International Journal of Transport Development and Integration
Vol. 9, No. 1, March, 2025, pp. 189-196
Journal homepage: http://iieta.org/journals/ijtdi

Optimized YOLO Approach for Drowsiness Detection in Automotive Safety: Parameter Tuning and Facial Expression Analysis

Dina Tri Utari[1*], Andrie Pasca Hendradewa[2], Mamta Anisa Bella[2]

[1] Department of Statistics, Universitas Islam Indonesia, Yogyakarta 55584, Indonesia
[2] Department of Industrial Engineering, Universitas Islam Indonesia, Yogyakarta 55584, Indonesia

Corresponding Author Email: dina.t.utari@uii.ac.id

https://doi.org/10.18280/ijtdi.090118

Received: 13 January 2025
Revised: 12 March 2025
Accepted: 17 March 2025
Available online: 31 March 2025

Keywords:
deep learning, drowsiness detection, object detection, image processing

ABSTRACT

Drowsy driving is a significant hazard, often leading to vehicular collisions, personal injuries, and fatalities. Detecting drowsiness signs quickly and accurately is crucial for reducing fatigue-related incidents. In recent years, the domain of artificial intelligence, especially the implementation of Convolutional Neural Network (CNN) frameworks in conjunction with the You Only Look Once (YOLO) algorithm, has attracted considerable academic scrutiny. These sophisticated methodologies enable the evaluation of driver fatigue through video footage or ongoing surveillance in real time. This study employs the YOLO algorithm integrated with a CNN to categorize detected drivers into drowsy and awake, utilizing bounding boxes during analysis. Model parameters, such as batch size (64), network size (416×416), subdivisions (16), max batch (4000), and filters (21), are configured for optimal performance. The dataset is split into four scenarios for training and testing, with learning rates set at 0.00261 and 0.001. Notably, the highest Intersection over Union (IoU) value is achieved with an 80%:20% split dataset and a learning rate of 0.00261, effectively identifying drowsiness in drivers and enhancing proactive safety measures.

1. INTRODUCTION

Ensuring driving safety is a fundamental expectation for anyone behind the wheel or a passenger. Nevertheless, unforeseen events, including accidents, can disrupt this expectation. In 2019, the World Health Organization reported a staggering 1.35 million global fatalities due to traffic accidents, with 20%-30% attributed to fatigue [1]. Further insight from a study by Yang et al. [2] reveals that those signs of driver fatigue manifest in distinct ways, including blinking eyes, subsequent head nodding, and frequent eye closures. However, the primary indicator of impending fatigue, as highlighted in the study, is the act of yawning. Despite the commonplace expectation of safety, understanding and addressing these subtle signs of fatigue is crucial in mitigating driving-related accidents and promoting safer transportation experiences.

Driving distractions, as identified by Yazdi and Soryani [3], are recognized triggers for accidents. Among these distractions, using mobile phones stands out, diverting drivers' attention and compromising their focus on the road. However, this study underscores that while distractions are acknowledged culprits, the primary contributors to driving accidents extend beyond mere diversions. Fatigue and drowsiness emerge as paramount factors influencing driver safety. Drowsiness, characterized as a state between wakefulness and sleep [4], induces slowed response times, suboptimal limb reflexes, and challenges in maintaining an upright head for clear visibility. This drowsy state escalates the dangers of driving, leading to thousands of annual fatalities and injuries. The compromised driving skills resulting from drowsiness increase the vulnerability to accidents, as highlighted by Higgins et al. [5]. This study sheds light on the multifaceted nature of driving risks, emphasizing the need to address distractions and the critical aspects of fatigue and drowsiness for comprehensive road safety.

Mitigating accidents caused by driver drowsiness hinges on early and precise detection [6]. This underscores the significance of drowsiness detection as a pivotal aspect of driving safety, seeking to diminish the frequency and severity of traffic accidents [7]. Crucial to effective drowsiness detection is the preliminary step of face and expression detection. Several researchers, including Sinha et al. [1], have delved into face detection to discern signs of drowsiness. Their work involved a demonstration analyzing the eye and mouth areas through a camera to determine a driver's alertness level. This study advocates for advancements, proposing using higher-quality cameras to enhance image capture for improved system performance. Additionally, incorporating audio in each video frame is suggested, marking an innovative approach to bolster the efficacy of drowsiness detection systems.

Advancements in technology have spurred significant research in recent decades, particularly in artificial intelligence applied to drowsiness detection in drivers [8]. This exploration

has given rise to diverse methodologies harnessing deep learning techniques, with drowsiness detection emerging as a prominent object detection application. Object detection, a vital facet of this technological evolution, employs deep learning to ascertain the presence of predefined object categories within an image. This process involves utilizing bounding boxes to precisely delineate the areas occupied by each detected object [9]. As technology continues to evolve, these developments in artificial intelligence and object detection methodologies underscore their pivotal role in enhancing safety systems, particularly in the critical domain of drowsiness detection for drivers. Despite these advancements, many existing drowsiness detection systems struggle with performance, accuracy, and robustness in handling ambiguous facial expressions or varying lighting conditions.

To address these limitations, this research presents an innovative approach to mitigating vehicle accidents from drowsiness symptoms. It introduces an early warning system designed to alert drivers promptly. The system employs image processing with a focus on analyzing facial images. It is built on a deep learning model using the You Only Look Once (YOLO) algorithm, implemented in Python. This combination of image analysis and advanced algorithmic modeling aims to provide an effective and timely alert mechanism, contributing to enhanced driver safety and accident prevention.

Unlike conventional methodologies that necessitate multiple iterations over a given image, YOLO executes the comprehensive analysis of the entire image through a singular forward pass within the neural network framework. This distinctive feature enables it to attain elevated detection velocities, rendering it particularly advantageous for applications requiring prompt feedback, such as surveillance systems and autonomous vehicular technology [10, 11]. Furthermore, the architectural design of YOLO is designed to concurrently predict bounding boxes and class probabilities, thus augmenting its operational efficiency. This concurrent prediction mechanism alleviates the computational load. It facilitates expedited processing times when juxtaposed with alternative object detection methodologies, such as Fast R-CNN and SSD, which depend on region proposal networks [11, 12]. The capability to generate predictions within a global context significantly minimizes background inaccuracies, enhancing overall detection precision [11]. Beyond rapidity, YOLO has evidenced resilience across diverse applications, encompassing aerial image analysis and environmental surveillance.

In the scenario of drowsiness detection, all these advantages become particularly significant. Traditional methods generally struggle to monitor subtle facial variations in real-time, leading to missed or time-delayed notifications. By capitalizing on the ability of YOLO to process images with a single pass, the resulting system effectively recognizes significant facial cues for drowsiness, such as eye closing, yawning, and head orientation, with minimal computational overhead. This efficiency enables the system to timely alert drivers, preventing reaction time and potential accidents. Second, incorporation of CNN-based feature extraction encourages the model to distinguish between drowsiness expressions and normal facial changes, improving detection accuracy. By overcoming the limitation of existing approaches, this work contributes to enhancing real-time drowsiness detection technology by offering a robust solution for enabling automotive safety.

2. LITERATURE REVIEW

The main objective of this paper is to minimize the risk of loss of driving focus caused by drowsiness and the need for an early detection process of drowsiness symptoms in the driver's activity through the utilization of physiological sign changes in facial expressions. The main concept of drowsiness detection is using image classification with deep learning. One of the most common deep-learning models for image classification is the YOLO approach.

Various solutions for analyzing and modeling the drowsy driving state have been studied using different methods. Redmon et al. [11] introduced the YOLO approach. This unified, real-time object detection method outperforms other detection methods when generalizing from natural images to other domains. This reference highlights the potential of YOLO for real-time object detection, which can be instrumental in detecting facial expressions indicative of drowsiness in automotive safety applications. Siddiqui et al. [13] proposed a non-invasive driver drowsiness detection system utilizing support vector machine and computer science techniques. The study emphasizes the severe consequences of drowsiness-related road accidents, underscoring the critical need for effective drowsiness detection systems in automotive safety. Han et al. [14] focused on the design of a scalable and fast YOLO for edge-computing devices, highlighting the potential for YOLO to be used in developing network models that are faster and easier than other methods. This reference underscores the applicability of YOLO in edge-computing environments, which is crucial for real-time drowsiness detection in automotive safety systems.

Research by Sinha et al. [1] conducted a comparative analysis of Viola Jones, DLib, and YOLO algorithms for detecting sleepiness. The conclusion favored YOLO, highlighting its superior accuracy to Viola Jones and DLib. Al-Sabban [15] introduced a real-time driver drowsiness detection system using DLib based on driver eye/mouth monitoring technology, employing a CNN model with inputs based on images related to eye and mouth openings and closings. This study demonstrates the feasibility of using CNN-based models for real-time drowsiness detection, aligning with the objectives of leveraging CNN-YOLO for facial expression analysis.

Another study explores YOLO used for agricultural monitoring, focusing on detecting plant diseases and calculating the infected area. The methodology combines the YOLOv4 object detection algorithm with ArUco Marker reference images to measure infected and healthy regions accurately. This approach helps localize infections, prevent their spread, and mitigate the risk of crop failure. The evaluation showed a high accuracy of 97.05% when comparing the detected area to the actual area, demonstrating the effectiveness of this technique [16].

The study on transportation systems utilizes surveillance cameras and YOLO object detection technology to analyze traffic information, ensuring accurate counting and classification in various traffic conditions [17]. The system achieves effective object detection by leveraging image processing and deep convolutional neural networks (CNNs). In the other study, YOLOv3-SPP demonstrates the highest classification accuracy among the tested models across all road conditions [18].

A machine learning-based driver monitoring system can achieve accuracy, precision, and recall values [19]. The study

underscores the effectiveness of machine learning in driver monitoring, which can be extended to drowsiness detection using CNN-YOLO for automotive safety. The utilization of CNN-YOLO in drowsiness detection systems marks a significant advancement in automotive safety. Leveraging the capabilities of CNN-YOLO, this approach aims to precisely identify and interpret facial expressions indicative of drowsiness, enabling real-time monitoring of driver alertness.

Recent advancements in drowsiness detection systems based on facial expressions have gained momentum, supported by deep-learning-based technology. The blink and yawn detection using CNNs has demonstrated high accuracy in real-world driving videos, enhancing the reliability of drowsiness detection systems [20]. Additionally, facial landmark detection combined with CNNs allows for more precise analysis of multiple facial traits, such as expressions and head poses, to assess a driver's drowsiness level. Further improvements in deep learning algorithms have addressed the limitations of traditional methods, enabling real-time detection through advanced image processing techniques [21, 22].

This study aims to explore the viability and effectiveness of harnessing YOLO based on CNN, combined with machine learning and artificial intelligence, to craft a cutting-edge drowsiness detection system for automotive safety. The facial features and expressions analyzed include key indicators such as eye closure, yawning, and head position, which are essential for detecting a driver's drowsiness.

3. MATERIALS AND METHODS

This study used secondary data sourced from YawDD: Yawning Detection Dataset [23] and Driver Drowsiness Dataset (D3S) [24], which are openly licensed for unrestricted use. The dataset employed is labeled into two categories: drowsy and awake. The entire image dataset is stratified into subsets for training and testing to optimize the model's performance. The chosen dataset comprises images featuring the driver's face on the car's steering wheel, categorized into expressions of drowsy and awake, as shown in Figure 1. Furthermore, additional data from diverse prior studies that align with similar research objectives enriches the study with a more comprehensive range of models and outcomes.

Figure 1. Images of drowsy and awake faces

3.1 You Only Look Once (YOLO)

The YOLO (You Only Look Once) algorithm, pioneered by [11], represents a notable advancement in object detection. Introduced alongside a proprietary framework called Darknet, YOLO is distinctive for its use of a single CNN, contributing to its remarkable speed in object detection. Unlike traditional methods, YOLO employs a specialized CNN for the simultaneous classification and localization of multiple objects within a single image. While compromising some accuracy, this efficiency still outperforms other object detection algorithms like R-CNN [25]. Compared to alternatives such as Faster R-CNN and SSD, YOLO's real-time processing capability makes it particularly well-suited for applications where speed is crucial. In the context of drowsiness detection for automotive safety, the ability to rapidly analyze facial features is essential for timely alerts. YOLO's balance between speed and accuracy makes it a practical choice for this study, ensuring efficient detection of drowsiness-related facial expressions with minimal computational overhead.

The YOLO combines candidate box extraction, feature extraction, and object classification within a unified neural network architecture. It revolutionizes object detection by directly extracting candidate boxes from images and discerning objects across the entire image feature space. The YOLO network strategically partitions the image into an $s \times s$ grid, evenly distributing bounding boxes along the X and Y axes. This approach facilitates region-based predictions, where the network analyzes each region's bounding box location, confidence level, and class probability, as delineated by reference [26]. The innovative YOLO framework thus optimizes the object detection process by unifying these critical elements into a cohesive and efficient neural network.

In the spatially segmented image with a grid size of sxs, each grid cell is pivotal in predicting bounding boxes and confidence scores [11]. The confidence score is a crucial indicator, reflecting the model's certainty in detection and measuring the accuracy of object identification within the bounding box. Within every bounding box, five predicted values are generated: x, y, w, h, and confidence. The (x, y) coordinates denote the center of the box relative to the grid cell boundary, while (w, h) represent the width and height of the entire image. The confidence score, in turn, quantifies the IoU between the predicted box and the ground truth box. IoU serves as an evaluation metric, gauging the accuracy of the object detector based on a pre-trained dataset. This comprehensive approach within the YOLO framework enhances the precision and reliability of object detection.

$$IoU = \frac{Area\ of\ overlap}{Area\ of\ union} \tag{1}$$

An IoU score of 1 indicates an entirely accurate prediction. Conversely, as the IoU score decreases, the prediction quality worsens. Figure 2 illustrates the variation in IoU scores in determining the prediction accuracy.

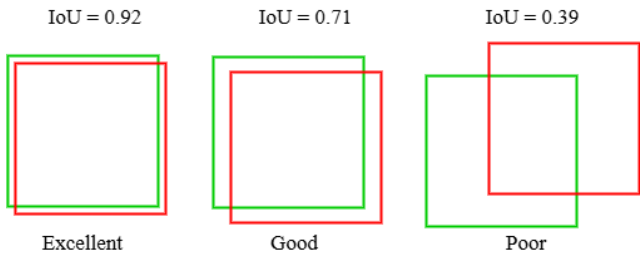


Figure 2. IoU illustration

If the bounding box does not detect any object, the confidence score is set to 0.

$$Confidence\ score = P(Class) * IoU \tag{2}$$

The class confidence score assesses the confidence level within each bounding box, indicating the probability of the class's presence in the box and specifying the predicted value. The computation of the class confidence score follows the formulation provided by the study [11].

The YOLOv4, introduced by reference [27], incorporates several enhancements to bolster detection accuracy and efficiency. This version adopts the CSPDark-net53 CNN backbone architecture at its core, boasting 162 layers meticulously configured for optimal performance with input images, as outlined in the research. The illustration of the YOLO process, its architecture with the Darknet framework, and the consecutive phase in detecting drowsiness are shown in Figures 3-5, respectively.

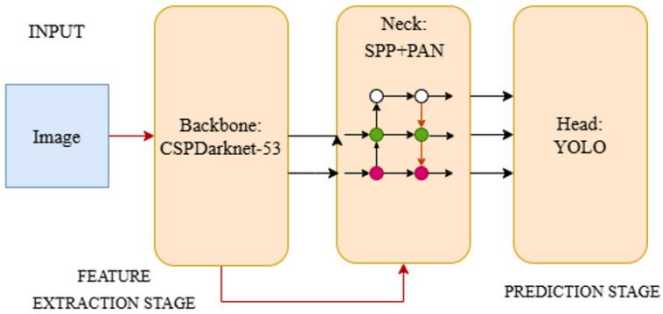


Figure 3. Simple YOLOv4 process

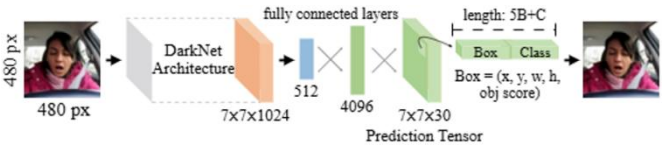


Figure 4. The YOLOv4 architecture with Darknet framework

Figure 5. Consecutive phases of the YOLOv4 model were implemented to detect drowsiness

YOLOv4 was unveiled following two years of incremental enhancements over YOLOv3 [28], capitalizing on the latest developments in deep learning. It attains an accuracy level of 43.5% Average Precision (AP) on the MS COCO dataset, outperforming YOLOv3, which achieves 33.0% AP. Notably, this heightened accuracy is achieved while maintaining a highly efficient inference time of 65 frames per second (FPS) on the Tesla V100. YOLOv4 is designed to ensure effective and seamless object detection on the cost-effective hardware commonly found in most edge devices [29]. The feature and architecture comparison between two different versions of YOLO is shown in Table 1.

Table 1. Comparison between YOLO version 3 and 4 [29]

	YOLOv3	YOLOv4
Stages	Simultaneous regression of bounding boxes and classification	Simultaneous regression of bounding boxes and classification
Type of neural network	Fully convolutional	Fully convolutional
Backbone feature extractor	Darknet-53 (53 convolutional layers)	CSPDarknet53 (53 convolutional layers)
Detection of location	Anchor-based (dimension clusters)	Anchor-based
Quantity of anchor boxes	A singular bounding box prior is assigned to each ground-truth object	Utilizing multiple anchors for one ground truth
Standard sizes of anchors.	(10,13), (16,30), (33,23), (30,61), (62,45), (59,119), (116,90), (156,198), (373,326)	(12,16), (19,36), (40,28), (36,75), (76,55), (72,146), (142,110), (192,243), (459,401)
IoU thresholds	One (at 0.5)	One (at 0.213)
Loss function	Binary cross-entropy loss	Complete IoU loss
Input size	Various potential input dimensions (n × n with n being a multiple of 32)	Various potential input dimensions (n × n with n being a multiple of 32)
Momentum	Default value: 0.9	Default value: 0.949
Weight decay	Default value: 0.0005	Default value: 0.0005.
Batch size	Default value: 64	Default value: 64

3.2 Dataset and workflow

The dataset used for training and evaluating the drowsiness detection model consists of labeled images depicting various facial expressions associated with drowsiness and alertness. The dataset was collected from publicly available sources and supplemented with additional real-world driving scenarios to enhance diversity. The collected data was then preprocessed to ensure the input quality. The overall workflow is shown in Figure 6.

To ensure a well-balanced dataset, the following preprocessing steps were applied: *Image Resizing*: All images were resized to 480×480 pixels, the input size required for YOLO. *Normalization*: Pixel values were normalized to the range [0,1] to improve training stability. *Dataset Splitting*: The dataset was split into different sets of training and testing.

The YOLO architecture was used due to its real-time object detection capabilities. The model was fine-tuned on our dataset using transfer learning from pre-trained weights. Training was conducted using the following hyperparameters:

- Batch Size: 16
- Optimizer: Adam with learning rate = 0.001
- Loss Function: Binary Cross-Entropy (for drowsiness classification)
- Epochs: 50 (with early stopping based on validation loss)

The hyperparameters were chosen based on previous research. A batch size between 16-32 balances convergence speed and accuracy, especially for facial recognition tasks that require fine feature extraction [21]. Recent studies suggest using adaptive learning rate methods like Adam, which adjusts

the learning rate according to the gradient [20].

To assess model performance, we used the following evaluation metrics:

- Precision (P): Measures the percentage of correctly predicted drowsy cases among all drowsy predictions.
- Recall (R): Indicates how well the model identifies actual drowsy cases.
- F1-Score: The harmonic means of Precision and Recall, providing a balanced measure of model performance.
- Accuracy: Overall percentage of correct predictions.

These metrics were calculated on the test set to evaluate generalization performance.

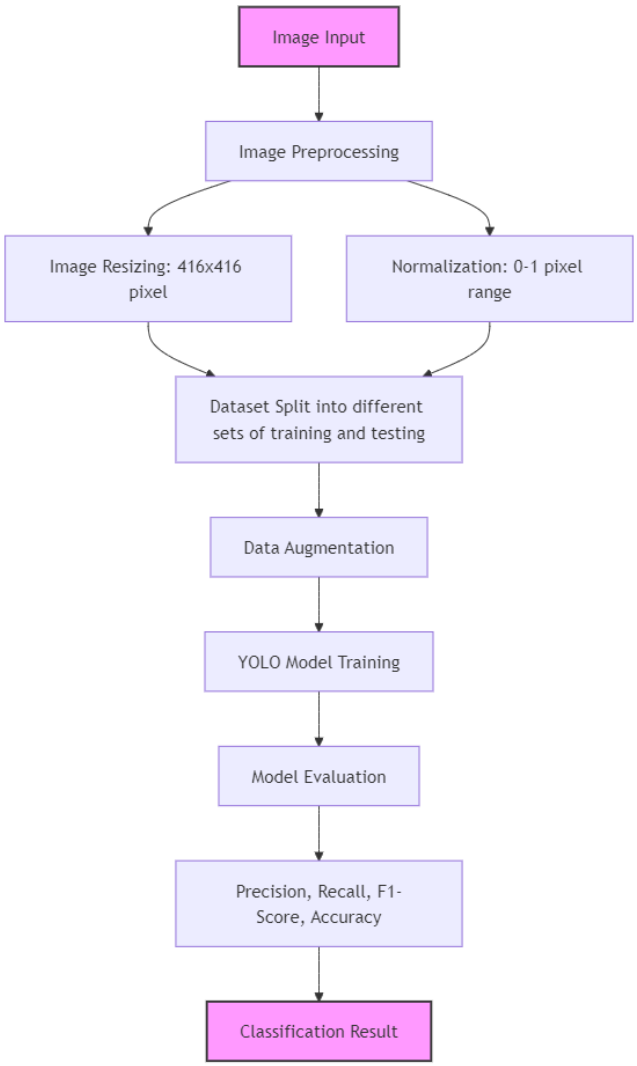


Figure 6. Image classification workflow

4. RESULTS AND DISCUSSION

The dataset is categorized into drowsy or awake, and the resulting labels are applied to images in the YOLO format, which are then saved as .txt files. Subsequently, the process involves generating bounding boxes using the "Create RectBox" feature in the LabelImg application. The dataset is annotated by drawing a box around the image, with images depicting drowsy drivers labeled accordingly and images featuring awake drivers marked with the appropriate label, as illustrated in Figures 7 and 8.

The Darknet framework undergoes model parameter configuration to align with the upcoming model training. In Table 2, modifications applied to both the original parameters of YOLOv4-tiny and the parameters adapted for this research in YOLOv4-tiny are outlined. This model parameter configuration is guided by the YOLOv4 developer's specifications [27]. The configured settings will be used for four data training sessions, each detailed as follows (Table 2).

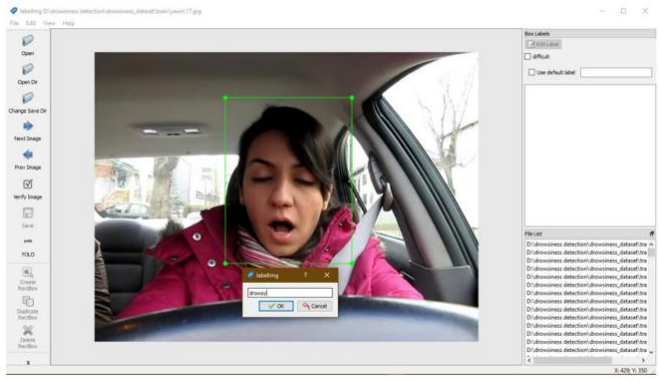


Figure 7. Making bounding boxes for drowsy class

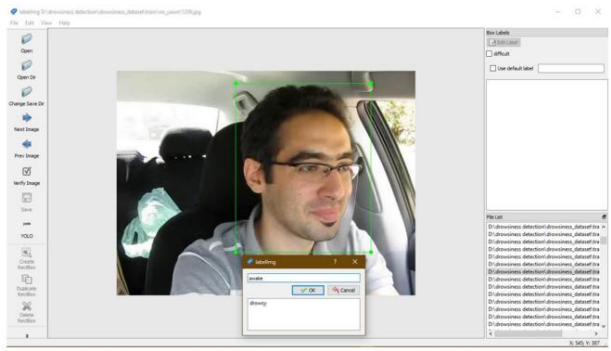


Figure 8. Making bounding boxes for awake class

Table 2. Model parameter configuration

Scenario	Learning Rate	Dataset Split
Scenario 1	0.00261	90:10
Scenario 2	0.001	90:10
Scenario 3	0.00261	80:20
Scenario 4	0.001	80:20

All training sessions utilized uniform network architectures (batch size: 64, image resolution: 416×416, subdivisions: 16, maximum batches: 4000, filters: 21, and class count: 2), whereas modifications were implemented in the learning rate (0.001 versus 0.00261) and dataset partition ratios (90:10 and 80:20).

The split dataset determines the percentage division between training and testing data. The batch size denotes the number of images processed per iteration during training. A higher batch value accelerates the training process and increases the GPU workload. Network size refers to the dimensions of images used for model training. Larger image sizes slow down training and demand more GPU memory, while smaller sizes speed up training but may compromise object recognition due to reduced size.

Subdivisions are utilized to divide batches into mini batches, optimizing GPU memory usage to prevent runtime crashes during training. The max batch represents the maximum

number of iterations for network training, set at 4000 iterations in this research. The number of filters in each layer pre-convolution is adjusted based on the classes being trained, with two classes utilized in this study: "drowsy" (class_id = 0) and "awake" (class_id = 1). The formula for determining the number of filters is expressed as follows:

$$F = (C + Co + P) * M \quad (3)$$

where, F is the number of filters, C is the number of classes, Co is the number of coordinates used (x, y, w, h), P is the object value of a proposed area, and M is the number of anchor boxes used. Thus, this study adjusts the number of filters to (2+4+1) * 3 = 21.

The learning rate, a crucial parameter controlling the model's learning pace during training, influences the convergence to optimal weights. A too-small learning rate prolongs the time to reach optimal weights, while a huge one hinders the attainment of optimal values. This study selected learning rate values of 0.001 and 0.00261 based on the training dataset scenario.

The training dataset is instrumental in enabling the model to conduct detection based on the established configuration outlined in Table 2. This study meticulously executes a training dataset scenario involving a comparative analysis of the split dataset and learning rate. This comparison yields output metrics, precisely average IoU values, and loss values, presented in detail in Table 3.

Table 3. Comparison of output

Scenario	Avg IoU (%)	Avg Loss
Scenario 1	78.15	0.0286
Scenario 2	74.40	0.0594
Scenario 3	84.62	0.0154
Scenario 4	79.14	0.0553

IoU serves as a metric assessing the system's precision in detecting objects within the trained dataset. It accomplishes this by comparing the ground truth or objects within the image with the predicted bounding box generated by the model. A higher IoU value signifies a closer alignment of the bounding box to the ground truth, indicating a superior detection performance. Conversely, the loss value gauges the errors made by the network, aiming to minimize these errors. A lower loss value signifies fewer errors committed by the model.

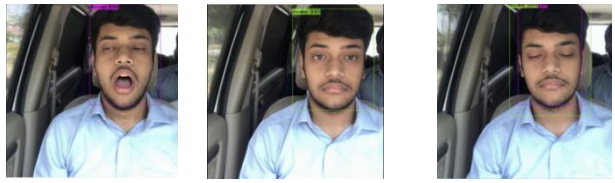
Upon scrutinizing the training data, as presented in the comparative analysis in Table 3, the split dataset configuration of 80%:20% coupled with a learning rate of 0.00261 yields the highest IoU value and the lowest loss value compared to other training datasets. This implies that, in this study, configuring parameters through a split dataset with specific learning values results in testing dataset values with enhanced accuracy compared to alternative configurations.

Image datasets of drowsy and awake face during driving [24] were utilized during the testing stages. The testing process involved feeding images into the model, and four specific scenarios were implemented using the parameter configuration outlined in Table 2. The result analysis of four scenarios showed in Table 3 is further explained as follows:

4.1 Scenario 1 (Split Dataset 90:10, Lr: 0.00261)

We have three test images prepared for evaluating the model.

The detection outcomes demonstrate a 94% accuracy in identifying drowsy characteristics, achieved at a prediction speed of 15.67 ms, as illustrated in Figure 9. The accuracy value reflects the system's confidence level in conducting detection. In the subsequent test using an image featuring open eyes and a closed mouth, indicative of an awake driver, the detection results exhibit an 83% accuracy, accompanied by a prediction speed of 15.48 ms. The following test involved images portraying closed eyes and mouth, characteristic of a drowsy driver. However, the system exhibited a dual detection, identifying 64% as awake and 34% as drowsy, with a prediction speed of 15.52 ms. This detection bias may arise from the system's unfamiliarity with specific image patterns, underscoring the importance of comprehensive training to mitigate instances of double detection.

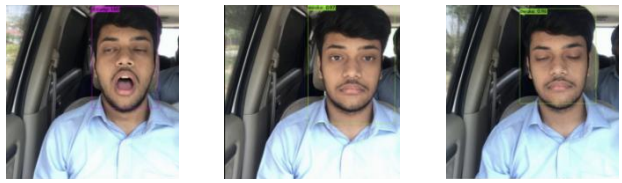


Drowsy 94% | Awake 83% | Drowsy 64%; Awake 34%

Figure 9. Testing on image using scenario 1

4.2 Scenario 2 (Split Dataset 90:10, Lr: 0.001)

The detection outcomes depicted in Figure 10 show a 100% accuracy in identifying drowsiness.



Drowsy 100% | Awake 97% | Awake 96%

Figure 10. Testing on image using scenario 2

In Figure 10, the detection results reveal a 100% accuracy in detecting a drowsy state and 97% accuracy in recognizing an awake state. However, a detection accuracy of 96% in the third image for an awake state was misclassified with the expected result, as the image should have been identified as drowsy. This discrepancy can be attributed to the system's criterion, which recognizes an image as drowsy when the eyes and mouth are closed. Consequently, mouth detection leads to the classification as an awake driver in the image featuring closed eyes.

4.3 Scenario 3 (Split Dataset 80:20, Lr: 0.00261)

Scenario 3 represents a configuration characterized by the highest IoU value and the slightest loss value, crucial factors influencing the system's accuracy in detection.

As illustrated in Figure 11, the results show 100% detection accuracy for drowsy drivers. Subsequently, the second image exhibits a 100% detection accuracy for awake drivers. However, in the last image featuring a drowsy facial expression, the system erroneously detects 100% wakefulness. This discrepancy arises from the system's reliance on detecting open eyes and a closed mouth to identify drowsiness,

leading to misclassification in cases where both the eyes and mouth are closed.

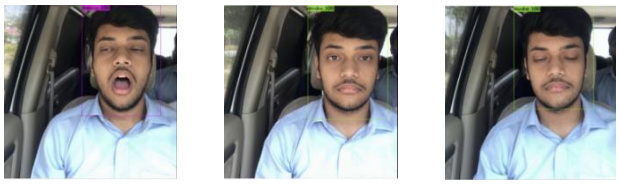


Figure 11. Testing on image using scenario 3

4.4 Scenario 4 (Split Dataset 80:20, Lr: 0.001)

Test in the fourth scenario demonstrates detection results with 98% accuracy for drowsy drivers as illustrated in Figure 12. The detection results indicate 95% accuracy in recognizing awake drivers. However, similar to scenarios 2 and 3, the system misclassifies the third image as 98% awake, as both the eyes and mouth are closed, leading to an incorrect detection of an awake driver.

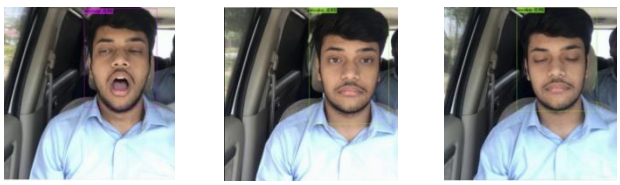


Figure 12. Testing on third image on scenario 4

The results demonstrate the model's ability to achieve high accuracy in detecting drowsiness and wakefulness in clear cases, with peak performance observed in Scenarios 1 and 2. High accuracy (94%-100%) for images featuring distinct features, such as fully closed eyes or open eyes, highlights the system's effective feature recognition. However, frequent misclassifications, particularly in cases where both eyes and mouth were closed, reveal a limitation in the model's decision-making criteria. These errors suggest the system overly relies on the absence of “awake” features (e.g., open eyes) rather than a nuanced analysis of "drowsy" characteristics, leading to confusion in ambiguous scenarios.

The impact of parameter configurations is also evident, with scenario 1 that uses a 90:10 dataset split and optimal learning rate (0.00261) outperformed others. Larger training datasets enhanced the model’s ability to generalize, while the ideal learning rate balanced learning speed and stability. Misclassifications in scenarios with a lower training ratio (80:20) or suboptimal learning rate (0.001) underline the importance of these parameters in fine-tuning the system’s performance. These findings emphasize the need for improved feature extraction and robust criteria to handle diverse or ambiguous facial expressions.

Although the proposed approach has promising results, there are certain limitations that need to be addressed. The system suffers from difficulties in classifying ambiguous facial expressions, such as slight eye closure or weak yawning, which can compromise the performance of drowsiness detection. Additionally, environmental factors like changing illumination and occlusions (e.g., glasses or headbands) can affect the performance of the model. These problems highlight the need for more research to enhance the robustness and generalizability of the model.

5. CONCLUSIONS

The results of the work support the efficient application of a CNN-based YOLO algorithm in detecting driver drowsiness. The performance analysis of different model configurations, i.e., different batch sizes, network sizes, subdivision levels, training scenario numbers, and learning rates, revealed that maximum performance was achieved by using an 80:20 dataset proportion and a learning rate of 0.00261, yielding the highest IoU value. These findings verify the effectiveness of YOLO in detecting real-time drowsiness facial expressions.

With the continued development of real-time camera technology, GPUs, and deep learning algorithms coming on the scene, the accuracy and responsiveness of sleepiness detection systems will presumably rise. Integration of this approach into existing motor vehicle safety infrastructure, such as Advanced Driver Assistance Systems (ADAS), may be used to enhance real-time monitoring of drivers as well as prevention measures for accidents. There may be some future work to further improve the model for embedded applications, detection stability in other environments, and integration of further sensor data for reliability improvement.

This work demonstrates the feasibility of YOLO-based drowsiness detection, which lends itself to the growth of AI-powered driver monitoring solutions. It presents a scalable and efficient solution for preventing fatigue-related road accidents and improving transport safety.

REFERENCES

[1] Sinha, A., Aneesh, R.P., Gopal, S.K. (2021). Drowsiness detection system using deep learning. In 2021 Seventh International Conference on Bio Signals, Images, and Instrumentation (ICBSII), Chennai, India, pp. 1-6. https://doi.org/10.1109/ICBSII51839.2021.9445132

[2] Yang, H., Liu, L., Min, W., Yang, X., Xiong, X. (2020). Driver yawning detection based on subtle facial action recognition. IEEE Transactions on Multimedia, 23: 572-583. https://doi.org/10.1109/TMM.2020.2985536

[3] Yazdi, M.Z., Soryani, M. (2019). Driver drowsiness detection by identification of yawning and eye closure. Automotive Science and Engineering, 9(3): 3033-3044. https://doi.org/10.22068/ijae.9.3.3033

[4] Chaabene, S., Bouaziz, B., Boudaya, A., Hökelmann, A., Ammar, A., Chaari, L. (2021). Convolutional neural network for drowsiness detection using EEG signals. Sensors, 21(5): 1734. https://doi.org/10.3390/s21051734

[5] Higgins, J.S., Michael, J., Austin, R., Åkerstedt, T., Van Dongen, H.P., Watson, N., Rosekind, M.R. (2017). Asleep at the wheel—The road to addressing drowsy driving. Sleep, 40(2): zsx001. https://doi.org/10.1093/sleep/zsx001

[6] Gwak, J., Hirao, A., Shino, M. (2020). An investigation of early detection of driver drowsiness using ensemble machine learning based on hybrid sensing. Applied Sciences, 10(8): 2890. https://doi.org/10.3390/app10082890

[7] Ma, Y., Zhang, S., Qi, D., Luo, Z., Li, R., Potter, T., Zhang, Y. (2020). Driving drowsiness detection with EEG using a modified hierarchical extreme learning machine algorithm with particle swarm optimization: A

pilot study. Electronics, 9(5): 775. https://doi.org/10.3390/electronics9050775

[8] Faraji, F., Lotfi, F., Khorramdel, J., Najafi, A., Ghaffari, A. (2021). Drowsiness detection based on driver temporal behavior using a new developed dataset. arXiv preprint arXiv:2104.00125. https://doi.org/10.48550/arXiv.2104.00125

[9] Liu, L., Ouyang, W., Wang, X., Fieguth, P., Chen, J., Liu, X., Pietikäinen, M. (2020). Deep learning for generic object detection: A survey. International Journal of Computer Vision, 128: 261-318. https://doi.org/10.1007/s11263-019-01247-4

[10] Andrean, D., Unik, M., Rizki, Y. (2023). Hotspots and smoke detection from forest and land fires using the YOLO algorithm (You Only Look Once). JIM-Journal International Multidisciplinary, 1(1): 46-56. https://doi.org/10.58794/jim.v1i1.410

[11] Redmon, J., Divvala, S., Girshick, R., Farhadi, A. (2016). You only look once: Unified, real-time object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, pp. 779-788. https://doi.org/10.1109/CVPR.2016.91

[12] Kadhum, A.N., Kadhum, A.N. (2023). Literature survey on YOLO models for face recognition in COVID-19 pandemic. Journal of Image Processing and Intelligent Remote Sensing, 3(4): 27-35. https://doi.org/10.55529/jipirs.34.27.35

[13] Siddiqui, H.U.R., Saleem, A.A., Brown, R., Bademci, B., Lee, E., Rustam, F., Dudley, S. (2021). Non-invasive driver drowsiness detection system. Sensors, 21(14): 4833. https://doi.org/10.3390/s21144833

[14] Han, B.G., Lee, J.G., Lim, K.T., Choi, D.H. (2020). Design of a scalable and fast YOLO for edge-computing devices. Sensors, 20(23): 6779. https://doi.org/10.3390/s20236779

[15] Al-Sabban, W.H. (2022). Real-time driver drowsiness detection system using Dlib based on driver eye/mouth monitoring technology. Communications in Mathematics and Applications, 13(2): 807-822. https://doi.org/10.26713/cma.v13i2.2034

[16] Masykur, F., Adi, K., Nurhayati, O.D. (2024). Measuring agricultural area using YOLO object detection and ArUco Markers. Ingénierie des Systèmes d'Information, 29(1): 95-106. https://doi.org/10.18280/isi.290111

[17] Huo, J.L., Shi, B.J., Zhang, Y.H. (2023). An object detection method for the work of an unmanned sweeper in a noisy environment on an improved YOLO algorithm. Signal, Image and Video Processing, 17(8): 4219-4227. https://doi.org/10.1007/s11760-023-02654-4

[18] Wu, J.D., Chen, B.Y., Shyr, W.J., Shih, F.Y. (2021). Vehicle classification and counting system using YOLO object detection technology. Traitement du Signal, 38(4): 1087-1093. https://doi.org/10.18280/ts.380419

[19] Ziryawulawo, A., Kirabo, M., Mwikirize, C., Serugunda, J., Mugume, E., Miyingo, S.P. (2023). Machine learning based driver monitoring system: A case study for the Kayoola EVS. SAIEE Africa Research Journal, 114(2): 40-48. https://doi.org/10.23919/SAIEE.2023.10071976

[20] Muralidharan, J., Sindhuja, L.P., Srinivasan, S., Sunshetha, K.V., Surendhran, P. (2023). Smart safety and accident prevention system. In E3S Web of Conferences, Tamil Nadu, India, pp. 01006. https://doi.org/10.1051/e3sconf/202339901006

[21] Singh, D., Singh, A. (2023). Enhanced driver drowsiness detection using deep learning. In ITM Web of Conferences, Kurukshetra, India, pp. 01011. https://doi.org/10.1051/itmconf/20235401011

[22] Albasrawi, R., Fadhil, F.F., Ghazal, M.T. (2022). Driver drowsiness monitoring system based on facial Landmark detection with convolutional neural network for prediction. Bulletin of Electrical Engineering and Informatics, 11(5): 2637-2644. https://doi.org/10.11591/eei.v11i5.3966

[23] Abtahi, S., Omidyeganeh, M., Shirmohammadi, S., Hariri, B. (2020). YawDD: Yawning Detection Dataset. IEEE Dataport. https://doi.org/10.21227/e1qm-hb90

[24] Gupta, I., Garg, N., Aggarwal, A., Nepalia, N., Verma, B. (2018). Real-time driver's drowsiness monitoring based on dynamically varying threshold. In 2018 Eleventh International Conference on Contemporary Computing (IC3), Noida, India, pp. 1-6. https://doi.org/10.1109/IC3.2018.8530651

[25] Tao, J., Wang, H., Zhang, X., Li, X., Yang, H. (2017). An object detection system based on YOLO in traffic scene. In 2017 6th International Conference on Computer Science and Network Technology (ICCSNT), Dalian, China, pp. 315-319. https://doi.org/10.1109/ICCSNT.2017.8343709

[26] Liu, C., Tao, Y., Liang, J., Li, K., Chen, Y. (2018). Object detection based on YOLO network. In 2018 IEEE 4th Information Technology and Mechatronics Engineering Conference (ITOEC), Chongqing, China, pp. 799-803. https://doi.org/10.1109/ITOEC.2018.8740604

[27] Bochkovskiy, A., Wang, C.Y., Liao, H.Y.M. (2020). YOLOv4: Optimal speed and accuracy of object detection. arXiv preprint arXiv:2004.10934. https://doi.org/10.48550/arXiv.2004.10934

[28] Redmon, J., Farhadi, A. (2018). YOLOv3: An incremental improvement. arXiv preprint arXiv:1804.02767. https://doi.org/10.48550/arXiv.1804.02767

[29] Ammar, A., Koubaa, A., Ahmed, M., Saad, A., Benjdira, B. (2021). Vehicle detection from aerial images using deep learning: A comparative study. Electronics, 10(7): 820. https://doi.org/10.3390/electronics10070820