**JOIN (Jurnal Online Informatika)**
p-ISSN: 2528-1682, e-ISSN: 2527-9165
Volume 10 Number 1 | June 2025: 93-104
DOI: 10.15575/join.v10i1.1390

# A Comparison of YOLOv8 Series Performance in Student Facial Expressions Detection on Online Learning

**Dewi Tresnawati[1], Shopi Nurhidayanti[2], Nina Lestari[3]**
[1,2] Department of Computer Science, Institut Teknologi Garut, Garut, Indonesia
[3] Department of Electrical Engineering, Universitas Sangga Buana, Bandung, Indonesia

**Article Info**

***Article history:***

Received June 15, 2024
Revised December 24, 2024
Accepted March 06, 2025
Published April 01, 2025

***Keywords:***

Emotion Recognition
Facial Expression Detection
Online Learning
Student Engagement
YOLOv8

**ABSTRACT**

Student engagement in online learning is an important factor that can affect learning outcomes. One indicator of engagement is facial expression. However, research on facial expression detection in online learning environments is still limited, especially in the use of the YOLOv8 algorithm. This study aims to compare the performance of several YOLOv8 variants, namely YOLOv8x, YOLOv8m, YOLOv8s, YOLOv8n, and YOLOv8l in recognizing six facial expressions: happy, sad, angry, surprised, afraid, and neutral. Student facial expression data was collected through the Moodle platform every 15 seconds during the learning process. All models were trained using 640x640 pixel images for 100 epochs to improve facial expression detection capabilities. The main contribution of this study is to provide a comprehensive analysis of the effectiveness of YOLOv8 in detecting student facial expressions, which can be used to improve the online learning experience. The evaluation results show that the YOLOv8s model has the best performance with the highest mAP of 0.840 and the fastest inference speed of 2.4 ms per image. YOLOv8m and YOLOv8x also performed well with mAP of 0.816 and 0.815, respectively. Although YOLOv8x had the slowest inference speed, it was superior in detecting fear, happiness, and sadness expressions with mAP above 0.9. YOLOv8n had mAP of 0.636, while YOLOv8l achieved mAP of 0.813 with an inference speed of 9.1 ms per image. This study shows that the YOLOv8 algorithm, especially YOLOv8s, can be an effective solution to analyze student engagement based on their facial expressions during online learning.

***Corresponding Author:***

Dewi Tresnawati,
Computer science Department, Institut Teknologi Garut
Jl. Mayor Syamsu No 1, Garut, Indonesia. 44151
Email: dewi.tresnawati@itg.ac.id

## 1. INTRODUCTION

The You Only Look Once (YOLO) algorithm is a CNN algorithm for identifying objects by creating bounding boxes and classifying fill from pictures or videos. YOLO is capable of processing pictures at the speed of 45 Frames Per Second (FPS), which is faster than the method of detecting objects [1]. Apart from that, YOLO can also be processed in real time with level good accuracy. Version latest from YOLO, namely YOLOv8, has enhancement in speed and accuracy detection objects compared to the previous version [2] .

The You Only Look Once version 5 (YOLOv5) algorithm successfully detected expressions like happy, sad, and surprised with the mean Average Precision (mAP) level of 96% and value accuracy reached 87%, which shows that the YOLOv5 algorithm is capable of detecting expressions with Good. However, YOLOv8, as the latest development from the YOLO algorithm, shows enhancement compared to the previous version, especially in matter precision and F1 score. The data shows that YOLOv8 has superiority with difference mark precision and F1 score reached 2.82% and 0.98%, respectively, in a row [2]. Further, deep study proves that YOLOv8 can classify itself as Good. By using the designed dataset specialty and labeling process classes conducted in a way independent, YOLOv8 succeeded detected 25 of 26 images with a level success of 96.5%, with level accuracy of 99.8%, precision of 99,4%, and recall of 99,8% [3].

Additionally, YOLOv8 can detect objects moving in complex conditions with reasonable accuracy, which compares the performance between YOLOv5 and YOLOv8 in detecting objects moving that display level complexity. Tested aspects including condition occlusion, change size minimal spatial, and rotation object. The result shows that YOLOv8 outperforms YOLOv5, especially in matter accuracy, with the folder reaching the 50-95 range of 0.835 after 190 epochs. Study This confirms that YOLOv8 can detect objects in various situations while maintaining performance for use in real-time [5]. YOLOv8 is used to detect and produce more performance compared to the previous YOLO variant, with mAP on level 0.5 confidence of 0.981 and at level confidence of 0.95 is 0.827. However, the use of YOLO in Facial Emotion Recognition (FER) is still ongoing and tends to be little used especially YOLO version 8 [6]. FER is one of the technologies that can be used for analysis to express facial "Good" from pictures and videos to obtain information about somebody's emotions [8]; its use still depends on the algorithm used [9] . In addition, other research also developed a face detection system and emotion classification using the YOLO and CNN algorithms. Using the FER2013 dataset, the results showed that the system was able to identify seven main emotions with 94% accuracy. These results demonstrate the potential of YOLO in providing high speed and accuracy for computer vision-based emotion recognition applications in various fields, such as security and health [7].

In learning, expression plays a crucial role in determining engagement and learning results. That matters because the expression of face gives an outlook on emotions experienced by students during the learning process [8] [9] [10] [11] [14]. The results of the Academic Emotions Questionnaire (AEQ) study have identified and measured various emotions experienced by the student during the study, and research finds that emotions like happiness, hopefulness, pride, and bored own role crucial. Positive emotions, happiness, and pride, related to suitable responses to material learning and success in overcoming challenges, show that positivity can increase understanding and achievement results for Study students. On the contrary, boredom reflects a lack of involvement or interest, which can lower understanding and potentially cause failure in processing assessment through change expression [13]. This changing emotion happens quickly, ranging between 3 to 15 seconds [14].

This study chose YOLOv8 because the algorithm has advantages compared to other YOLO versions, such as YOLOv4 and YOLOv5, in terms of precision and speed of facial expression detection. YOLOv8 has faster detection capabilities with higher precision and F1 scores, as shown in a study by Sary [2], which makes it more efficient to use in dynamic learning environments. Compared to YOLOv4 or YOLOv5 which have lower detection speeds and slightly lower accuracy, YOLOv8 provides more optimal results in real-time facial expression detection. In this study, six facial expressions (angry, fearful, happy, sad, surprised, and neutral) were chosen because of the relevance of these emotions in the learning context which can provide a clearer picture of students' emotional involvement during the learning process based on the results of testing conducted by Gupta [15].

FER research, which identifies seven expression bases, such as happy, sad, angry, surprised, disgusted, afraid, and neutral, is considered an indicator important for level involvement students [15]. During the learning process, test results in distribution data expression students discovered that expression neutral was the most dominant, reaching 87% of the total expression detected. This result shows that big students display more neutral expressions in context learning. The expression like is the second largest, with 7% of the total, indicating moments of happiness and engagement for students. Other expressions like sad, surprised, angry, and afraid only appear in a tiny percentage, between 1% to 3%. Expression of disgust is not detected in testing, which shows that negative emotions or very intense situations seldom happen.

Based on analysis problems and literature review, the YOLO algorithm, especially YOLOv8 in detecting objects, has proven its high accuracy. However, its use in FER is still limited, especially in environment learning. Therefore, the research aims to compare the YOLOv8 series in FER to recognize six expression faces based on results testing, namely happy, sad, angry, surprised, afraid, and neutral, using expression data for students during assessment through the Moodle Learning Management System platform [15]. Where the dataset used in this study consists of images of students' facial expressions taken during assessments through the Moodle Learning Management System platform. These images were collected at intervals of every 15 seconds a method based on research [14], using devices such as mobile phones and laptops with varying camera resolutions. For mobile phones, camera resolutions generally range from 1920x1080 pixels (Full HD) to 3840x2160 pixels (4K), while for laptops, camera resolutions are usually around 1280x720 pixels (HD) to 1920x1080 pixels (Full HD). All images are saved in PNG format to ensure high image quality without losing visual data, with RGB (Red, Green, Blue) colors, the standard color format for digital images. This dataset consists of 18,236 images taken from 146 students, each depicting facial expressions in six categories, namely happy, sad, angry, surprised, afraid, and neutral. This dataset was chosen because of its relevance in observing students' facial expressions during technology-based learning processes, which can provide insights into students' emotional engagement.

## 2. METHOD

### *2.1. Framework Research*

In general, this research flow uses the Machine Learning Life Cycle method [14] which consists of six main stages: acquisition, inspection, preparation, modeling, evaluation, and deployment. This cycle ensures that any deficiencies in the model can be corrected by returning to the initial stage for improvement and adjustment. This method is not a new concept, but its application in a facial expression recognition-based proctoring system provides novelty in the context of implementation. In this study, the data approach used is an augmentation system for data balance and mAP (mean Average Precision)-based evaluation to improve the accuracy of facial expression detection.

#### 2.1.1. Acquisition

The acquisition phase begins with the use of webcams to record students’ activities during the assessment. This process provides three attempts for each student. The proctoring system automatically stores all activities during the assessment in a database, including students’ identity information and their facial images.

The data collected from 146 students amounted to 18,236 images, which were then stored in a separate database to facilitate the extraction process. This step ensures that the data can be accessed efficiently for the next stage. The acquisition process flow is illustrated in Figure 2.

#### 2.1.2. Inspection

This stage consists of data cleaning and data selection for modeling. Validation was carried out by the PPTIK ITB research team based on six main facial expressions: angry, afraid, happy, sad, surprised, and neutral. The result of this process was a data selection of 11,419 images that met the validation criteria, namely displaying the entire area of the student's face (eyes, nose, forehead, mouth, eyebrows, lips, and cheeks). However, to ensure data balance in the model, the final amount of data used was angry (130 images), afraid (16 images), happy (250 images), neutral (450 images), sad (350 images), and surprised (210 images). The dominance of neutral expression data is the main reason for balancing the amount of data in each emotion class.

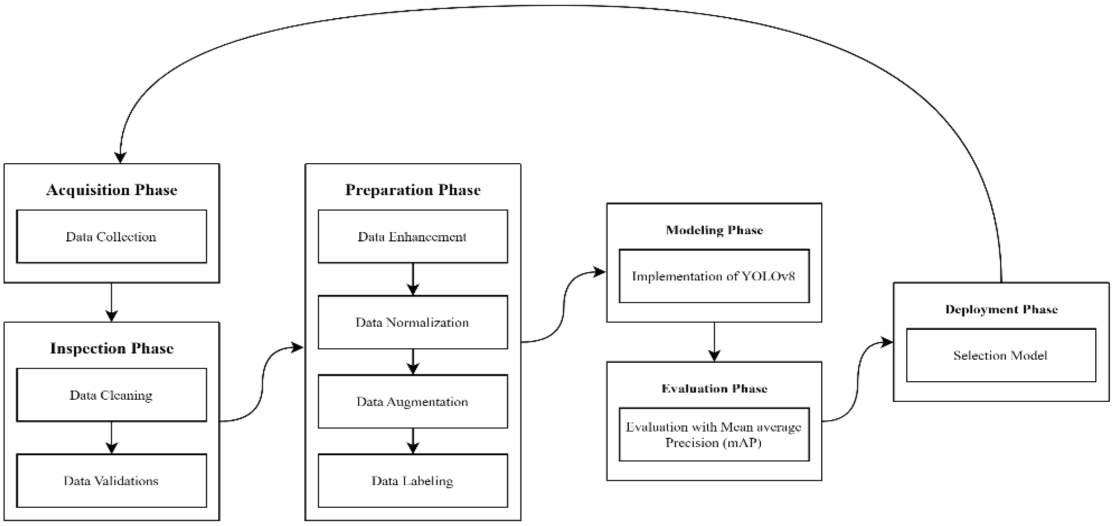


Figure 1. Research Framework

### 2.1.3. Preparation

After the data is validated, the preparation stage is carried out to make it suitable for modeling. The processes carried out include:

1. Enhancement: Improving image quality by increasing lighting intensity.
2. Normalization: Adjusting image dimensions to a scale of 300x300 pixels to facilitate model processing.
3. Augmentation: Techniques used include rotation, cropping, scaling, brightness adjustment, and flipping. Augmentation is done to balance the amount of data in each facial expression class.
4. Labeling: Done with the LabelImg tool from the Label Studio community, which provides appropriate bounding boxes and expression labels.

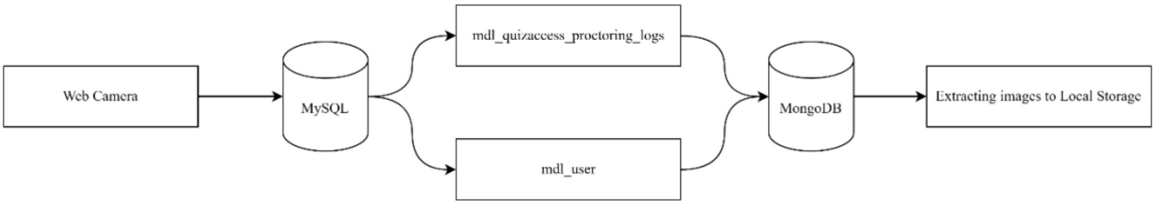

Web Camera
MySQL
mdl_quizaccess_proctoring_logs
mdl_user
MongoDB
Extracting images to Local Storage

Figure 2. Flow Acquisition

### 2.1.4. Modelling

At this stage, the algorithm used is YOLOv8 (You Only Look Once version 8), which is an object detection algorithm that uses processed data to train the model. The basic step taken is to divide the data into training data (80%) and validation data (20%). The detection models used include several variants of YOLOv8, namely YOLOv8n (nano), YOLOv8s (small), YOLOv8m (medium), YOLOv8l (large), and YOLOv8x (extra large). The training process is carried out for 100 epochs with an image size of 640x640 pixels. The purpose of this process is to improve the model's ability to detect students' facial expressions. [19].

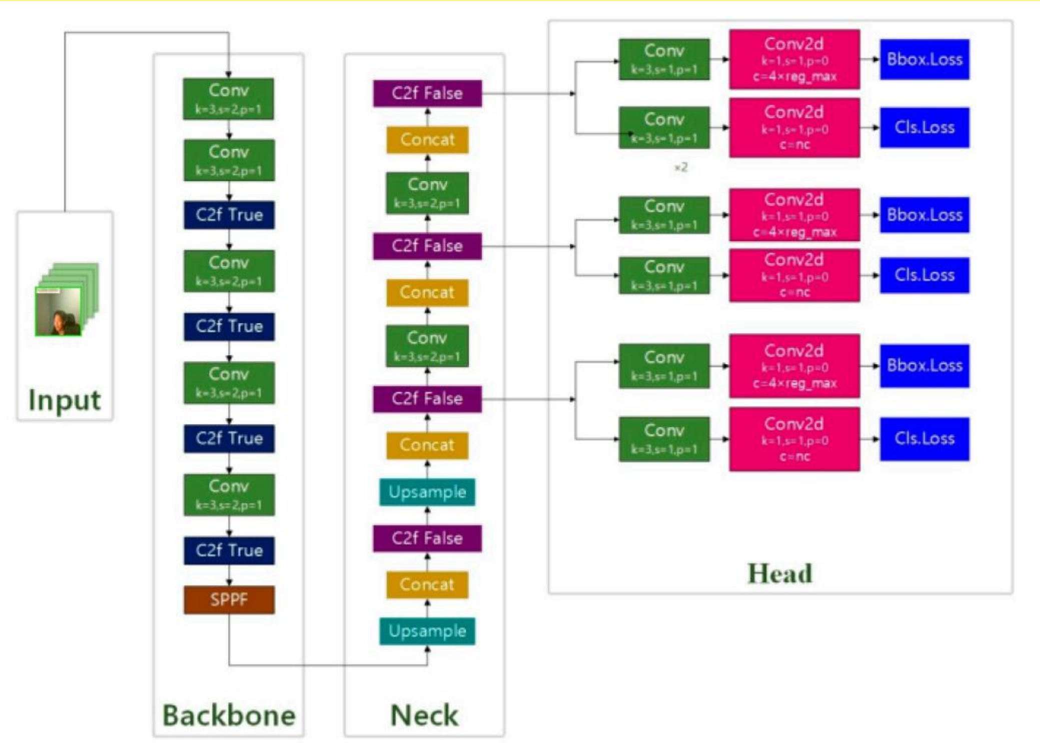


Figure 3. YOLO Model Architecture

#### 2.1.5. Evaluation

Model evaluation is done by calculating the mAP (mean Average Precision) value for each facial expression class. The mAP value measures the accuracy of object detection by comparing the model's prediction results to the ground truth dataset [20]. This is done to determine the extent to which the model is able to identify each facial expression accurately. The higher the mAP value (closer to 1), the better the model performance [21]. The evaluation results will be used to select the best model to be implemented.

#### 2.1.6. Deployments

In this study, the deployment process only includes the selection of 5 trained YOLOv8 models. The model that shows the best performance in detecting six facial expressions will be selected for implementation in a wider environment.

Table 1. Research Data

| Phase | Process | Results |
|---|---|---:|
| Acquisition | Data retrieval | 18,236 |
| Inspection | Cleaning | 11,419 |
| | Validation by PPTIK ITB Researchers | 1,406 |
| Preparation | Augmentation | 2,250 |
| Modelling | Training | 1,800 |
| | Validation | 450 |

### *2.2. Facial Expression*

In learning, students often utilize body language, posture, and facial expressions to support explanations about certain concepts or events. The influence of communication through facial expressions is significant because the face provides a picture of a person's identity, mood, and outlook on life, which makes it possible to understand a person's emotional processes [22]. The following are examples of facial expressions based on the dataset used:

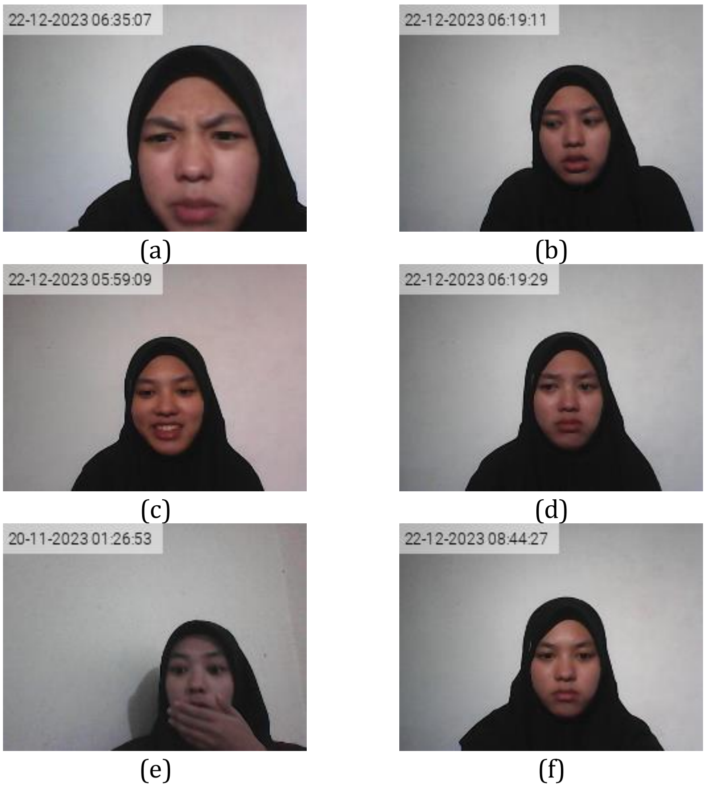


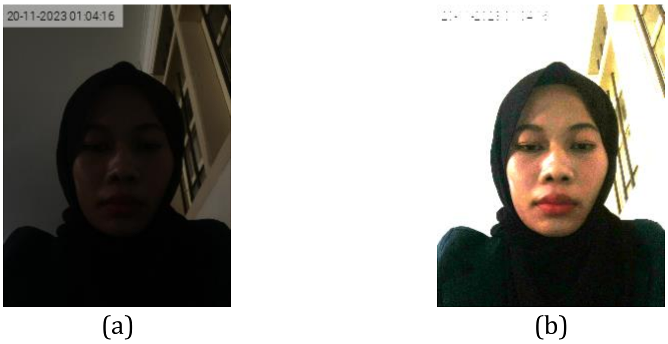
Figure 2. Facial Expression (a) angry (b) fear (c) happy (d) sad (e) surprise (f) neutral

According to psychological scientist Paul Ekman, human facial expressions can be grouped into six basic emotions: surprise, sadness, joy, fear, disgust, and anger. Raised eyebrows, wide-open eyes, and an open mouth indicate surprise. Sadness is characterized by lowered eyebrows, corners of the lips, and teary or crying eyes. Happy is characterized by raised lip corners, shining eyes, and raised cheeks. Raised eyebrows, wide-open eyes, and an open mouth with pulled-back lips indicate fear. Disgust is characterized by a twitching nose, a raised upper lip, and an expression indicating nausea. Lowered eyebrows, bulging eyes, and an open mouth with pulled-back lips characterize anger [10] .

## 3. RESULTS AND DISCUSSION

### *3.1. Data Processing*

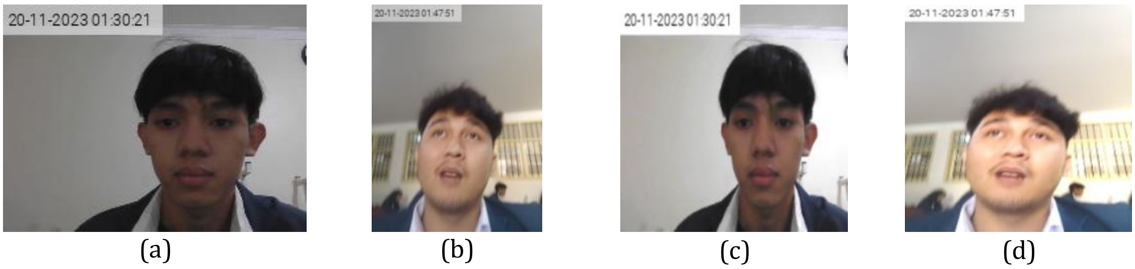
The enhancement process involves increasing the lighting intensity for several images with dark image quality but still recognizable faces. This is done using the sci-kit-image library via rescale intensity with varying values depending on the image conditions. Figure 5 shows the differences in results before and after enhancement.



(a) (b)

Figure 5. Enhancement Process (a) Before (b) After

Figure 6 is the result of the normalization process. This process was carried out because the data held had different variations in image dimensions, which was caused by differences in the devices used by students. Normalization is done by changing all image dimensions to the same size, namely 300x300 pixels to make the next process easier. This process is carried out using the OpenCV library and resize function.

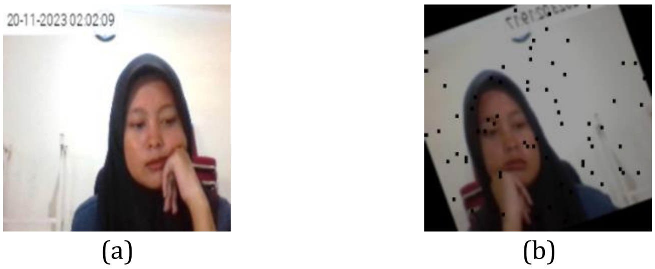


(a) (b) (c) (d)

Figure 6. Normalization Process (a) (b) Before (c) (d) After

Meanwhile, the meticulous augmentation process using imgaug involves several precise steps. First, it employs the Fliplr function with a probability of 0.5, which generates a random horizontal rotation of the image. Next, a random rotation between -25 to 25 degrees is executed using the Affine function. The subsequent step is a random brightness adjustment using the Multiply function, ranging between 0.5 to 1.5 times the initial brightness. A Gaussian blur effect is applied using the GaussianBlur function, with a random sigma between 0 and 3.0. Finally, using the CoarseDropout function, some pixels are randomly deleted from the image with varying sizes and intensities, namely 1% to 10% between 2% and 25% of the image dimensions. This augmentation process is implemented using the imgaug augmenters library, where the results of the augmentation process can be seen in Figure 7.

Augmenting less numerous classes, such as angry, afraid, happy, sad, and surprised, is a strategic move. By expanding the variety and number of samples in these classes, the model can better learn the features representing each class, thereby improving overall facial expression detection performance. Despite the initial data imbalance, after augmentation, each class will have a more even amount of data, which can significantly enhance the accuracy and reliability of the facial expression detection model.



(a) (b)

Figure 7. Augmentation Process (a) Before (b) After

Labeling is done using the LabelImg tool from the Label Studio community. The goal is to mark objects in the image by providing bounding boxes and appropriate expression labels. The results of the labeling process are saved in the form of a txt file, which contains information about the expression class, coordinates of the x-axis and y-axis, and the width and height of each given bounding box.

Labeling is very important in using the YOLO (You Only Look Once) algorithm because YOLO requires specific information about the location and class of objects in the image. Labeling provides information about the location and class of facial expressions that the model wants to detect. The YOLO model requires this information during training to understand and learn the patterns associated with each expression class.

### *3.2. Implementation Model*

The performance of several YOLOv8 models is evaluated based on model size, speed, and accuracy. These models have the same input image size, namely 640x640 pixels, but have differences in the number of parameters and FLOPs (floating point operations per second). YOLOv8n, the smallest model, has 3.2 million parameters and 8.7 billion FLOPs, while YOLOv8x, the largest model, has 68.2 million parameters and 257.8 billion FLOPs. Although YOLOv8x has the highest accuracy with a mean Average Precision (mAP) of 53.9%, the larger model size causes an increase in inference time, especially when using CPU, which took 479.1 ms on the A100 TensorRT CPU. In addition, the inference speed is also affected by the model size. YOLOv8n, despite fast inference performance with 80.4 ms on an A100 TensorRT CPU, has a lower mAP compared to larger models [21] .

Table 2. Detection Model of YOLOv8

| Model | Size (pixels) | mAPval (50-95) | Speed CPU ONNX (ms) | Speed A100 TensorRT (ms) | Params (M) | FLOPs (B) |
|---|---|---|---|---|---|---|
| YOLOv8n | 640 | 37.3 | 80.4 | 0.99 | 3.2 | 8.7 |
| YOLOv8s | 640 | 44.9 | 128.4 | 1.20 | 11.2 | 28.6 |
| YOLOv8m | 640 | 50.2 | 234.7 | 1.83 | 25.9 | 78.9 |
| YOLOv8l | 640 | 52.9 | 375.2 | 2.39 | 43.7 | 165.2 |
| YOLOv8x | 640 | 53.9 | 479.1 | 3.53 | 68.2 | 257.8 |

The training process was carried out using all series of YOLOv8 detection models, consisting of YOLOv8n, YOLOv8s, YOLOv8m, YOLOv8l, and YOLOv8x. It was carried out for 100 epochs with an image size of 640x640 pixels. This process aims to improve the model's detection ability for facial expressions to better understand and recognize various expressions through thorough training.

After the training process, the evaluation results of the YOLOv8n model succeeded in getting a mAP (mean Average Precision) of 0.636. This model has an inference speed of 3.3 ms per image. However, the detection performance of happy and neutral expressions has a relatively low mAP, namely 0.444 and 0.226. For expressions of anger, fear, sadness, and surprise, the YOLOv8n model performs well with mAPs of 0.391, 0.557, 0.171, and 0.308, respectively.

The YOLOv8s model has a greater number of parameters compared to YOLOv8n and achieves a mAP of 0.840 with an inference speed of 2.4 ms per image. Expression detection in all categories, except neutral expression, has mAP above 0.7. For neutral expression detection, the model can detect with an AP of 0.585.

The YOLOv8m model, with 25,843,234 parameters, achieves a mAP of 0.816 and has an inference speed of 5.4 ms per image. Detection of angry, fearful, and happy expressions produces mAP above 0.8, with values of 0.786, 0.966, and 0.886, respectively. However, the detection performance for neutral, sad, and surprise expressions still needs improvement, with mAPs of 0.613, 0.734, and 0.912, respectively.

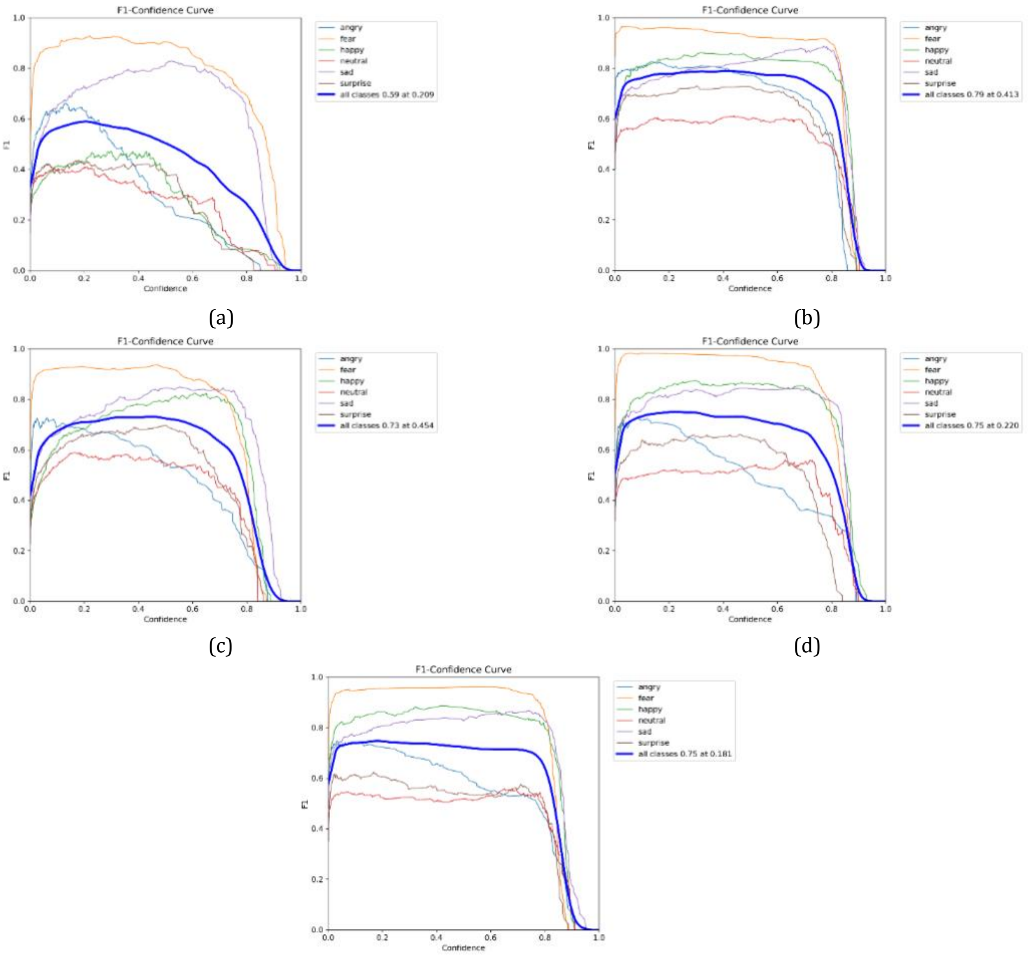
YOLOv8l has 43,611,234 parameters and achieves an mAP of 0.813 with an inference speed of 9.1 ms per image. This model performs well in detecting angry, fearful, happy, and sad expressions with mAP of 0.799, 0.988, 0.94, and 0.537, respectively. However, the detection performance of neutral and surprise expressions is at a lower AP, namely 0.699 and 0.913.

Finally, YOLOv8x, with 68,129,346 parameters, produces an mAP of 0.815. This model's inference speed is the slowest among the other model series, 14.0 ms per image. The detection of fear, happy, and sad expressions has an mAP above 0.9, with values of 0.968, 0.932, and 0.547, respectively. However, there was a decrease in performance in detecting neutral and surprise expressions, with AP of 0.699 and 0.926, respectively.

The YOLOv8s model performs best, with the highest mAP of 0.840 and the fastest inference speed of 2.4 ms per image. The YOLOv8m and YOLOv8x models also perform well, although with slower inference speeds than YOLOv8s.

Table 4. Result Model of YOLOv8

| Model | Speed ( ms ) | AP Angry | AP Fear | AP Happy | AP Neutral | AP Sad | AP Surprise | folder |
|---|---|---|---|---|---|---|---|---|
| YOLOv8n | 3.3 | 0.391 | 0.444 | 0.226 | 0.171 | 0.557 | 0.308 | 0.636 |
| YOLOv8s | 2.4 | 0.861 | 0.984 | 0.917 | 0.585 | 0.939 | 0.754 | 0.840 |
| YOLOv8m | 5.4 | 0.786 | 0.966 | 0.886 | 0.613 | 0.912 | 0.734 | 0.816 |
| YOLOv8l | 9.1 | 0.799 | 0.988 | 0.94 | 0.537 | 0.913 | 0.699 | 0.813 |
| YOLOv8x | 14.0 | 0.816 | 0.968 | 0.932 | 0.547 | 0.926 | 0.699 | 0.815 |



(a)

(b)

(c)

(d)

(e)

Figure 8. F1-Confidence Curve (a) YOLOv8n (b) YOLOv8s (c) YOLOv8m (d) YOLOv8l (e) YOLOv8x

Figure 8 shows the value of the F1-Confidence Curve, which illustrates the relationship between the F1-score value and the confidence threshold in the YOLO object detection model. YOLOv8x, even though it has high performance with an F1-score reaching 0.75, reaches this point in confidence with a relatively low threshold, namely 0.181. This illustration shows that this model can produce highly accurate detections even with low confidence levels.

YOLOv8s performs slightly better with an F1-score of 0.79 but reaches a higher confidence threshold of 0.413. The F1-score indicates that this model requires a higher confidence level to produce detections equivalent to YOLOv8x. Meanwhile, YOLOv8n has a lower F1-score, namely 0.59, and reaches that point at a confidence threshold of 0.209. This F1-score shows that this model performs less than YOLOv8x and YOLOv8s at the same confidence level.

YOLOv8m and YOLOv8l have almost equivalent F1-scores, reaching 0.73 and 0.75 respectively. However, both reach this point at different confidence thresholds, with YOLOv8m at 0.454 and YOLOv8l at 0.220. Although both have almost the same performance, YOLOv8m is more confident in its predictions than YOLOv8l at the same confidence level.

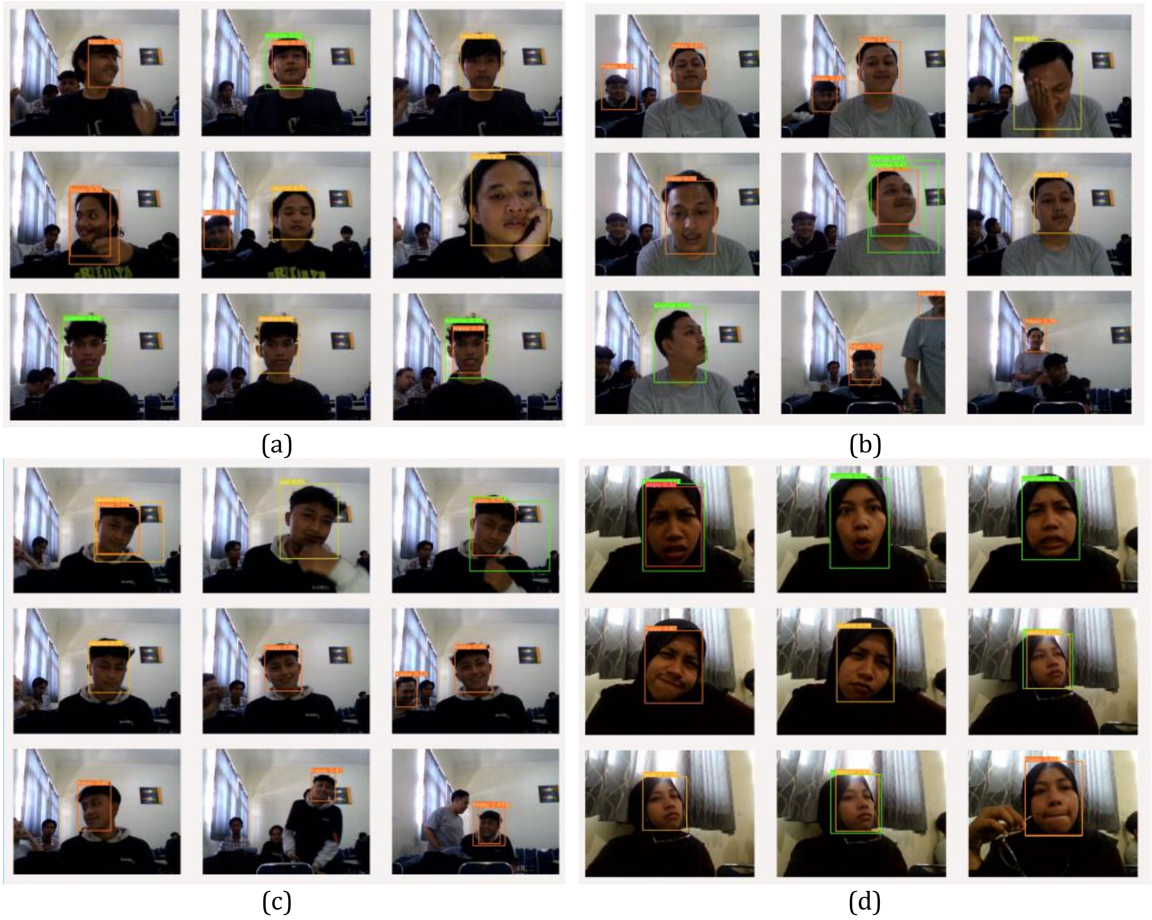

(a) (b)
(c) (d)

Figure 9. YOLOv8l Testing Results

The testing of the models on students during learning showed that the five models resulting from training and validation successfully recognized student facial expressions such as happy, sad, surprised, and neutral. However, fearful expressions tend not to appear during the testing process during learning, with the threshold value used in model testing being the default value, 0.25. However, each model faces several problems. YOLOv8x tends to detect surprised and sad expressions erroneously. YOLOv8m has difficulty recognizing angry expressions. YOLOv8s experienced errors in recognizing expressions of surprise and sadness identified as anger. YOLOv8n has an error in detecting happy expressions detected as neutral or surprise. However, compared to other models, YOLOv8l performs best in recognizing several expressions, including angry, happy, sad, surprised, and neutral. This comparison shows that the YOLOv8l model can recognize student facial expression variations accurately. Overall, the YOLO object detection algorithm, especially YOLOv8, successfully recognized angry, happy, sad, surprised, and neutral expressions well, although fear expressions tend not to appear in the learning process.

### *3.3. Discussion*

This study highlights the importance of balancing model size and inference speed with detection accuracy in real-time applications such as facial expression recognition in educational settings. While YOLOv8x provides the highest accuracy, its slow inference time makes it less practical for environments where real-time processing is critical. On the other hand, YOLOv8s offers a good balance of speed and accuracy, making it the most suitable choice for facial expression detection in dynamic learning environments.

The findings also suggest that augmenting less-represented classes, such as fear and surprise, can significantly enhance the model's ability to generalize across various facial expressions. Further investigation is needed to improve the detection of neutral and happy expressions, which presented challenges in the lower-performing models.

## 4. CONCLUSION

In the performance comparison between various YOLOv8 model series—YOLOv8n, YOLOv8s, YOLOv8m, YOLOv8l, and YOLOv8x—several key conclusions can be drawn regarding performance and inference speed. The YOLOv8s model demonstrates the best balance of performance and speed, achieving the highest mean Average Precision (mAP) of 0.840, with the fastest inference speed of 2.4 ms per image. This makes YOLOv8s the most efficient model in terms of both accuracy and speed. YOLOv8m and YOLOv8x also perform well, with mAPs of 0.816 and 0.815, though their inference speeds are slower, at 5.4 ms and 14.0 ms per image, respectively. YOLOv8n, while showing good performance with an mAP of 0.636 and a relatively fast inference speed of 3.3 ms per image, has lower accuracy compared to the other models. YOLOv8l, with an mAP of 0.813 and an inference speed of 9.1 ms per image, shows consistent detection capabilities but is slightly slower than YOLOv8m.

All models were able to detect key facial expressions—angry, happy, sad, surprised, and neutral—though with varying levels of accuracy across different expressions. YOLOv8s and YOLOv8m performed strongly across most expression categories, with mAPs above 0.7 for most expressions. YOLOv8x, while having the slowest inference speed, excelled in detecting fear, happy, and sad expressions with an mAP above 0.9. However, each model had some challenges: YOLOv8n struggled with detecting happy and neutral expressions, YOLOv8m had difficulty with angry expressions, YOLOv8s often misclassified surprised and sad expressions as angry, and YOLOv8x tended to err in detecting surprised and sad expressions. On the other hand, YOLOv8l demonstrated the most accurate recognition across various facial expressions, outperforming the other models in this regard.

During the learning and testing phases with students, all models, including YOLOv8l, showed high accuracy in recognizing common expressions such as happy, sad, surprise, and neutral. Fear expressions, being less common or intense in the learning context, were less frequently detected, but the models still performed well in identifying them. YOLOv8l, in particular, stood out for its ability to capture a wide range of facial expression variations, which further highlights the effectiveness of YOLOv8, especially in educational settings where accurate facial expression recognition is crucial for assessing student engagement and emotions.

## ACKNOWLEDGEMENTS

This research was supported by Institut Teknologi Garut. We thank our colleagues at the Department of Computer Science, Institut Teknologi Garut, and the Department of Electrical Engineering, Universitas Sangga Buana, who provided insight and expertise that greatly assisted the research, although they may not agree with all of the interpretations/conclusions of this paper.

## REFERENCES

[1] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, “You only look once: Unified, real-time object detection,” Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit., vol. 2016-Decem, pp. 779–788, 2016, doi: 10.1109/CVPR.2016.91.

[2] I. P. Sary, S. Andromeda, and E. U. Armin, “Performance Comparison of YOLOv5 and YOLOv8 Architectures in Human Detection using Aerial Images,” Ultim. Comput. J. Sist. Komput., vol. 15, no. 1, pp. 8–13, 2023, doi:

10.31937/sk.v15i1.3204.

[3] A. Ma'aruf and M. Hardjianto, "Application of the You Only Look Once Version 8 Algorithm for Indonesian Sign Language Alphabet," Semin. Nas. Mhs. Fak. Teknol. Inf., vol. 2, no. September, pp. 567–576, 2023.

[4] D. Reis, J. Kupec, J. Hong, and A. Daoudi, "Real-Time Flying Object Detection with YOLOv8," 2023, [Online]. Available: http://arxiv.org/abs/2305.09972

[5] M. Safran, A. Alajmi, and S. Alfarhood, "Efficient Multistage License Plate Detection and Recognition Using YOLOv8 and CNN for Smart Parking Systems," J. Sensors, vol. 2024, pp. 1–18, 2024, doi: 10.1155/2024/4917097.

[6] A. L. Cîrneanu, D. Popescu, and D. Iordache, "New Trends in Emotion Recognition Using Image Analysis by Neural Networks, A Systematic Review," Sensors, vol. 23, no. 16, 2023, doi: 10.3390/s23167092.

[7] P. Sharma, P. Sharma, V. Deep, and V. K. Shukla, "Facial Emotion Recognition Model," Lect. Notes Mech. Eng., no. 1, pp. 751–761, 2021, doi: 10.1007/978-981-15-9956-9_73.

[8] Jian-Ming Sun, Xue-Sheng Pei, and Shi-Sheng Zhou, "Facial emotion recognition in modern distant education system using SVM," in 2008 International Conference on Machine Learning and Cybernetics, IEEE, Jul. 2008, pp. 3545–3548. doi: 10.1109/ICMLC.2008.4621018.

[9] P. Ekman and H. Oster, "Facial Expressions of Emotion," Annu. Rev. Psychol., vol. 30, no. 1, pp. 527–554, Jan. 1979, doi: 10.1146/annurev.ps.30.020179.002523.

[10] M. R. Reyes, M. A. Brackett, S. E. Rivers, M. White, and P. Salovey, "Classroom emotional climate, student engagement, and academic achievement.," J. Educ. Psychol., vol. 104, no. 3, pp. 700–712, Aug. 2012, doi: 10.1037/a0027268.

[11] K. Seashore Louis, Cultivating Teacher Engagement: Breaking the Iron Law of Social Class, no. 7. 2020. doi: 10.4324/9780203012543-16.

[12] R. Pekrun, "The Control-Value Theory of Achievement Emotions: Assumptions, Corollaries, and Implications for Educational Research and Practice," Educ. Psychol. Rev., vol. 18, no. 4, pp. 315–341, Nov. 2006, doi: 10.1007/s10648-006-9029-9.

[13] O. Stanley and G. Hansen, ABSTUDY: An Investment forTomorrow's Employment A Review ofABSTUDY forthe AboriginalandTorres StraitIslanderCommission by Owen Stanley andGeoffHansen. 1998.

[14] R. Pekrun et al., "Academic Emotions in Students ' Self-Regulated Learning and Achievement : A Program of Qualitative and Quantitative Research Academic Emotions in Students ' Self-Regulated Learning and Achievement : A Program of Qualitative and Quantitative Research," no. July 2013, pp. 37–41, 2010, doi: 10.1207/S15326985EP3702.

[15] H. Gunes and M. Pantic, "Automatic, Dimensional and Continuous Emotion Recognition," Int. J. Synth. Emot., vol. 1, no. 1, pp. 68–99, 2010, doi: 10.4018/jse.2010101605.

[16] S. Gupta, P. Kumar, and R. K. Tekchandani, "Facial emotion recognition based real-time learner engagement detection system in online learning context using deep learning models," Multimed. Tools Appl., vol. 82, no. 8, pp. 11365–11394, 2023, doi: 10.1007/s11042-022-13558-9.

[17] A. Combs and M. Roman, Python Machine Learning Blueprints, Second. 2019.

[18] Z. He, L. Xie, X. Chen, Y. Zhang, Y. Wang, and Q. Tian, "Data Augmentation Revisited :," 2019.

[19] G. Jocher, A. Chaurasia, and J. Qiu, "YOLO by Ultralytics.," https://github.com/ultralytics/ultralytics.

[20] P. Henderson and V. Ferrari, "End-to-end training of object class detectors for mean average precision," Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), vol. 10115 LNCS, pp. 198–213, 2017, doi: 10.1007/978-3-319-54193-8_13.

[21] S. Yohananda, "What is Mean Average Precision (MAP) and how does it work," xailient.com. Accessed: Mar. 12, 2024. [Online]. Available: https://xailient.com/blog/what-is-mean-average-precision-and-how-does-it-work/

[22] D. Rosmala and V. Setyaningsih, "Tlemc (Teaching & Learning English in Multicultural Contexts) Classroom English Learning Activities: Students' Facial Expressions With a Focus on Interpersonal Meanings," vol. 5, no. 2, 2021, [Online]. Available: http://jurnal.unsil.ac.id/index.php/tlemc/index