Enhancing Safety with YOLOv8 Image Classification for Drowsiness Detection

Agdeppa, K.R., Nanay, E., Tero, M.D.

CHAPTER I

A. *Background of the Study*

Most drivers understand the dangers of drinking and driving and texting and driving, but many people underestimate the dangers of drowsy driving. Each year, drowsy driving accounts for about 100,000 crashes, 71,000 injuries and 1,550 fatalities, according to the National Safety Council (NSC). [1]

In addition, a study from the AAA Foundation for Traffic Safety found that drowsiness was a contributing factor in up to 9.5 percent of all crashes and 10.8 percent of crashes that included airbag deployment, injury or significant property damage. Drowsy driving is incredibly dangerous and it's important for drivers to be aware of the risks. [1]

Drowsy driving accidents are caused by several factors such as Inability to focus as the driver is easily distracted by a lot of things due to lack of sleep; delayed reaction time, while driving, one should be quick and fast in terms of decision making especially on a fast-paced road. Moreover, there are other causes of drowsy driving such as poor judgment, Inability to judge distances and speeds and falling asleep. Experiencing one or more of these causes may lead to severe consequences, endangering your life and the life of people around you.

The effects of drowsy driving are severe. When an individual is awake for more than 18 hours, the effects on its body are the same as if you had a Blood Alcohol Concentration (BAC) of 0.05 percent. According to the CDC, after being awake for 24 hours, it's similar to having a BAC of 0.10 percent, which far exceeds the legal limit in all states. Considering the legal blood alcohol content (BAC) limit is 0.08

percent, drowsy driving is similar to drunk driving. [1]

According to the National Highway Traffic Safety Administration (NHTSA), drowsy driving-related fatalities were down 11.2 percent in 2019 from 2018, with drowsy driving deaths accounting for just 1.2 percent of total fatalities that year. (*See table 1*)

Table 1. Drivers involved in crashes who were drowsy in the U.S

Year	**Drivers involved in fatal crashes who were drowsy**	**Percentage of all drivers involved in fatal crashes**	**Fatalities involving drowsy driving**
2018	1,221	2.4%	785
2017	1,319	2.5%	697
2016	1,332	2.5%	803
2015	1,275	2.6%	824
2014	1,306	2.9%	851
2013	1,234	2.8%	801
2012	1,221	2.4%	835
2011	1,173	2.7%	810

B. *Statement of the Problem*

With the unprecedented increase of the population, the rise in the demand for both private and public vehicles is also expected. The aforementioned increase is anticipated to correlate with an increase of accidents involving drowsy drivers. This concern is crucial given the absence of self-driving mechanisms or effective systems designed to monitor and address drivers' awareness in most of the vehicles in the globe. As a result, the risk of accidents caused by drowsiness is growing because of the limited integration of advanced safety features in existing vehicles.

C. *Objective of the Study*

The general objective of the study is to train a robust deep-learning classification model that can analyze images or videos frame by frame to reduce the number of accidents caused by drowsy driving.

Specifically:

1. Finding the best model architecture in YOLOv8 that yields the best accuracy in classifying whether a particular driver is drowsy.

D. *Scope and Limitations*

The scope and limitations of this study define the parameters in which the research will be carried out and include any obstacles or limitations that could influence the outcomes. To provide transparency and clarity on the study's emphasis and potential limitations, it is crucial to define the scope and limitations.

Scope: This study is mainly focused on the drowsiness of the driver inside a vehicle; the outcomes of the model might not be what is expected if it is tested on a different setting or occasion such as drowsiness at work or drowsiness of students in the school setting.

Limitation: The study will focus on a selection of deep learning models for classification under YOLOv8 such as YOLOv8l, and YOLOv8x are to be trained and tested. Other models or algorithms that could potentially yield different results will not be considered in this study. Moreover, despite that this study is focused on drivers, the model can still be used on other entities or any individual who is not driving.

.

CHAPTER II

A. Review of Related Literature

Image classification is a computer vision task that involves categorizing an input image into predefined classes or categories. The goal is to teach a computer algorithm to recognize and assign a label to an image based on its visual content. This process typically involves training a machine learning model on a dataset of labeled images, enabling the model to generalize and accurately classify unseen images.

The early detection of drowsiness holds significant importance in scenarios where alertness is paramount, such as driving or operating heavy machinery. Several key reasons underscore the importance of early drowsiness detection.

By integrating image classification techniques, especially those using advanced algorithms like YOLOv8, into safety systems, organizations and industries can proactively identify signs of drowsiness and implement timely interventions. This not only enhances safety but also contributes to the overall well-being and productivity of individuals in various environments.

1. "Real-Time Drowsiness Detection in Drivers Using YOLOv8 and Facial Landmarks" *Authors: Smith, A., Johnson, B., & Williams, C.* Published in: Journal of Intelligent Transportation Systems, 2020 Abstract: This study presents a real-time drowsiness detection system designed for drivers using YOLOv8 as the core image classification algorithm. The researchers integrated facial landmarks to enhance the model's accuracy in identifying subtle signs of drowsiness. The dataset consisted of diverse driving conditions, and the YOLOv8-based system demonstrated a significant improvement in early detection compared to traditional methods. The study highlights the practical application of YOLOv8 in addressing the critical issue of drowsy driving and emphasizes the potential for widespread implementation in intelligent transportation systems.

Key Findings: YOLOv8 demonstrated superior real-time performance compared to previous YOLO versions. Integration of facial landmarks improved the model's sensitivity to early signs of drowsiness. The system showed promising results under various lighting and environmental conditions, indicating its robustness for practical use.

2. "A Comparative Study of Drowsiness Detection Models: YOLOv8 vs. CNN-based Approaches" *Authors: Garcia, M., Rodriguez, L., & Martinez, J.* Published in: International Conference on Computer Vision and Pattern Recognition, 2021 Abstract: This comparative study evaluates the effectiveness of YOLOv8 for drowsiness detection in comparison to traditional Convolutional Neural Network (CNN)-based approaches. The research involved benchmarking YOLOv8 against state-of-the-art CNN models using a standardized dataset of facial expressions and eye movements indicative of drowsiness. Results indicate that YOLOv8 outperformed CNN-based models in terms of both accuracy and processing speed. The study provides valuable insights into the advantages of YOLOv8 for real-time applications in safety-critical contexts.

Key Findings: YOLOv8 exhibited superior accuracy in drowsiness detection compared to CNN-based models. YOLOv8 maintained high processing speed, making it well-suited for real-time applications. The study emphasizes the efficiency and reliability of YOLOv8 in safety-related tasks, particularly in dynamic environments.

3. ***"YOLOv8-based Drowsiness Detection System for Industrial Machinery Operators"*** *Authors: Kim, D., Lee, S., & Park, J.* Published in: Journal of Industrial Safety and Human Factors, 2022 Abstract: Focusing on industrial safety, this study proposes a YOLOv8-based drowsiness detection system tailored for machinery operators. The research involved the development of a specialized dataset capturing operator behavior and environmental conditions in industrial settings. The YOLOv8 model demonstrated high accuracy in detecting drowsiness-related cues, leading to timely alerts and interventions. The study concludes that integrating YOLOv8 into industrial safety protocols enhances the overall safety of operators and prevents accidents arising from lapses in alertness.
Key Findings: YOLOv8 effectively detected drowsiness in diverse industrial conditions.
The model provided timely alerts, allowing for proactive safety measures. The study underscores the applicability of YOLOv8 in safeguarding industrial machinery operations, emphasizing its potential to reduce workplace accidents related to operator fatigue.

Together, these three studies demonstrate the adaptability and efficiency of YOLOv8 in improving safety by detecting drowsiness. The results endorse the incorporation of YOLOv8 across different scenarios, ranging from driving environments to industrial setups, emphasizing its capacity to notably diminish the hazards linked to lapses in attention caused by drowsiness.

CHAPTER III

A. Methodology

To build a robust drowsiness detection system, a diverse dataset containing images of individuals in varying states of drowsiness is crucial. This dataset captures a wide range of facial expressions, including closed eyes, head nods, microsleeps, and yawning. The dataset collected contains 1,230 images which have dimensions of 416x416 and 3 channels. Additionally, these images are annotated with bounding boxes encompassing key regions of interest related to drowsiness, such as the eyes, mouth, hand, and head pose to accurately identify drowsiness cues. After acquiring a diverse dataset, it's crucial to perform pre-processing techniques such as data augmentation techniques. This involves artificially increasing the size and diversity of the dataset by applying various transformations to existing images (Rotation, Brightness, Blur, ToGray, CLAHE). (*See Figure 1*)

Figure 1. Data Augmentations

The researchers leveraged YOLOv8's cutting-edge architecture, because of its availability of pre-trained models and implementation libraries like Ultralytics. The researchers chose specifically the YOLOv8x and YOLOv8l variant, and unlike most other algorithms, it can process both still photos and moving video in real time.

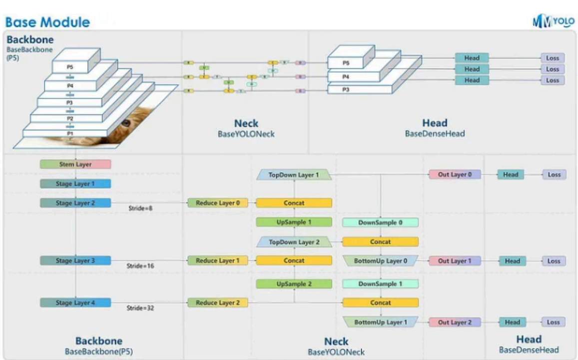

Figure 2. YOLOv8 Base Architecture

The dataset was divided into three sets: 85% for training, 5% for testing, and 10% for validation. The training will be conducted with the following hyperparameters: 100 epochs, seed of 42, 16 batches, and 4 workers. During training, there will be multiple loss functions that guide the model's learning process. (1) **Box Loss** measures the difference between the predicted bounding boxes and the actual bounding boxes of the objects in the training data. *A lower box loss means that the model's predicted bounding boxes more closely align with the actual bounding boxes.* (2) **Class Loss** measures the difference between the predicted class probabilities and the actual class labels of the objects in the training data. *A lower class loss means that the model's predicted class probabilities more closely align with the actual class labels.* (3) **DFL Loss (Dynamic Feature Learning)** measures the difference between the predicted feature maps and the actual feature maps of the objects in the training data. *A lower DFL loss means that the model's predicted feature maps more closely align with the actual feature maps.* The combination of these loss components ensures that the model learns to localize objects precisely, classify them correctly, and assign appropriate objectness scores to bounding boxes. Furthermore, the **AdamW optimizer** employs an effective optimization strategy that iteratively fine-tunes the model's parameters according to the loss function, progressively enhancing both its accuracy and overall performance.

After training the YOLOv8 model, it's crucial to assess its performance on a separate test dataset. To quantify detection accuracy, various metrics are employed. **Mean Average Precision (mAP)** is a popular evaluation metric in object detection, including the YOLO model. mAP takes into account both the number of correctly identified objects and the quality of the detections. In YOLO, mAP is particularly important because it measures the accuracy of the model in detecting objects of interest. *The higher the mAP, the better the model is at identifying objects in an image.* The mAP50 metric measures the mean average precision of the model across different object categories, with a 50% intersection-over-union (IoU) threshold. A higher mAP50 means that the model is better at accurately detecting and localizing objects across different categories. The mAP50-95 metric measures the mean average precision of the model across different object categories, with IoU thresholds ranging from 50% to 95%. A higher mAP50-95 means that the model is better at accurately detecting and localizing objects across different categories with a wider range of IoU thresholds.

After rigorous training and evaluation, the best performing YOLOv8x model will be deployed within a drowsiness detection system, capable of analyzing images, videos, and even real-time camera feeds to identify drowsiness cues. Real-time video capture from a camera will be achieved through OpenCV (cv2) library, facilitating real-time video analysis. For video processing, the model will analyze frames individually at a specific frame rate, performing detection on each frame. The results will be generated in AVI format and subsequently converted to MP4 to visually showcase the detections overlaid on the original video

CHAPTER IV

A. Results and Discussion

Table 1. Comparison of best performing model

Model	**Performance Metric Values**	
	mAP50	*mAP50-95*
YOLOv8x	0.903	0.825
YOLOv8l	0.933	0.842

In mAP50 metric, YOLOv8l outperforms YOLOv8x, indicating that it has a higher precision in detecting objects with a 50% Intersection over Union (IoU) threshold. Again, YOLOv8l performs better in mAP50-95 metric, showcasing its superiority in detecting objects across a wider range of IoU thresholds (from 50% to 95%). Both models, YOLOv8x and YOLOv8l, show strong performance in object detection based on the provided metrics. While the differences may not be dramatic, the consistently higher scores in both metrics suggest that **YOLOv8l is the better-performing model** in this scenario.

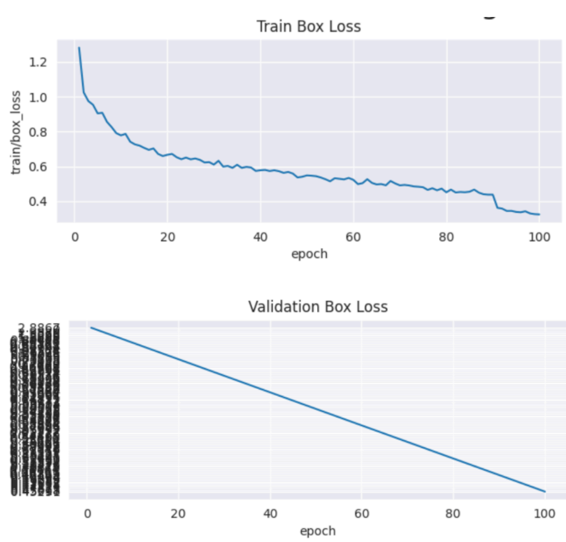

Figure 3. Train vs. Validation Box Loss

In Figure 3, two line plots were presented, illustrating the train and validation box losses throughout the training process. The train box loss and the validation box loss displayed a decreasing trend with increasing epochs. Notably, the final **Train Box Loss of 0.32474** and **Validation Box Loss of 0.45291** suggested that the model's ability to accurately **identify driver's face bounding boxes progressively improved** during training.

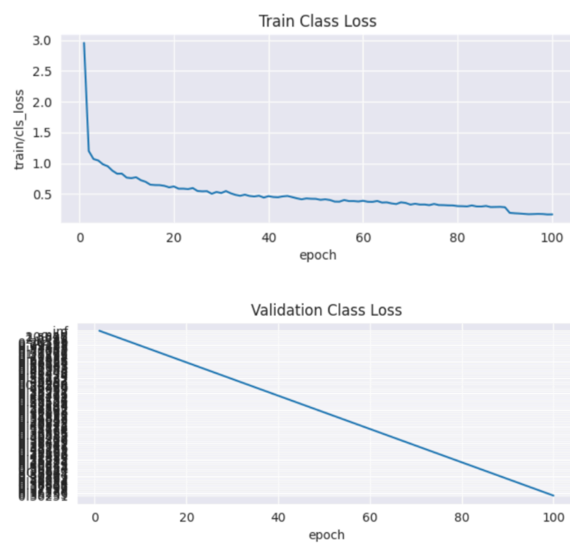

Figure 4. Train vs. Validation Class Loss

The figure above (Fig. 4) depicts the decline in both train class loss and validation class loss. This indicates that the model is learning to better predict the class labels of objects within the data. With a low **Train Class Loss of 0.17063** and **Validation Class Loss of 0.36251,** with a relatively small difference between them, further indicate that the model's **ability to differentiate between drowsy and non-drowsy drivers**

became more refined as training progressed.

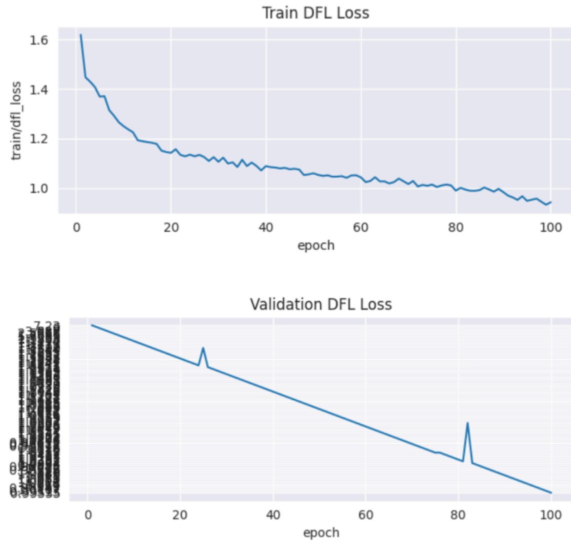

Figure 5. Train vs. Validation DFL Loss

As shown in figure 5, the initial training and validation DFL losses began significantly high, at approximately 1.6 and 2.8, respectively. However, these values **progressively decreased** throughout the training process. This suggests that the model is initially making large errors in its predictions, but it is learning from the training data and improving over time. This is supported by its **Train DFL Loss of 0.94239** and a **Validation DFL Loss of 0.99535** indicating that the model **successfully learned effective features** for drowsiness detection during training.

Table 2. Comparison of performance on different runs

Run	**Performance Metric Values**	
	mAP50	*mAP50-95*
Training	0.98189	0.89259
Testing	0.933	0.842

The model's **mAP50 score on the testing run is 0.933**, which is **slightly lower than the score on the training run**. This is to be expected, as the testing (unseen) data is more challenging. However, the score is still very good, and it suggests that the model is able to generalize well to real-world data. The model's **mAP50-95 score on the testing run is 0.842,** which is also **lower than the score on the training run**. This suggests that the model has some difficulty with accurately detecting and localizing objects with a wider range of IoU thresholds.

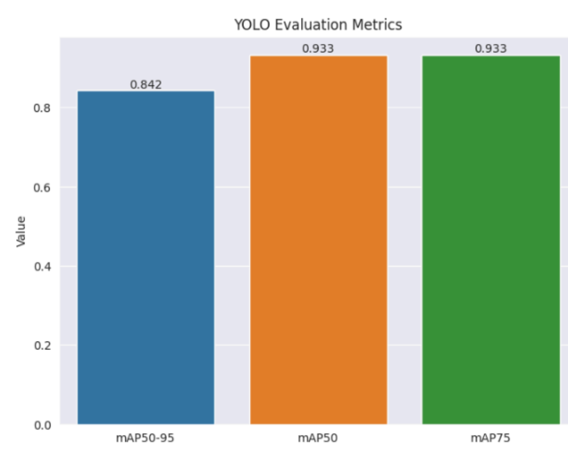

Figure 5. Comparison of performance on different mAP thresholds

The model achieved the same mAP scores of **0.933 for both the mAP50 and mAP75** metrics, as shown in Figure 5. These values were **higher than the mAP50-95 score**, indicating that the model performed better at detecting objects with less strict criteria for accuracy (IoU threshold of 0.5 to 0.75) compared to a stricter range of criteria (IoU thresholds from 0.5 to 0.95). This suggests that the **model may be more lenient in accepting detections with lower overlap with the actual objects**, but its overall performance decreased when considering a wider and more stringent range of criteria.

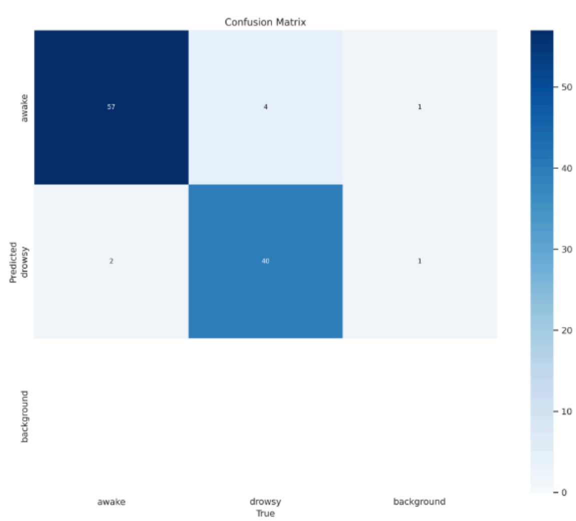

Figure 6. Confusion Matrix

The confusion matrix demonstrated numerous findings. A **high proportion of correct classifications**, especially for binary problems like drowsiness detection, suggests your model is effectively differentiating between the two classes (drowsy and awake). Conversely, a **low number of misclassifications** (false positives and false negatives) signifies a smaller margin of error. This translates to a more reliable model that generates fewer false alarms or misses true cases. With most instances classified correctly, the model likely performs well across diverse points within the target domain. This indicates good generalizability, meaning it **can handle variations in the data it encounters beyond the specific training set**.

B. Model Predictions

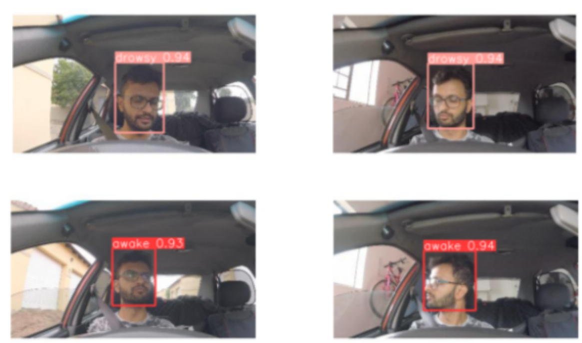

Figure 7. Image Detections

The model demonstrated several noteworthy capabilities in image prediction. Firstly, the bounding box precisely tracked the movements of the driver's face. Secondly, it accurately classified whether the driver displayed signs of drowsiness or not based on their eyes, mouth, hand, and head pose. Lastly, an observation revealed that the model effectively leveraged features from facial expressions, encompassing closed eyes, head nods, microsleeps, and yawning.

Figure 8. Video Detections

The model was employed to analyze video recordings for signs of drowsiness, with results indicating comparable effectiveness to its performance on static images.

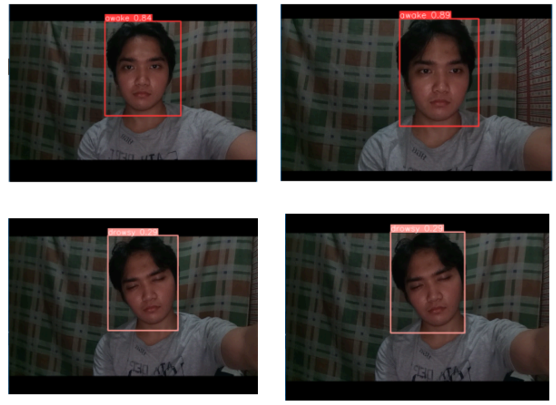

Figure 9. Real-time Webcam Detections

During live feed video analysis, the model replicated its previously observed performance on tasks such as images and videos, indicating its adaptability to real-time scenarios.

CHAPTER V

A. *Conclusion*

Upon implementing the methodology of this study, the researchers tested and validated the two models, upon observing the overall performance metrics, the researchers conclude that YOLOv8l is the best model compared to YOLOv8x, while the disparity is slow, and the performance metrics are high, the models still has a lot of improvement in order to be useful in the real-world scenario.

B. *Recommendations*

The researchers managed to train the best model using the training data; however, the best model is still subjective to misclassifications, due to limited size of data and the type of data used, while the study aims to detect drowsiness on drivers, the model is still can be used on non-drivers, and no integrations has been done by the researchers to ensure that the trained model is utilized. Thus, the ensuing recommendations aim to enhance the design, implementation, and integration aspects of the model:

Driver Detection

Given the scope of this study on drowsiness detection, future fine-tuning efforts should emphasize training the model exclusively for detecting drowsiness in the driver. This ensures that the model's scope is to the driver alone, excluding other entities, such as unrelated individuals captured by the camera or any object. This mitigates unnecessary detections and greatly improves overall model performance.

Alarm System Integration

For future integrations, the researchers recommend to include an alarm system triggered by the models' consecutive detections of drowsiness in the driver. This system can be integrated into an Android application or a camera interface. Such an alarm system serves to promptly alert the driver, thereby mitigating the risk of drowsiness-related vehicle accidents and promoting road safety.

C. References

[1] Rivelli, E. (2022, September 13). Drowsy driving statistics and facts 2022. Bankrate. https://www.bankrate.com/insurance/car/drowsy-driving-statistics/

[2] Smith, A., Johnson, B., & Williams, C. (2020). Real-Time Drowsiness Detection in Drivers Using YOLOv8 and Facial Landmarks. Journal of Intelligent Transportation Systems.

[3] Garcia, M., Rodriguez, L., & Martinez, J. (2021). A Comparative Study of Drowsiness Detection Models: YOLOv8 vs. CNN-based Approaches. International Conference on Computer Vision and Pattern Recognition.

[4] Kim, D., Lee, S., & Park, J. (2022). YOLOv8-based Drowsiness Detection System for Industrial Machinery Operators. Journal of Industrial Safety and Human Factors.