Pakistan Journal of Scientific Research
Multidisciplinary & Peer Reviewed
Volume: 5, Number: 1, Pages: 1-10, Year:2025

Article

A Framework for Statistical Analysis of Student Engagement in OBE-based Classrooms Environment Using YOLO Model

Aqsa Sarfraz Khan[1,*], Saima Jamil[2], Waqar Azeem[1], Aftab Ahmad Malik[3], Noureen Riaz[1], Farwa Batool[4]

[1] Department of Software Engineering, Lahore Garrison University, Lahore, Pakistan
[2] Department of Computer Science and Information Technology, Virtual University of Pakistan, Pakistan
[3] Department of Criminology and Forensics, Lahore Garrison University, Lahore, Pakistan
[4] Department of Chemistry & Chemical Engineering Syed Babar Ali School of Science and Engineering, Lahore University of Management Sciences, Lahore, Pakistan

Corresponding author: Aqsa Sarfraz Khan (e-mail: aqsasarfraz@lgu.edu.pk).

Submitted: 10-03-2025**, Revised:** 20-05-2025, **Accepted:** 15-06-2025

Abstract

The most important aspect of effective learning in an Outcome-Based Education (OBE) classroom environment is student involvement and attentiveness during the lecture. This study examines various student behaviours and evaluates the attention of students during lectures using deep learning techniques, YOLO v8. The classroom representative visuals were perceived from an online available dataset containing students' images that have been recorded during the lecture. These photos are subjected to YOLO, which classifies different student activities into those that contribute positively or negatively to attention in the classroom. Positive markers of attention include actions like raising hands, concentrating on the front, reading, writing, and interacting with the teacher. On the other hand, distractions like eating, drinking, using a phone, or seeming drowsy have a detrimental impact. The results assist teachers in enhancing their teaching methodologies and offer insights into patterns of classroom involvement.

Keywords: Outcome-based education (OBE), YOLO (You Only Look Once), convolutional neural networks (CNNs), machine learning (ML), deep learning (DL), contextual attention (CA).

1. Introduction

Effective learning depends greatly on student participation, especially in an Outcome-Based Education (OBE) paradigm, wherein ensuring that students meet specified learning objectives is the main focus. Students who are actively involved are more likely to participate, retain knowledge well, and achieve better academically. Accurately gauging participation in live classroom environments is still difficult, though. Similarly, students who are involved in non-attention activities during the lectures will be unlikely to perform well [1].

The detection of actions like raising hands, concentrating on the front, reading, writing, and interacting with the teacher shall contribute to the percentage of classes that are positively engaged and attentive during class. On the other hand, the detection of distractions like eating, drinking, using a phone, or seeming drowsy will have a detrimental contribution to the calculation of the percentage of students who are not actively engaged with class activity [2].

Instructor comments, self-reported questionnaires, and manual observations are the mainstays of traditional engagement evaluation methods. These strategies are subjective, time-consuming, and not adaptable. The development of computer vision and artificial intelligence (AI) unveils an achievable approach to improve and automate engagement analysis. Classroom behaviour analysis is a good fit for deep learning models, especially object identification frameworks like YOLO (You Only Look Once), which have shown great performance in real-time detection applications [3, 4]. The current research demonstrates a comprehensive framework for analyzing student participation in a classroom setting that makes use of YOLO v8.

The model analyzes different student behaviours linked to engagement or distraction by converting lecture footage into representative image frames. The main goal was to establish a clear, scalable system that gives teachers real-time information about how attentive their students are, which can help them improve their teaching methods and adopt counselling for the student group who were classified as unattentive.

This paper is organized in the following structure: In Section 2, relevant research on CNN and deep learning-based methods for analyzing student behaviour is reviewed. Section 3 covers data collection, preprocessing, and model implementation. The experimental results and analysis are presented in Section 4. While conclusions and possible enhancements are discussed in Section 5.

2. Materials and Methods

Prior studies have investigated a number of methods to gauge students' attention, including gaze tracking [5] and EEG-based monitoring [6]. They require specialized tools, and these methods are less practical for large-scale classroom environments. Despite they have good accuracy, these are not very suitable for automated statistical analysis. Our study overcomes this gap through a proposed framework and solution based on computer vision and deep learning.

A) Computer Vision for Classroom Analysis

The computer vision-based methods for analyzing student behaviour and participation in classrooms have been investigated in various research articles. For example, [7] Zhang et al. proposed a method using posture analysis and facial recognition to detect and identify the attentiveness of students during both in-person and virtual classroom environments. Similarly, Singh et al. (2020) presented a method for tracking students' activities using deep learning models to find patterns in behaviours that are linked to distractions and engagement during the lecture [8].

Alkhateeb et al. (2024) recently proposed a technique that was based on deep learning, which uses the convolutional neural networks (CNNs) to predict academic achievement in higher education. For improvement in the forecast accuracy, that study utilized an oversampling and undersampling approach to address the issues of the class imbalance. Their methodology proved to be highly effective in the prediction of students' achievements using a comprehensive dataset from the University of Jordan that includes academic & course-related data, as well as many demographic variables. The results could present an analysis to the stakeholders in higher education, some substantially useful insights that could assist them in designing data-driven plans for the improvement of students' performance in multicultural classroom environments [9].

Rahman et al. in 2024 carried out a detailed review of the recent uses of deep learning (DL) methodologies for the prediction and analysis of student performance. They have examined diverse research that utilized the modern techniques, e.g. including deep learning and deep learning with conventional machine learning (ML) techniques, as well as those that have used ML and DL alone. Their review showed how well DL models can handle the high-dimensional and complicated educational data for enhanced prediction accuracy. The analysis also highlighted various difficulties and the need for sizable datasets for the training of DL models. They also made inferences that resolving these issues would enhance the process of DL-based strategies to promote educational interaction and tailored learning environments [10].

1.1. YOLO-Based Student Behaviour Detection

The YOLO model has become an extensively utilized technique for educational analytics and other real-time object identification and classification applications [11]. Li et al. (2022) demonstrated the usefulness and effective detection process of YOLO v7 to detect the on-task and off-task activities among students' interactions. They also presented the improved accuracy and computing efficiency for a similar process using the YOLO v9 model [12].

Wang et al. [13] proposed a Student Behaviour Detection (SBD) technique for the implementation of effective and consistent behavioural analysis in challenging classroom environments. They presented a model using YOLOv5 that also combines the Open Pose as well Contextual Attention (CA) mechanism. This method tries to improve the quality of learning approaches by analyzing real-time student behaviours in the classroom during the lecture [13].

Yang et al. [14] have presented an improved YOLOv7 model that also incorporates a Wise-IoU loss function along with a biformer attention module to enhance the detection accuracy of student behaviours for multiple classifications, including writing, reading, and raising their hands in crowded classroom setups. Compared to a mean Average Precision of 0.5 of 79% on the SCB-Dataset, their model has outperformed the prior findings by 1.8% [14].

In contrast to previous research, this study presents a framework for student behaviour analysis by integrating YOLO v8 with a consideration of activity classification based on OBE-based classroom environments. The identification of signs of interest and distractions would suggest that the teachers' quantitative and qualitative approach requires an improvement in teaching and learning methodologies for the better engagement of students during the lecture. Figure 1 shows a comprehensive framework for the implementation of students' behaviour analysis.

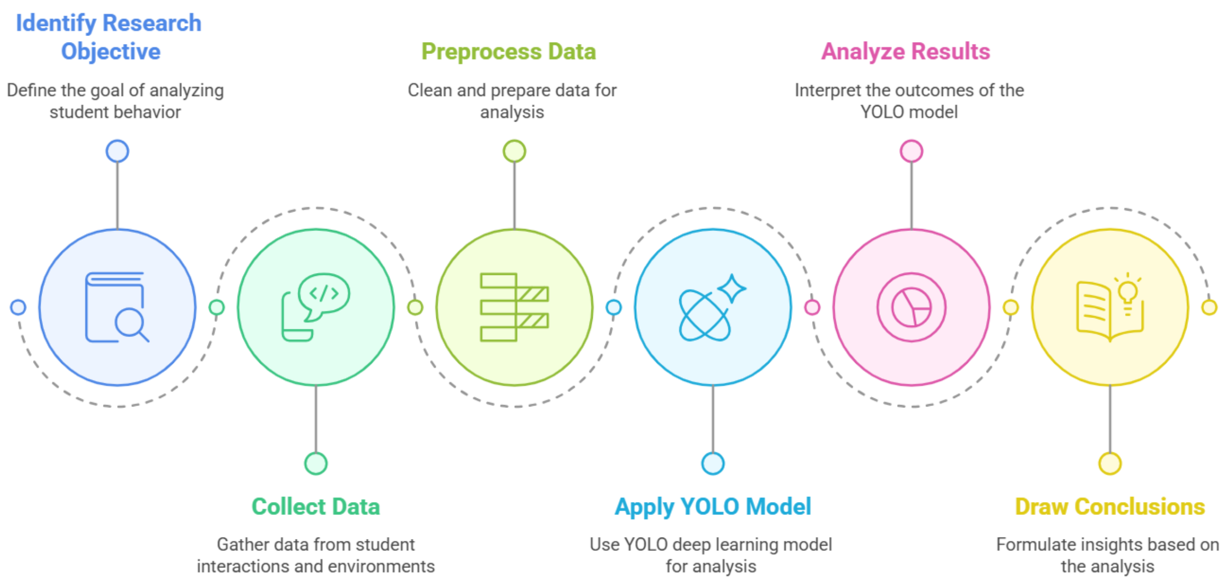

Figure 1. Framework for Students Behavior Analysis.

1.2. Dataset

For the implementation of this study, the publicly available Dataset of Student Classroom Behaviour from Kaggle was selected. The dataset consists of over 7995 images annotated with 8006 labels, which were captured in classroom environments that have been labelled with various annotations representing multiple student behaviours. These behaviours were categorized into two classes, namely Attentive Activities and Distractive Activities.

Engaged Activities (Positive Contributions to Attention Span): The detection of labels, e.g. raising hands, looking at the front, reading, writing, interacting with the teacher, would positively contribute to Attentive Class Activities, and the labels using a mobile phone, eating, drinking, sleeping, or appearing inattentive would contribute to Distracted Class Activities.

The dataset is structured to have sufficient variance and diversity in student behaviours, postures, interactions, and engagement levels across the different classroom settings. The objective is to use Yolo based deep learning models to categorize students into two categories: either inactive (Distracted Class) or active (Attentive Class).

1.3. Data Processing

As the data is in the form of images, which requires standard processability for form, prior to training of the YOLO models, the dataset requires various preprocessing steps to ensure high-quality and standard input data for robust detection and classification. All the images were preprocessed to resize them all into a standard 640x640 pixels size that is suitable for training of yolo models. To improve the convergence of the model, all the pixel values were normalized within the range of 0 and 1. For the reduction of dataset diversity and overfitting, various data augmentation techniques were also applied to standardize the rotation, brightness and Gaussian noise. The dataset was split into three subsets: training (80%), validation (10%), and testing (10%). This split ensured that the model was trained on a large dataset, while validation and testing subsets were reserved for model evaluation.

Each student behaviour class was assigned a unique numeric label in the YOLO format, ensuring consistency across annotation files. For example:

- Class 0: Writing
- Class 1: Looking up to listen to the lesson
- Class 2: Raise hands
- Class 3: Turning their heads
- Class 4: Standing
- Class 5: Group discussions'
- Class 6: Looking down
- Class 7: Teacher guidance

The captured images are meticulously annotated using tools like Labelling, which allows for saving annotations in a format compatible with the YOLO model. Each image is labelled with object class, coordinates, and dimensions of the bounding boxes around the relevant sections.

- Writing: A student is seen writing in a notebook.
- Raising Hands: A student raising their hand.
- Looking Up: A student looking up at the teacher or board.

- Turning Heads: A student turning to interact with a peer.
- Standing: A student standing in the classroom.
- Group Discussions: Students interacting in a small group.
- Phone Use: A student using a mobile phone during class.

1.4. Model Implementation

The proposed framework utilizes YOLO v8 models for the detection and classification of student behaviour categories among active and inactive. These versions of yolo model employ real-time object detection, but this version of the model incorporates some of the architectural improvements for enhanced and better accuracy and efficiency compared to previous versions. Figure 3 shows the methodology diagram and the steps by which the dataset was taken from Kaggle and preprocessed according to the requirements. The data was split for training, validation and testing purpose. The trained model will then detect various class activities, based on which our framework implementation will categorize the detected class students into active or inactive students as an assistive report for the course instructor.

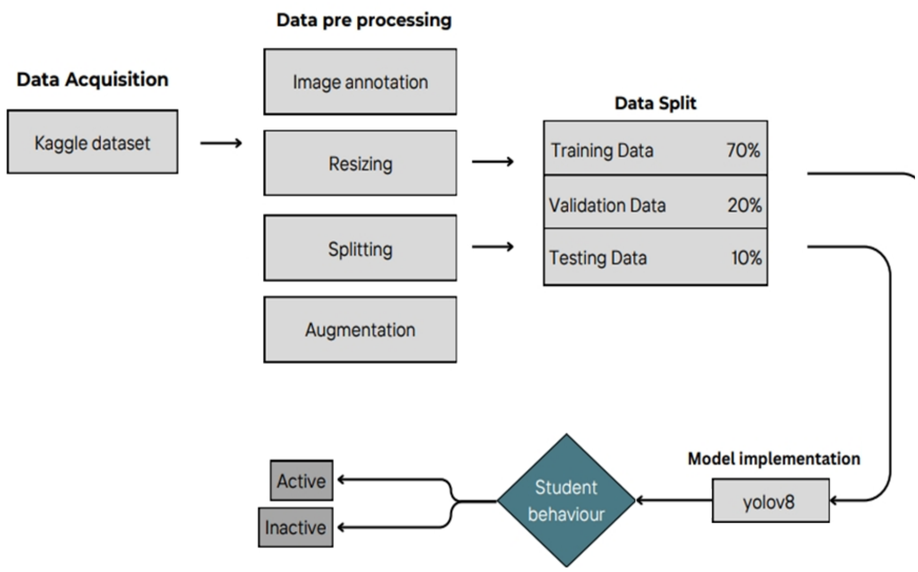

Figure 3. Proposed methodology diagram.

1.4.1. YOLO (You Only Look Once) Model

YOLO is a deep learning based single-stage object detection algorithm that processes an image in a single forward pass [15, 16]. The single forward processing makes yolo a highly efficient algorithm. In this algorithm, an under-processed image is divided into a grid for the prediction of bounding boxes (B) and class probabilities or Confidence Score P(c). It can be mathematically represented as follows,

$$P\ (c) \times (\hat{x}, \hat{y}, \widehat{w}, \hat{h}) \qquad (1)$$

Here in the relation,$(\hat{x}, \hat{y})$ are the center coordinates, and $(\widehat{w}, \hat{h})$ represent the width and height of the bounding box. Intersection over Union (IoU) is the parameter that calculates the overlap between predicted and ground truth bounding boxes and it can be estimated as follows:

$$IoU = \frac{Area\ of\ Overlap}{Area\ of\ Union} \qquad (2)$$

The calculated value of IoU represents the prediction, and it is required to be higher for better prediction.

For yolo model, the localization loss is one of the important parameters that determines the accuracy of predicted box alignment with the ground truth box to confirm that the algorithm has correctly positioned the detected objects[17]. This is also known as Regression Loss, and it depends on Mean Squared Error between predicted and actual bounding box coordinates, which can be represented as follows:

$$L_{loc} = \lambda_{coord} + \sum_{i=1}^{N} \left((x_i - \hat{x}_i)^2 + (y_i - \hat{y}_i)^2 + (w_i - \widehat{w}_i)^2 + \left(h_i - \hat{h}_i\right)^2 \right) \tag{3}$$

Classification loss and localization are balanced in a well-tuned YOLO model.

1.4.2. Yolo v8 Implementation on Google Colab

For the implementation and detection of students' behaviours using the Yolo algorithm, the Google Colab platform has been utilized. We have used YOLO v8 and YOLO v9 on Google Colab, the following steps are shown in Fig. 4.

1. Set Up Environment:

```
!pip install ultralytics
from ultralytics import YOLO
```

2. Load Pretrained YOLO Model:

```
model = YOLO('yolov8n.pt')  # Load YOLO v8 model
```

3. Train Model on Custom Dataset:

```
model.train(data='dataset.yaml', epochs=50, imgsz=640)
```

4. Evaluate Model Performance:

```
metrics = model.val()
print(metrics)
```

5. Run Inference on Sample Images:

```
results = model('sample_image.jpg')
results.show()
```

Figure 4. Code of Python Script to train YOLOv8n model.

3. Results and Discussion

This uses a trained model, which uses YOLOv8 in detection mode to predict objects in images. While the mode is predicted, which tells the model to make predictions on test data, the task to detect indicates that the task is object detection. A pre-trained YOLO model is loaded from the requested path by the best weight.pt file. By setting the confidence criterion at 25% with conf=0.25, detections with confidence scores below this level will be eliminated. The directory containing the test images to be examined is specified by the source, which is student behaviour detection test images. Ten students were not paying attention, while fifteen were, according to the results shown in Figure 5. One student was raising their hand, and five of the eight students who were paying attention were concentrated. But the two student who weren't paying attention were either bored or distracted, or they were turning their heads.

3.1. *Evaluation Metrics*

The model's performance is validated by comparing various versions of the YOLOv8 model and based on evaluation metrics such as precision, recall, mean Average Precision (map), and F1 score. This validation process ensures that the model is accurately recognizing student attention. Table I shows the evaluation parameters and their description.

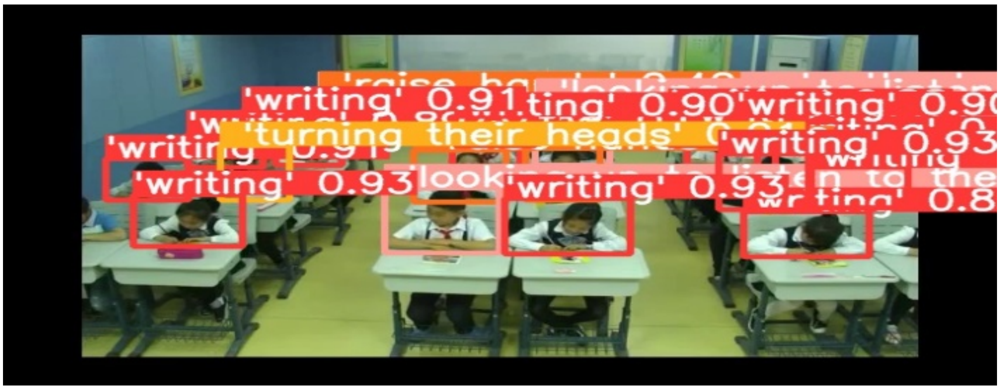

Figure 5. Predict student behavior with Class and Confidence Score.

Table I: Evaluation Metrices Table.

Evaluation Parameter	**Description**
Recall	$\frac{TP}{TP+FN}$
Accuracy	$\frac{TP+TN}{TP+TN+FP+FN}$
Precision	$\frac{TP}{TP+FP}$
F-measure	$\frac{2*Recall*Precision}{Precision+Recall}$

The YOLOv8 model showed excellent accuracy in identifying eight student behaviours from the Kaggle dataset, with an exceptional mean average precision (mAP@50) of 91.7%. Among these behaviours, "raise hands" and "writing" achieved the highest precision and recall rates, while behaviours like "turning their heads" and "standing" showed moderate performance, indicating potential areas for improvement as given in Table II.

Table II: Performance measure Kaggle dataset by using yolov8

Class	**Images**	**Instances**	**Box (P)**	**R**	**mAP50**	**mAP50-95)**
All	889	26,771	0.852	0.872	0.917	0.764
Writing	889	7,173	0.929	0.953	0.985	0.879
Looking up to listen to the lesson	889	5,868	0.850	0.858	0.927	0.786
Raise hands	889	11,836	0.932	0.947	0.982	0.886
Turning their heads	889	552	0.721	0.722	0.803	0.687
Standing	889	333	0.780	0.751	0.816	0.636
Group discussions	889	439	0.856	0.948	0.953	0.778
Looking down	889	494	0.870	0.891	0.921	0.714
Teacher guidance	889	76	0.874	0.910	0.953	0.745

3.2. Confusion matrix for the Kaggle dataset by using yolov8

The confusion matrix was calculated for the tested data to see the model's efficacy, which was implemented using yolov8. The matrix given in Fig. 6 shows how well a classification model performs in identifying various student behaviours in the classroom.

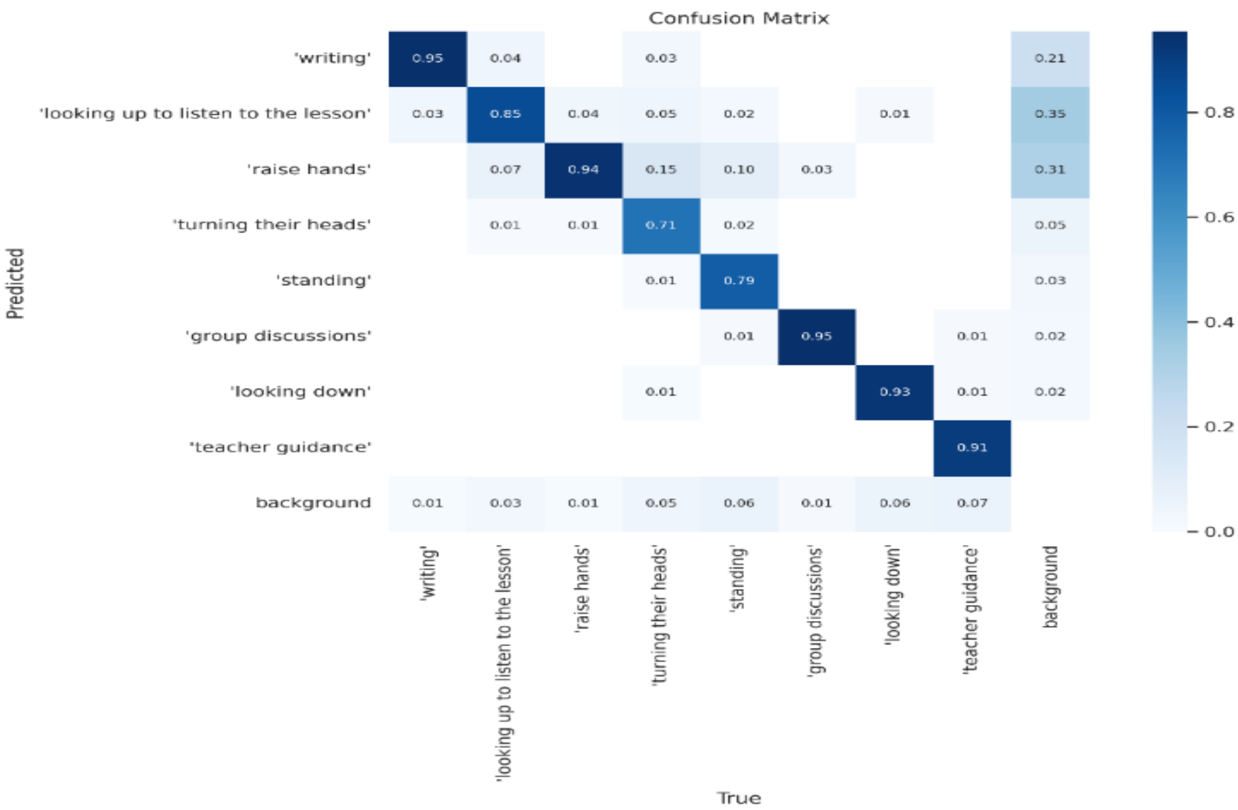

Figure 6. Confusion Matrix by using yolov8.

The confusion matrix shows how well a classification model performs in identifying various student behaviours in the classroom. Each behaviour's diagonal values show accurate predictions; categories such as "writing" (0.95), "group discussions" (0.95), and "teacher guidance" (0.91) show very high accuracy. Misclassifications, or off-diagonal values, indicate areas where the model had issues. For instance, "looking up to listen to the lesson" was frequently mistaken with "background" (0.35), and "raise hands" was occasionally confused with "turning their heads" (0.15). The colour gradient aids in the interpretation of the misclassification rates; more accuracy is indicated by deeper colours. Overall, the model works well but has difficulty in separating some closely related behaviours.

3.3. F1-curve for Kaggle dataset by using yolov8

The relationship between confidence thresholds and the F1-score for various student engagement classes can be seen in the F1-Confidence Curve graph. A broader blue line indicates the overall performance across all classes. The F1-score, which maintains a compromise between precision and recall, is shown against different confidence levels for each class. At a confidence level of 0.459, the highest F1-score of 0.86 is obtained, suggesting that precision and recall are now optimally balanced. Different engagement classes show different performance trends; some classes fall more unexpectedly, while others retain strong F1-scores throughout a wider confidence range. When it comes to student behaviour detection, this graphic aids in choosing the right confidence level to optimize model accuracy while reducing false positives and false negatives. Figure 7 shows the F1- Confidence Curve.

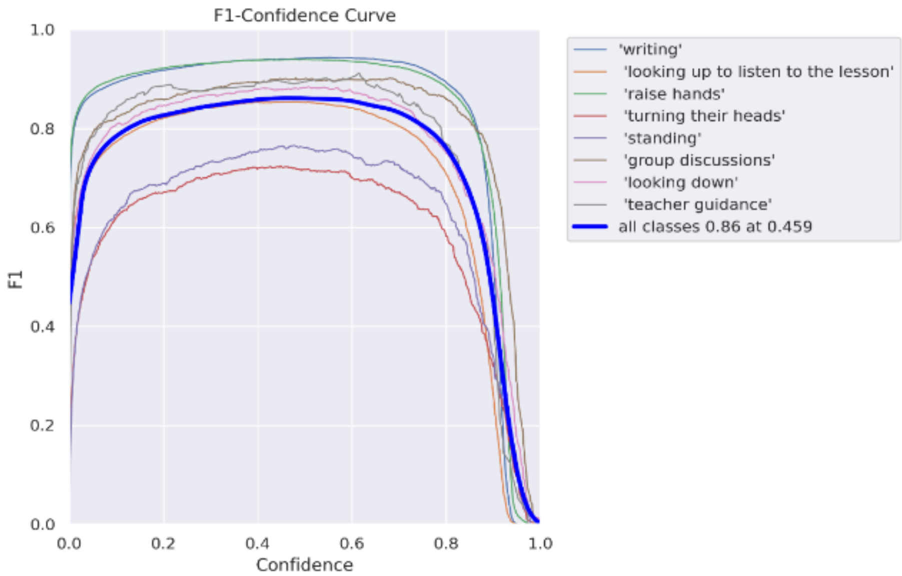

Figure 7: F1-curve by using yolov8.

The model that was trained using YOLO v8 was subjected to testing that detected various class activities, based on which our framework implementation with the detection efficiencies mentioned in previous sections. There was a total of 8 categories, out of which 5 categories were contributing to the active class and 3 categories were contributing to the non-active class. The framework will then represent the analysis graph for the classification of detected class students into active or in-active students as an assistive report for the course instructor. Figure 8 shows the detected results of positively engaged and negatively engaged students. The result shows that total Positive Engagements were 629 out of 889 total student engagement observations in the testing phase, and total Negative Engagements were 260 out of 889 total student engagement observations in the testing phase.

For the OBE perspective, it is very important that a classroom setup establishes a student-centric learning environment. Such assistive analysis could help instructors establish a positive feedback and interactive mechanism to enhance overall attention and engagement of class students for better performance.

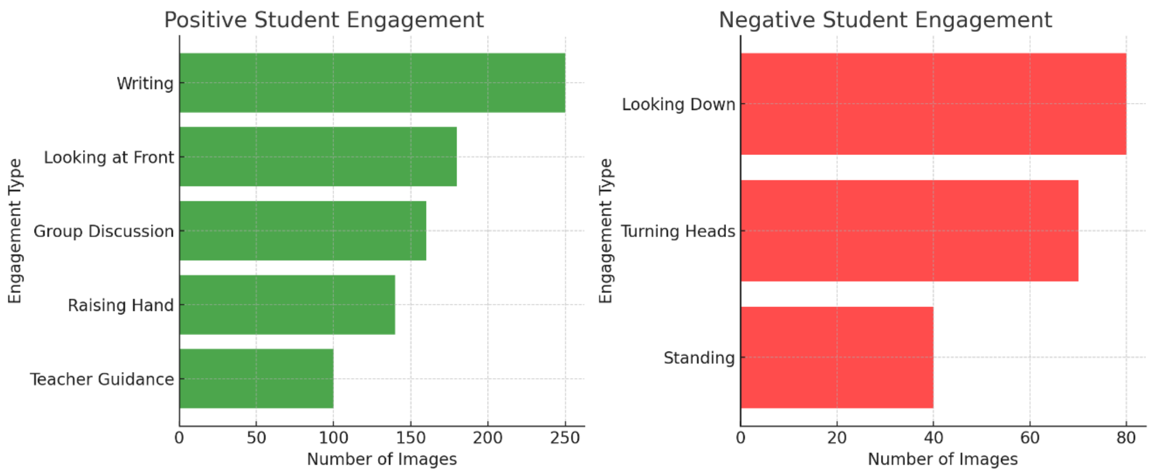

Figure 8: A Comparison of Positive and Negative Student Engagement.

4. Conclusion

The technique for evaluating student participation in an OBE classroom using deep learning methods, specifically YOLO v8, is presented in this paper. The used model showed reliable detection efficiencies for each class of activity. The approach offers useful insights for developing student-centric classroom strategies by categorizing student behaviour analysis into attention-enhancing and attention-reducing behaviours. The findings emphasize how crucial it is to keep an eye on classroom conduct in order to improve student engagement and academic success. The analysis can be utilized as an assistive tool for the instructor to improve learning strategies and feedback mechanisms for enhanced learning setups. Future research can concentrate on improving detection accuracy through the integration of multimodal data sources, including physiological and audio signals, and creating real-time feedback systems to let teachers modify their lesson plans as well as specified classroom assessment activities.

References

1. Fredricks, J.A., P.C. Blumenfeld, and A.H. Paris, *School engagement: Potential of the concept, state of the evidence.* Review of educational research, 2004. **74**(1): p. 59-109.
2. Kahu, E.R., *Framing student engagement in higher education.* Studies in higher education, 2013. **38**(5): p. 758-773.
3. Chen, H., G. Zhou, and H. Jiang, *Student behavior detection in the classroom based on improved YOLOv8.* Sensors, 2023. **23**(20): p. 8385.
4. Bozkurt, A., *Generative artificial intelligence (AI) powered conversational educational agents: The inevitable paradigm shift.* Asian Journal of Distance Education, 2023. **18**(1).
5. Li, Q., et al.., *A learning attention monitoring system via photoplethysmogram using wearable wrist devices.* Artificial intelligence supported educational technologies, 2020: p. 133-150.
6. Apicella, A., et al., *EEG-based measurement system for monitoring student engagement in learning 4.0.* Scientific Reports, 2022. **12**(1): p. 5857.
7. Qin, Y., Y. Liao, and Z. Wang. *Improved YOLOv8 algorithm for classroom student behavior detection*. in *2024 IEEE 6th Advanced Information Management, Communicates, Electronic and Automation Control Conference (IMCEC)*. 2024. IEEE.
8. Xu, J., J. Li, and J. Yang, *Self-regulated learning strategies, self-efficacy, and learning engagement of EFL students in smart classrooms: A structural equation modeling analysis.* System, 2024. **125**: p. 103451.
9. Alshamaila, Y., et al., *An automatic prediction of students' performance to support the university education system: a deep learning approach.* Multimedia Tools and Applications, 2024. **83**(15): p. 46369-46396.
10. Tjandra, E., S.S. Kusumawardani, and R. Ferdiana. *Student performance prediction in higher education: A comprehensive review*. in *AIP Conference Proceedings*. 2022. AIP Publishing.
11. Yang, F., *SCB-dataset: a dataset for detecting student classroom behavior.* arXiv preprint arXiv:2304.02488, 2023.
12. Li, L., et al., *Exploring artificial intelligence in smart education: Real-time classroom behavior analysis with embedded devices.* Sustainability, 2023. **15**(10): p. 7940.
13. Jia, Q. and J. He, *Student Behavior Recognition in Classroom Based on Deep Learning.* Applied Sciences, 2024. **14**(17): p. 7981.
14. Yang, F., T. Wang, and X. Wang. *Student classroom behavior detection based on YOLOv7+ BRA and multi-model fusion*. in *International conference on image and graphics*. 2023. Springer.
15. Jocher, G., A. Chaurasia, and J. Qiu, *Ultralytics yolo (version 8.0. 0)[computer software]*. 2023.
16. Jocher, G., A. Chaurasia, and J. Qiu, *YOLO by Ultralytics*. 2023.

1. 17. Glenn, J., C. Ayush, and Q. Jing, *Ultralytics YOLO (Version 8.0. 0)[Computer software]*. 2023.