations

2.6. Evaluation Methodology

2.6.1. Target Accuracy Setting

A target accuracy of 80% was established as the system performance benchmark, considering thefollowing factors:

Technical Considerations:

a. Precision-Recall Balance: 80% target achieves optimal balance between precision and recall
b. Real-time System Tolerance: Accounts for noise, lighting variations, partial occlusion, andunpredictable subject movements
c. Computational Efficiency: Maintains optimal performance (25-30 FPS) without sacrificing speed
d. Computer Vision Standards: 80% accuracy is a reliable threshold in similar pose estimation research

Practical Considerations:

a. Real-world Applicability: System remains useful in actual conditions and provides meaningfulfeedback
b. User Experience: Sufficient accuracy without excessive false positives
c. Clinical Relevance: Adequate for identifying posture patterns at risk of musculoskeletal disorders

2.6.2. Performance Metrics

The following quantitative metrics are used to measure target achievement:

a. Accuracy: Percentage of correct classifications from total classifications
b. Precision: Percentage of correct positive classifications from total positive classifications
c. Recall: Percentage of correct positive classifications from total actual positive cases

d. F1-Score: Harmonic mean of precision and recall

2.6.3. Performance Categories

Results are grouped into four performance categories for qualitative analysis:

a. Excellent: 90% ≤ Accuracy ≤ 100% - System works exceptionally well
b. Good: 80% ≤ Accuracy < 90% - System meets expectations
c. Fair: 75% ≤ Accuracy < 80% - Performance approaching target, needs minor improvement
d. Poor: Accuracy < 75% - Requires significant improvement

3. Experiments and Analysis

3.1. Experimental Setup

This research was conducted using a controlled laboratory environment with standardized hardware andsoftware configurations to ensure consistent and reproducible results.

3.1.1. Hardware Configuration

The experimental setup utilized the following hardware specifications:

Table 4: Hardware specifications

Component	Specification
Laptop	AMD Ryzen 5 2500U, 12GB RAM, 256GB SSD
GPU	Radeon Vega Mobile Gfx 2.00 GHz
Webcam	1080p resolution digital camera
Tripod	Camera stabilization support

3.1.2. Software Environment

a. Operating System: Windows 10
b. Programming Language: Python 3.10.0
c. Libraries: OpenCV 4.5.4, MediaPipe 0.8.10, NumPy 1.21.4, Matplotlib 3.5.0, Pandas 1.3.4
d. IDE: Visual Studio Code

3.1.3. Testing Dataset

The system evaluation was conducted using 10 videos containing dynamic posture changes. Each framewas classified based on head and body angles, with results stored in CSV format for comparison withmanual labels.

3.2. Experimental Results and Analysis

3.2.1. Overall System Performance

The posture detection system, based on rule-based classification using head and body angles fromMediaPipe landmarks, achieved the following results:

Performance Summary:

a. Average Overall Accuracy: 81.65%
b. Accuracy Standard Deviation: 12.6%
c. Target Achievement Rate: 60% of videos met the 80% minimum threshold

3.2.2. Individual Video Performance Results

Table 5: Individual video accuracy results

Video	Duration (s)	Correct Detection	Incorrect Detection	Accuracy (%)	Status	Category
video1	63	63	0	100.00	✓	Excellent
video2	63	52	11	82.54	✓	Good
video3	63	62	1	98.41	✓	Excellent
video4	64	44	20	68.75	✗	Poor
video5	67	58	9	86.57	✓	Good
video6	67	41	26	61.19	✗	Poor
video7	64	49	15	76.56	✗	Fair
video8	71	65	6	91.55	✓	Excellent
video9	78	54	24	69.23	✗	Poor

video10	71	58	13	81.69	✓	Good

3.2.3. Performance Category Distribution

Table 6: Performance category distribution

Category	Number of Videos	Percentage	Description
Excellent (≥90%)	3	30%	Exceeds target significantly
Good (80–89%)	3	30%	Meets target requirements
Fair (75–79%)	1	10%	Approaches target
Poor (<75%)	3	30%	Below target threshold

The system successfully met the minimum accuracy target (80%) on 60% of test videos. However, 40% ofvideos remained below the threshold due to challenging environmental conditions and postureconfigurations.

3.2.4. Performance Category Distribution

Using confusion matrix analysis, the following performance metrics were obtained for classification:

Table 7: Classification performance metrics

Status	Precision	Recall	F1-Score
Focused	90.20%	90.00%	90.10%
Less Focused	81.10%	83.20%	82.10%
Not Focused	90.60%	88.70%	89.60%

Error Pattern Analysis: The most common classification errors occurred when distinguishing between"Focused" and "Less Focused" status. This typically happens when head or body positions are nearclassification thresholds (e.g., head_angle around 125° or body_angle around 90°). Minor movements ordetection noise can cause the system to switch status classification between "Focused" and "LessFocused" states.

3.3. Testing Models and Detailed Analysis

3.3.1. Real-time Performance Evaluation

Beyond classification accuracy, the system was evaluated for real-time processing performance on theutilized hardware:Performance Summary:

Real-time Performance Metrics:

a. Frame Rate: 25-30 FPS at 640x480 resolution
b. CPU Usage: 35% on AMD Ryzen 5 2500U
c. Memory Usage: Average 250MB
d. Latency: < 33ms per frame

These results demonstrate efficient computational performance suitable for real-time applications.

3.3.2. Detailed Case Analysis per Video

Excellent Category Analysis (video1, video3, video8):

a. Optimal Conditions: Even lighting, contrasting background, stable subject movement
b. High Landmark Visibility: Above 95% landmark detection rate
c. Minimal Noise: Very low environmental interference
d. Stable Transitions: Smooth posture changes without extreme movements

Good Category Analysis (video2, video5, video10):

a. Consistent Performance: Reliable and dependable system behavior
b. Adaptive Capability: Good adaptation to dynamic posture variations
c. Error Patterns: Errors mainly occur during rapid posture transitions or sudden posture changes

Fair Category Analysis (video7):

a. Near-target Performance: 76.56% accuracy approaching the 80% target
b. Environmental Challenges: Unstable lighting, landmark occlusion, challenging postures
c. Borderline Cases: Frequent postures near classification thresholds

Poor Category Analysis (video4, video6, video9):

a. Environmental Difficulties: Poor lighting, high noise, detection instability

b. High Error Rate: 61.19-69.23% accuracy due to challenging conditions
c. Landmark Issues: Frequent landmark loss or misdetection, especially during rapid movement or indark areas

3.3.3. System Limitations and Error Sources

The system successfully met the minimum accuracy target (80%) on 60% of test videos. However, 40% ofvideos remained below the threshold due to challenging environmental conditions and postureconfigurations.

Primary Error Sources Identified:

1. Threshold Sensitivity: Postures near classification boundaries causing unstable classifications
2. Lighting Variations: Poor or changing lighting conditions affecting landmark detection
3. Rapid Movements: Fast posture transitions exceeding system tracking capability
4. Partial Occlusion: Body parts hidden from camera view reducing landmark accuracy
5. Environmental Noise: Background interference affecting pose estimation quality

Robustness Analysis: While the system demonstrates solid proof-of-concept with effectivemethodology, particularly under ideal conditions (Excellent and Good categories), significant performancedegradation occurs under challenging environmental conditions. For reliable practical deployment, the system requires further optimization in filtering, threshold adjustment, and improved robustness againstlighting variations, occlusion, and diverse environmental conditions.

4. Conclusion

This study successfully demonstrated the application of rule-based classification in monitoring student activities based on body movement using BlazePose architecture. The MediaPipe BlazePose algorithm model showed the best performance. This model achieved a peak accuracy of 81.65%, while individual videos reached accuracy as high as 100.00%. This shows excellent capability in posture classification tasks. The analysis also revealed a significant difference in model performance, using geometric angle calculation with dot product method and rule-based thresholds produced a model accuracy that is worthy of being developed and utilized for the required purposes, indicating excellent model performance. This indicates the efficiency of landmark-based analysis, where the model is able to recognize posture patterns in real-time processing well without relying on complex computational resources. This approach is an effective and efficient choice, especially in the context of limited computational power and real-time processing requirements. This emphasizes the importance of optimizing the effectiveness of rule-based classification, where the MediaPipe BlazePose framework is proven to be the best choice to achieve high accuracy and low computational cost, but also minimize errors in posture prediction.

Acknowledgement

The authors sincerely appreciate all individuals and institutions who contributed to the successful completion of this study, particularly those who provided constructive feedback, technical assistance, and moral support throughout the research process

References

[1] J. A. Fredricks, P. C. Blumenfeld, and A. H. Paris, "School engagement: Potential of the concept, state of the evidence," Rev. Educ. Res., vol. 74, no. 1, pp. 59–109, 2004, doi: 10.3102/00346543074001059.
[2] Z. Sun, Q. Huang, H. Zhu, and D. Wu, "Classroom behavior recognition using computer vision: A systematic review," Sensors, vol. 25, no. 2, pp. 373, 2025, doi: 10.3390/s25020373.
[3] V. Bazarevsky, I. Grishchenko, Y. Kartynnik, A. Vakunov, M. Grundmann, and I. Tkachenka, "BlazePose: On-device real-time body pose tracking," Google AI Blog, Aug. 2020. [Online]. Available: https://ai.googleblog.com/2020/08/on-device-real-time-body-pose-tracking.html
[4] V. Bazarevsky and I. Grishchenko, "Real-time 3D body pose estimation with BlazePose," Google Research Blog, Mar. 2021. [Online]. Available: https://research.googleblog.com/2021/03/real-time-3d-body-pose-estimation-with.html
[5] Z. Sun, Q. Huang, H. Zhu, and D. Wu, "Classroom behavior recognition using computer vision: A systematic review," Sensors, vol. 25, no. 2, pp. 373, 2025, doi: 10.3390/s25020373.
[6] M. Tsiakmaki, G. Kostopoulos, S. Kotsiantis, and O. Ragos, "A computer-vision based application for student behavior monitoring in classroom," Appl. Sci., vol. 9, no. 22, pp. 4729, 2019, doi: 10.3390/app9224729.
[7] C. Lugaresi, J. Tang, H. Nash, C. McClanahan, E. Uboweja, M. Hays, F. Zhang, C. L. Chang, M. G. Yong, J. Lee, W. T. Chang, W. Hua, M. Georg, and M. Grundmann, "MediaPipe: A framework for building perception pipelines," arXiv preprint arXiv:1906.08172, 2019.F.-Z. Younsi, A. Bounnekar, D. Hamdadou, and O. Boussaid, “SEIR-SW, Simulation Model of Influenza Spread Based on the Small World Network,” *Tsinghua Sci. Technol.*, vol. 20, no. 5, pp. 460–473, 2015.