Available online at http://scik.org
Commun. Math. Biol. Neurosci. 2026, 2026:32
https://doi.org/10.28919/cmbn/9820
ISSN: 2052-2541

SKELETAL KEYPOINT-BASED PIPELINE AS A COMPUTER VISION-BASED APPROACHES

GABRIEL ASAEL TARIGAN[1,*], KUNCAHYO SETYO NUGROHO[2], BENS PARDAMEAN[2,3]

[1]Computer Science Department, School of Computer Science, Bina Nusantara University, Jakarta 11480, Indonesia

[2]Bioinformatics and Data Science Research Center, Bina Nusantara University, Jakarta 11480, Indonesia

[3]Computer Science Department, BINUS Graduate Program - Master of Computer Science, Bina Nusantara University, Jakarta 11480, Indonesia

Abstract: Classroom engagement analysis plays an important role in understanding students' learning behaviors in a non-intrusive manner. However, many existing computer vision-based approaches rely on complex deep learning architectures and time-consuming dataset construction, which limit their applicability in practical classroom settings. This study proposes a skeletal keypoint-based pipeline for classroom engagement analysis that combines person detection and single-person pose estimation. The extracted keypoints are represented in a low-dimensional form and used as input features for lightweight engagement classification models. Two classifiers, namely Ridge Classifier and Gradient Boosting, are evaluated to assess the effectiveness of the proposed pipeline. Experimental results on a test set of 432 samples show that the Ridge Classifier achieves an accuracy of 0.90 with a macro-average recall of 0.94. In contrast, Gradient Boosting achieves an accuracy of 0.96 and a macro-average precision of 0.98. In addition, the proposed pipeline significantly improves data collection efficiency, achieving approximately a sixteen-fold reduction in pose data acquisition time compared to conventional approaches. These results demonstrate that the proposed

---

[*]Corresponding author
E-mail address: gabriel.tarigan@binus.ac.id
Received February 13, 2026

pipeline provides a practical solution for individual-level classroom engagement analysis under controlled classroom.

Keywords: classroom engagement analysis; skeletal keypoints; pose-based classification; computer vision.

2020 AMS Subject Classification: 68T07, 68T45.

1. INTRODUCTION

Artificial Intelligence (AI) has increasingly permeated various aspects of daily life, with education among the most significantly impacted fields. AI now plays a central role in developing innovative methods that support and enhance learning processes. Education itself is widely regarded as a crucial indicator of global progress and societal development, with strong implications for individuals, institutions, and communities [1, 2]. In higher education, institutional performance indicators such as rankings, revenue, and reputation are closely linked to student retention, making the monitoring and evaluation of student engagement an important concern. Consequently, the application of machine learning techniques in education is not only beneficial but increasingly necessary [3, 4].

To ensure that learners effectively grasp instructional material, the interaction between teachers and students must be carefully observed and managed. Such observations enable informed educational decisions, allowing instructors to sustain effective learning practices or intervene when disengagement is detected. Non-verbal cues captured from classroom video data offer a practical and non-intrusive means of assessing student participation without disrupting the teaching–learning process [5]. However, a significant challenge is the scarcity of datasets that accurately reflect engagement levels in traditional, on-site classroom environments, which hampers the development and deployment of vision-based engagement analysis systems.

Existing research has explored a wide range of non-intrusive approaches for detecting student engagement using computer vision. Many studies focus on pose-based representations, in which human posture serves as an indicator of attention and participation [6, 7]. For example, prior work has employed person detection, skeleton-based pose estimation, and deep neural networks to classify student behaviors, demonstrating improvements over earlier skeleton-based methods [8]. Other approaches have combined OpenPose detection with manual annotation and rule-based classification derived from joint angles to recognize both individual and collective classroom behaviors [2]. These studies highlight the effectiveness of pose-based representations in capturing meaningful engagement-related cues.

More recent advancements have extended pose-based methods to multi-object and spatiotemporal settings. Some studies incorporate temporal semantics across video frames to improve body-part localization, often relying on object detectors such as YOLOv3 in conjunction with lightweight pose estimation networks [9]. Similar object-detection-driven approaches have been applied in smart campus environments to recognize student activities using deep neural network architectures [10]. To address challenges arising from low-resolution classroom imagery, researchers have also proposed enhanced detection models, such as upgraded Faster R-CNN architectures with locality-preserving loss functions [11] and multi-scale feature fusion strategies designed to handle object size variation and class imbalance [12].

In addition to pose-centric approaches, transfer learning and convolutional neural network (CNN)-based methods have been widely explored. For instance, VGG-based architectures have been used to classify engagement and disengagement from high-density classroom recordings [13]. At the same time, spatiotemporal representation learning has been shown to outperform benchmark three-dimensional CNN models for recognizing anomalous classroom behaviors [14]. Other studies have investigated covert analysis techniques that integrate non-verbal cues, including body posture and gestures, achieving moderate classification performance [15, 16]. Attention-enhanced YOLO-based models, such as YOLOv5s combined with Squeeze-and-Excitation mechanisms, have further demonstrated robustness in complex classroom backgrounds. Hybrid CNN architectures have also been proposed to infer both individual emotional states and overall classroom affective conditions [17].

Despite these advances, several practical limitations remain. Many existing approaches rely on computationally intensive models, require large volumes of manually annotated data, or focus primarily on maximizing classification accuracy rather than addressing end-to-end deployment considerations. In particular, primary data collection for classroom engagement analysis is often time-consuming and labor-intensive, reducing the adaptability of these methods to new learning environments. Moreover, some approaches continue to depend on earlier-generation object detection architectures, such as YOLOv3 or VGG-based models, which have since been surpassed in efficiency and performance by more recent alternatives [8–11, 18, 19].

Motivated by these challenges, this study focuses on the design and analysis of a skeletal keypoint-based pipeline for classroom engagement analysis. Rather than proposing a new pose estimation or detection algorithm, this work emphasizes the construction of a practical, modular pipeline that integrates object detection with skeletal keypoint extraction to support engagement analysis in classroom settings. The pipeline combines MediaPipe [20] for skeletal keypoint extraction with a YOLOv9-based [19] person detection stage, allowing individual-level pose analysis in classroom scenes containing multiple students. Engagement inference is performed using lightweight classification models, namely Ridge Classifier [21, 22] and Gradient Boosting [23–25], allowing an empirical comparison of their effectiveness when applied to low-dimensional skeletal representations.

Among the key contributions, this study provides an empirical evaluation of lightweight classification models applied to skeletal pose features, demonstrating their potential for practical engagement inference. It also offers insights into data collection strategies and pipeline design considerations, emphasizing the importance of scalable, efficient systems for deploying educational analytics in real classroom environments. These insights aim to bridge the gap between research and real-world application, addressing concerns about computational efficiency and deployment feasibility.

2. RESEARCH METHODOLOGY

Figure 1 presents an overview of the skeletal keypoint-based pipeline adopted for classroom engagement analysis in this study. The pipeline comprises three main stages: data collection and preprocessing, model training, and evaluation. Classroom video inputs are first processed to localize individual students in classroom scenes, after which skeletal keypoints are extracted using MediaPipe and represented as low-dimensional feature vectors. These features are subsequently used to classify engagement with lightweight machine learning models, and the overall pipeline is assessed using classification metrics, including accuracy, macro-average precision, and macro-average recall.

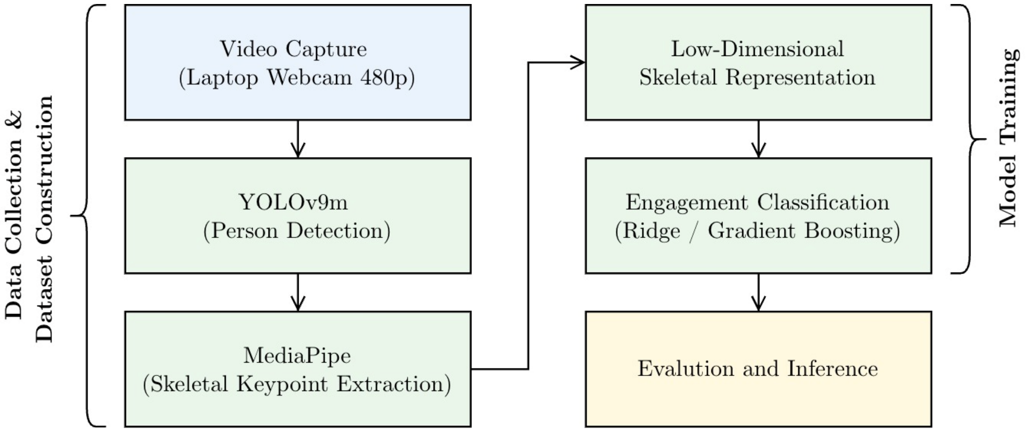


FIGURE 1. Proposed pipeline for classroom engagement analysis based on skeletal keypoints.

2.1. Data Collection and Dataset Construction. The data collection process captured human pose information in video format using a laptop webcam. Video recordings were obtained in a controlled classroom-like setting and processed at 480p resolution. Data acquisition was conducted across five recording sessions, with each session capturing approximately 500 video frames.

During the recording process, a volunteer performed a set of three predefined classroom-related poses, including *raise_hand*, *bored*, and *engaged*. For poses involving hand movement, namely *raise_hand* and *bored*, both right- and left-hand variations were recorded to capture natural variations in posture and movement.

The recorded video data were subsequently processed to construct a skeletal pose dataset. Each video frame was converted into a numerical representation based on extracted skeletal information, enabling further analysis and classification. The resulting dataset contains 8,100 instances, evenly distributed across three engagement-related classes: 2,700 for *raise_hand*, 2,700 for *bored*, and 2,700 for *engaged*. This balanced class distribution was intentionally maintained to support fair training and evaluation of the classification models.

2.2. Skeletal Keypoint Representation. Skeletal keypoint representation was obtained using

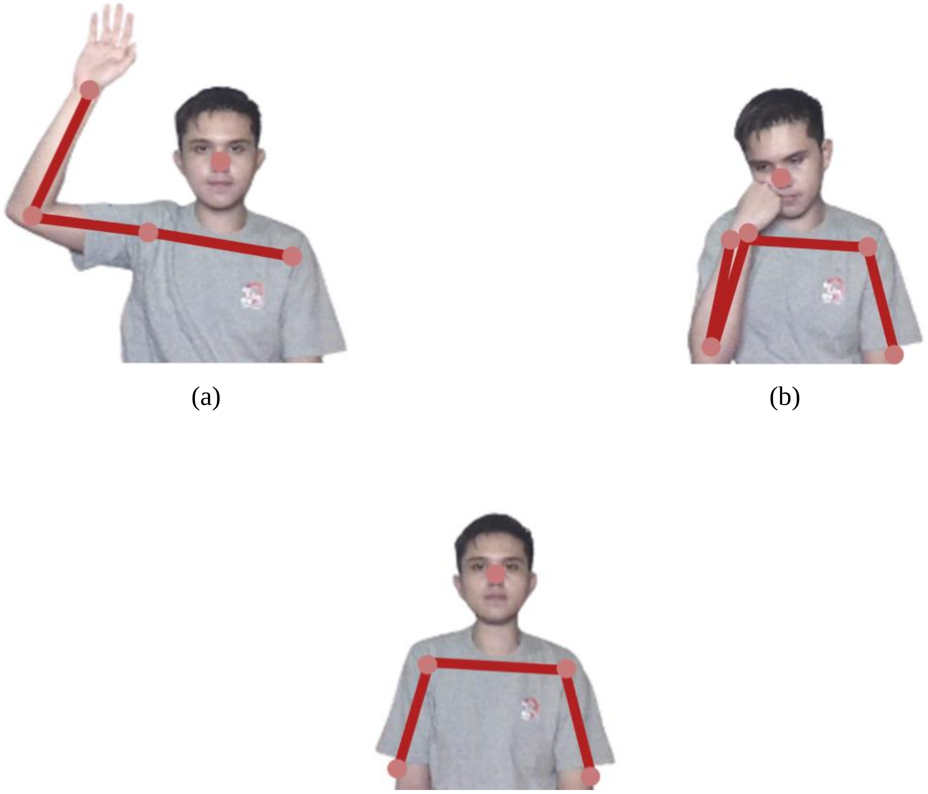
MediaPipe, which was employed to extract body keypoint coordinates from each video frame. MediaPipe provides a set of anatomical landmarks corresponding to major body joints, which can be used to describe human posture in a compact, structured manner. Since MediaPipe is designed to operate on a single detected person, an object detection stage was applied prior to keypoint extraction to enable processing of multi-person classroom scenarios.

Following keypoint extraction, a filtering process was applied to retain only the keypoints that contribute most significantly to engagement-related pose recognition. This filtering step focuses on upper-body joints that are commonly associated with classroom engagement behaviors. The selected keypoints include the nose, shoulders, elbow, and wrist, providing a concise representation of posture while reducing feature dimensionality.

Each pose instance is represented by 14 numerical features, corresponding to the x- and y-coordinates of seven selected skeletal keypoints. An example of this coordinate-based representation is presented in Table 1. Figure 2 provides illustrative examples of classroom postures with overlaid skeletal keypoints extracted using the proposed pipeline. The visualizations highlight how key anatomical points are captured to represent different engagement-related postures in a non-intrusive manner. The use of a low-dimensional skeletal representation enables efficient utilization of pose information as input features for subsequent engagement classification.

TABLE 1. Example of low-dimensional skeletal keypoint features extracted from classroom pose data.

class	$x0$	$y0$	**…**	$y16$
raise_hand	0.5035	0.3465	…	0.2615
raise_hand	0.5039	0.3490	…	0.2603
…	…	…	…	…
engaged	0.4727	0.3565	…	0.9087



(a) (b) (c)

FIGURE 2. Representative examples of classroom engagement postures with overlaid skeletal keypoints: (a) *raise-hand*, (b) *bored*, and (c) *engaged*.

2.3. Data Preprocessing. Prior to model training, data preprocessing was performed to prepare the dataset for classification. The dataset was first shuffled to reduce potential ordering effects and to improve the statistical robustness of the learning process [26]. Following shuffling, the dataset was partitioned into training and test subsets. An 80:20 split was adopted, with 80% of the data used for training and 20% for test. This partitioning strategy supports reliable model evaluation and generalization performance assessment, particularly for datasets of moderate size [27].

2.4. Engagement Classification Models. Two classification models were selected to infer engagement-related states from the extracted skeletal keypoint features, namely the Ridge Classifier and Gradient Boosting. Both models were trained and evaluated using the same dataset to allow a consistent comparison of their performance within the proposed pipeline.

2.4.1. *Ridge Classifier*. The Ridge Classifier was selected for its suitability for supervised classification tasks with low- to moderate-dimensional feature representations. By incorporating L2 regularization, ridge-based models can mitigate overfitting and provide stable solutions when working with limited or imbalanced datasets. In addition, ridge classifiers offer a closed-form solution, making them computationally efficient for training and evaluation. Formally, given a feature vector $x \in \mathbb{R}^d$, the Ridge Classifier predicts the class label using a linear decision function, as defined in Equations (1):

$$\hat{y} = w^{\top}x + b \tag{1}$$

Where w denotes the weight vector and b represents the bias term. The model parameters are obtained by minimizing least-squares objective, as defined in Equations (2):

$$L(w) = \sum_{i=1}^{N} (y_i - w^{\top}x_i)^2 + \lambda \parallel w \parallel_2^2 \tag{2}$$

Where λ controls the strength of the L2 regularization. Previous studies have demonstrated the effectiveness of ridge-based classifiers in handling high-dimensional and imbalanced data scenarios [21, 22].

2.4.2. *Gradient Boosting*. Gradient Boosting was employed as an alternative classification approach to model non-linear relationships within the skeletal keypoint features. Ensemble-based gradient boosting methods are known for their ability to handle complex decision boundaries and to perform effectively in datasets with class imbalance. While performance may vary across implementations, gradient boosting techniques have been widely adopted for classification tasks due to their flexibility and robustness [24, 25]. In Gradient Boosting, the final prediction model is constructed as an additive ensemble of weak learners, as defined in Equation (3):

$$F(x) = \sum_{m=1}^{M} \alpha_m f_m(x) \tag{3}$$

where $f_m(x)$ denotes the m-th weak learner and α_m is its corresponding weighting factor. Each weak learner is trained sequentially to minimize the loss function by fitting the residuals of the previous ensemble, as defined in Equation (4):

$$F_m(x) = F_{m-1}(x) + \alpha_m f_m(x) \tag{4}$$

This iterative optimization process enables Gradient Boosting to capture non-linear patterns in the skeletal keypoint features and to adapt effectively to imbalanced data distributions.

2.5. Model Evaluation Metrics and Inference. Inference in the proposed pipeline follows a two-stage process. First, YOLOv9m is used to detect students in classroom video frames and generate bounding boxes for each individual. Skeletal poses are then inferred for each detected student using MediaPipe, and the resulting keypoint representations are used for engagement classification. Evaluation was conducted on a test set comprising 432 data instances, all captured under consistent classroom conditions and recorded at 480p resolution. Model performance was assessed using accuracy, macro-average precision, and macro-average recall, as defined in Equations (5)–(7):

$$Accuracy = \frac{1}{N}\sum_{i=1}^{N} \mathbb{I}\left(\hat{y}_i = y_i\right) \tag{5}$$

$$Macro - Precision = \frac{1}{K}\sum_{k=1}^{K} Precision_k \tag{6}$$

$$Macro - Recall = \frac{1}{K}\sum_{k=1}^{K} Recall_k \tag{7}$$

In addition, confusion matrices were employed to analyze class-wise classification behavior and misclassification patterns.

3. RESULTS AND DISCUSSION

The experimental results reflect the classification performance of the proposed skeletal keypoint-based pipeline under controlled classroom conditions. One notable observation is the reduced time required for data collection and pose extraction. Each recording session lasted approximately 2 minutes and 30 seconds, resulting in a total data acquisition time of around 12 minutes for the complete dataset. This reduced collection time indicates the practicality of the proposed pipeline for constructing skeletal pose datasets in classroom-like environments, where minimizing manual intervention is an important consideration.

Figure 3 shows the classification outcomes obtained using the Ridge Classifier across the engagement-related classes. The model achieved perfect classification for the *bored* class, correctly identifying all 191 instances. In the *engaged* class, 121 of 163 instances were correctly classified, resulting in 42 misclassifications. The *raise_hand* class was also classified with high accuracy, with 77 correct predictions and a single misclassification out of 78 instances. These results suggest that the Ridge Classifier performs consistently for posture patterns associated with boredom and hand-raising behaviors, while engagement-related poses exhibit greater variability.

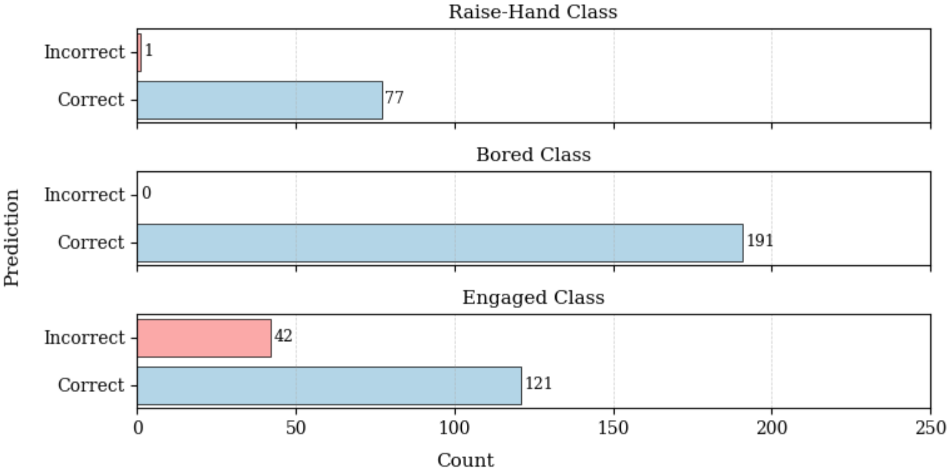


FIGURE 3. Prediction outcomes of the Ridge Classifier across engagement classes, showing correct and incorrect predictions.

FIGURE 4. Prediction outcomes of the Gradient Boosting across engagement classes, showing correct and incorrect predictions.

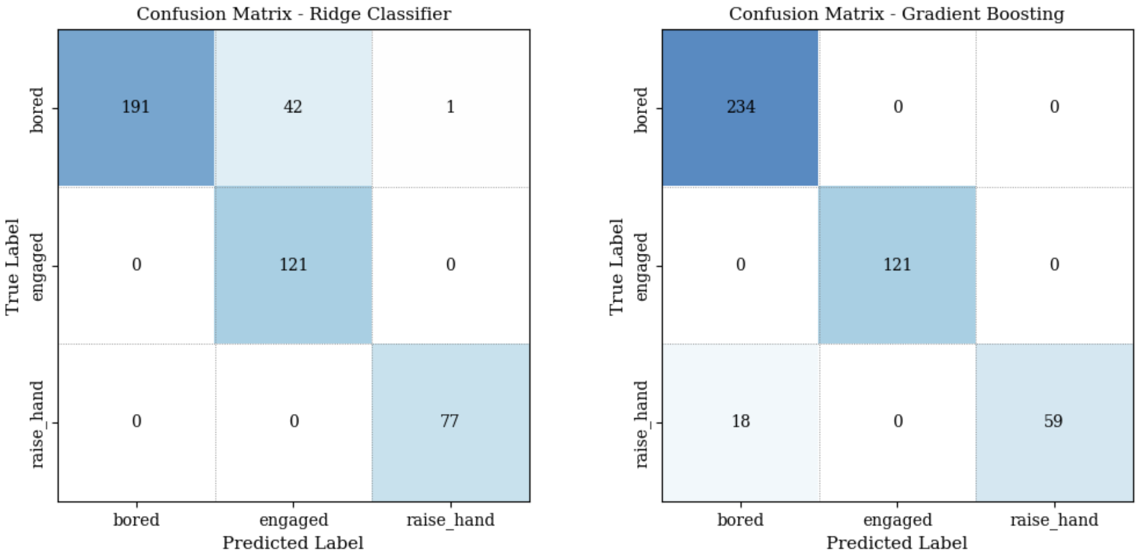
Figure 4 presents the classification results obtained using the Gradient Boosting model. The classifier achieved perfect prediction for both *engaged* and *raise_hand* classes, correctly identifying all evaluated instances in these categories. However, its performance on the *bored* class was slightly lower, with 234 correct classifications out of 252 instances and 18 misclassifications. This behavior indicates that ensemble-based models are effective at capturing non-linear patterns associated with active engagement while remaining sensitive to posture variations related to boredom.

A quantitative summary of classification performance is provided in Table 2, which reports the evaluation metrics computed on the 432 test instances. Both classifiers achieved high overall performance. Gradient Boosting achieved a higher accuracy (0.96) than the Ridge Classifier (0.90). In contrast, the Ridge Classifier achieved a higher macro-average recall (0.94) than Gradient Boosting (0.92), indicating a more balanced sensitivity across engagement classes. These results highlight a trade-off between maximizing overall classification accuracy and maintaining class-wise sensitivity, depending on the chosen model.

TABLE 2. Classification performance of the Ridge Classifier and Gradient Boosting evaluated on the test dataset.

	Ridge Classifier	**Gradient Boosting**
Accuracy	0.90	0.96
Macro-average Precision	0.91	0.98
Macro-average Recall	0.94	0.92

Further insight into class-wise behavior is illustrated in Figure 5, which highlights distinct misclassification patterns between the two classifiers. The Ridge Classifier shows strong performance in identifying *bored* and *raise_hand* classes, with most misclassifications occurring between *bored* and *engaged*. This suggests that engagement-related postures may be more challenging to model using linear decision boundaries. In contrast, the Gradient Boosting classifier exhibits more consistent predictions for *engaged* and *raise_hand* classes, while showing a limited number of misclassifications in the bored category. These patterns reinforce the ability of ensemble-based models to capture non-linear variations in skeletal posture features, while linear models exhibit more stable behavior across certain posture categories.



(a) (b)

FIGURE 5. Confusion matrices of classroom engagement classification using (a) the Ridge Classifier and (b) the Gradient Boosting classifier.

Overall, the results indicate that lightweight classification models can effectively utilize skeletal keypoint representations for classroom engagement analysis. While Gradient Boosting provides higher overall accuracy, the Ridge Classifier offers competitive performance with more balanced recall across classes. These findings suggest that classifier selection may be guided by application-specific priorities, such as prioritizing overall accuracy or sensitivity to particular engagement behaviors. The results also demonstrate the feasibility of the proposed pipeline under controlled classroom conditions, while acknowledging that further evaluation across a broader range of subjects and environments is necessary to assess generalizability.

4. CONCLUSION

This study investigated a skeletal keypoint-based pipeline for classroom engagement analysis, integrating MediaPipe for pose estimation with YOLOv9m for person detection. Experimental results on a test set of 432 samples show that the proposed pipeline effectively supports engagement classification using lightweight models. The Ridge Classifier achieved an accuracy of 0.90 with balanced class-wise sensitivity (macro-average recall of 0.94), while Gradient Boosting achieved higher overall performance with an accuracy of 0.96 and a macro-average precision of 0.98, highlighting a trade-off between overall accuracy and recall balance [28–32].

Beyond classification performance, the proposed pipeline significantly improves data acquisition efficiency by reducing the time required for pose extraction and dataset construction in controlled classroom settings. This efficiency supports practical deployment scenarios where minimizing manual effort and computational complexity is essential. From an applied perspective, the results indicate that low-dimensional skeletal keypoint representations, combined with lightweight classifiers, provide a feasible foundation for classroom engagement analysis without relying on complex deep learning architectures.

Several directions for future work can be identified. Improving robustness under challenging visual conditions, such as motion blur and occlusion, remains an important area for further investigation. Extending the pipeline to support a broader range of engagement-related poses and multi-pose detection may further enhance its applicability. In addition, exploring hybrid classification strategies that combine the strengths of different learning models could provide a more balanced solution for engagement analysis in diverse classroom environments.

AUTHOR CONTRIBUTIONS

Gabriel Asael Tarigan: Conceptualization, Methodology, Formal analysis, Software, Writing – Original Draft. Kuncahyo Setyo Nugroho: Formal analysis, Validation, Writing – Review & Editing. Bens Pardamean: Validation, Supervision.

DATA AVAILABILITY

The video data used in this study are not publicly available due to privacy considerations involving human participants. However, the extracted skeletal keypoint data supporting the findings of this study are available for download at: https://github.com/Zayphen/pose_coordinate.

CONFLICT OF INTERESTS

The authors declare that there is no conflict of interest.

REFERENCES

[1] Z. MacHardy, K. Syharath, P. Dewan, Engagement Analysis Through Computer Vision, in: Proceedings of the 8th IEEE International Conference on Collaborative Computing: Networking, Applications and Worksharing, IEEE, 2012. https://doi.org/10.4108/icst.collaboratecom.2012.250429.

[2] P. Vanneste, J. Oramas, T. Verelst, T. Tuytelaars, A. Raes, et al., Computer Vision and Human Behaviour, Emotion and Cognition Detection: A Use Case on Student Engagement, Mathematics 9 (2021), 287. https://doi.org/10.3390/math9030287.

[3] B. Pardamean, Measuring Change in Critical Thinking Skills of Dental Students Educated in a PBL Curriculum, J. Dent. Educ. 76 (2012), 443-453. https://doi.org/10.1002/j.0022-0337.2012.76.4.tb05276.x.

[4] D. Hao, Y. Xiaoqi, Q. Taoyu, Hybrid Machine Learning Models Based on CATBoost Classifier for Assessing Students' Academic Performance, Int. J. Adv. Comput. Sci. Appl. 15 (2024), 94-106. https://doi.org/10.14569/IJACSA.2024.0150709.

[5] T. S. Ashwin, R.M.R. Guddeti, Automatic Detection of Students' Affective States in Classroom Environment Using Hybrid Convolutional Neural Networks, Educ. Inf. Technol. 25 (2019), 1387-1415. https://doi.org/10.1007/s10639-019-10004-6.

[6] G.A. Tarigan, G. Natanael Elwirehardja, K.S. Nugroho, B. Pardamean, Computer Vision-Based Pose Estimation for Student Engagement Detection: Trends and Challenges, in: 2025 7th International Conference on Cybernetics and Intelligent System (ICORIS), IEEE, 2025, pp. 1-5. https://doi.org/10.1109/ICORIS67789.2025.11296082.

[7] G.A. Tarigan, G.N. Elwirehardja, K.S. Nugroho, B. Pardamean, Lighter Student Engagement Recognition in a Classroom Environment Using Skeletal Keypoints, IAENG Int. J. Comput. Sci. 52 (2025), 1997-2014.

[8] F.C. Lin, H.H. Ngo, C.R. Dow, K.H. Lam, H.L. Le, Student Behavior Recognition System for the Classroom Environment Based on Skeleton Pose Estimation and Person Detection, Sensors 21 (2021), 5314. https://doi.org/10.3390/s21165314.

[9] J. Liu, X. Mu, Z. Liu, H. Li, Human Skeleton Behavior Recognition Model Based on Multi-Object Pose Estimation with Spatiotemporal Semantics, Mach. Vis. Appl. 34 (2023), 44. https://doi.org/10.1007/s00138-023-01396-0.

[10] M. Rashmi, T.S. Ashwin, R.M.R. Guddeti, Surveillance Video Analysis for Student Action Recognition and Localization Inside Computer Laboratories of a Smart Campus, Multimed. Tools Appl. 80 (2020), 2907-2929. https://doi.org/10.1007/s11042-020-09741-5.

[11] Tang, L., Gao, C., Chen, X., Zhao, Y., Pose detection in complex classroom environment based on improved Faster R-CNN, IET Image Processing. 13 (2019), 451–457. https://doi.org/10.1049/iet-ipr.2018.5905.

[12] C. Gao, S. Ye, H. Tian, Y. Yan, Multi-Scale Single-Stage Pose Detection with Adaptive Sample Training in the Classroom Scene, Knowl.-Based Syst. 222 (2021), 107008. https://doi.org/10.1016/j.knosys.2021.107008.

[13] H.T. Binh, N. Quang Trung, H.A.T. Nguyen, B. The Duy, Detecting Student Engagement in Classrooms for Intelligent Tutoring Systems, in: 2019 23rd International Computer Science and Engineering Conference (ICSEC), IEEE, 2019, pp. 145-149. https://doi.org/10.1109/ICSEC47112.2019.8974739.

[14] Y. Xie, S. Zhang, Y. Liu, Abnormal Behavior Recognition in Classroom Pose Estimation of College Students Based on Spatiotemporal Representation Learning, Trait. Du Signal 38 (2021), 89-95. https://doi.org/10.18280/ts.380109.

[15] J. Terven, D.M. Córdova-Esparza, J.A. Romero-González, A Comprehensive Review of YOLO Architectures in Computer Vision: From YOLOv1 to YOLOv8 and Yolo-Nas, Mach. Learn. Knowl. Extr. 5 (2023), 1680-1716. https://doi.org/10.3390/make5040083.

[16] J.L. Chung, L.Y. Ong, M.C. Leow, Comparative Analysis of Skeleton-Based Human Pose Estimation, Futur. Internet 14 (2022), 380. https://doi.org/10.3390/fi14120380.

[17] T.S. Ashwin, R.M.R. Guddeti, Affective Database for E-Learning and Classroom Environments Using Indian Students' Faces, Hand Gestures and Body Postures, Futur. Gener. Comput. Syst. 108 (2020), 334-348. https://doi.org/10.1016/j.future.2020.02.075.

[18] Z. Wang, J. Yao, C. Zeng, W. Wu, H. Xu, et al., Learning Behavior Recognition in Smart Classroom with Multiple Students Based on YOLOv5, arXiv:2303.10916, 2023. https://doi.org/10.48550/arXiv.2303.10916.

[19] C.Y. Wang, I.H. Yeh, H.Y. M. Liao, YOLOv9: Learning What You Want to Learn Using Programmable Gradient Information, in: A. Leonardis, E. Ricci, S. Roth, O. Russakovsky, T. Sattler, G. Varol, (eds) Computer Vision – ECCV 2024. ECCV 2024. Lecture Notes in Computer Science, vol 15089. Springer, Cham, 2025. https://doi.org/10.1007/978-3-031-72751-1_1.

[20] C. Lugaresi, J. Tang, H. Nash, C. McClanahan, E. Uboweja, et al., MediaPipe: A Framework for Building Perception Pipelines, arXiv:1906.08172, 2019. https://doi.org/10.48550/arXiv.1906.08172.

[21] C. Peng, Q. Cheng, Discriminative Ridge Machine: A Classifier for High-Dimensional Data or Imbalanced Data, IEEE Trans. Neural Netw. Learn. Syst. 32 (2021), 2595-2609. https://doi.org/10.1109/TNNLS.2020.3006877.

[22] D. Arpit, S. Wu, P. Natarajan, R. Prasad, P. Natarajan, Ridge Regression Based Classifiers for Large Scale Class Imbalanced Datasets, in: 2013 IEEE Workshop on Applications of Computer Vision (WACV), IEEE, 2013, pp. 267-274. https://doi.org/10.1109/WACV.2013.6475028.

[23] Z. He, D. Lin, T. Lau, M. Wu, Gradient Boosting Machine: A Survey, arXiv:1908.06951, 2019. https://doi.org/10.48550/arXiv.1908.06951.

[24] C. Bentéjac, A. Csörgő, G. Martínez-Muñoz, A Comparative Analysis of Gradient Boosting Algorithms, Artif. Intell. Rev. 54 (2020), 1937-1967. https://doi.org/10.1007/s10462-020-09896-5.

[25] M.H.L. Louk, B.A. Tama, Revisiting Gradient Boosting-Based Approaches for Learning Imbalanced Data: A Case of Anomaly Detection on Power Grids, Big Data Cogn. Comput. 6 (2022), 41. https://doi.org/10.3390/bdcc6020041.

[26] K. Lee, M. Lam, R. Pedarsani, D. Papailiopoulos, K. Ramchandran, Speeding up Distributed Machine Learning Using Codes, IEEE Trans. Inf. Theory 64 (2018), 1514-1529. https://doi.org/10.1109/TIT.2017.2736066.

[27] A. Rácz, D. Bajusz, K. Héberger, Effect of Dataset Size and Train/Test Split Ratios in QSAR/QSPR Multiclass Classification, Molecules 26 (2021), 1111. https://doi.org/10.3390/molecules26041111.

[28] R.E. Caraka, S.A. Bakar, B. Pardamean, A. Budiarto, Hybrid Support Vector Regression in Electric Load during National Holiday Season, in: 2017 International Conference on Innovative and Creative Information Technology (ICITech), IEEE, 2017, pp. 1-6. https://doi.org/10.1109/INNOCIT.2017.8319127.

[29] T.W. Cenggoro, A. Budiarto, R. Rahutomo, B. Pardamean, Information System Design for Deep Learning Based Plant Counting Automation, in: 2018 Indonesian Association for Pattern Recognition International Conference (INAPR), IEEE, 2018, pp. 329-332. https://doi.org/10.1109/INAPR.2018.8627019.

[30] R.E. Caraka, S. Shohaimi, I.D. Kurniawan, R. Herliansyah, A. Budiarto, et al., Ecological Show Cave and Wild Cave: Negative Binomial Gllvm’s Arthropod Community Modelling, Procedia Comput. Sci. 135 (2018), 377-384. https://doi.org/10.1016/j.procs.2018.08.188.

[31] M.F. Kacamarga, B. Pardamean, H. Wijaya, Lightweight Virtualization in Cloud Computing for Research, in: R. Intan, C.-H. Chi, H.N. Palit, L.W. Santoso (Eds.), Intelligence in the Era of Big Data, Springer, Berlin, Heidelberg, 2015: pp. 439–445. https://doi.org/10.1007/978-3-662-46742-8_40.

[32] R.E. Caraka, R.C. Chen, H. Yasin, S. Suhartono, Y. Lee, et al., Hybrid Vector Autoregression Feedforward Neural Network with Genetic Algorithm Model for Forecasting Space-Time Pollution Data, Indones. J. Sci. Technol. 6 (2021), 243-266. https://doi.org/10.17509/ijost.v6i1.32732.