Malaysian Journal of Computing, 8 (1): 1375-1387, 2023

eISSN: 2600-8238

DRIVER DROWSINESS DETECTION SYSTEM THROUGH FACIAL EXPRESSION USING CONVOLUTIONAL NEURAL NETWORKS (CNN)

Nipa Das Gupta[1*], Rajesvary Rajoo[2], and Patricia Jayshree Jacob[3]

[1,2] School of Computing, [3] School of Applied Sciences, Nilai University, Negeri Sembilan, Malaysia*

[1*] nipantika.nipa@gmail.com, [2]rajes_e@gmail.com, [3]trishashree77@gmail.com

ABSTRACT

Driver drowsiness or fatigue is a significant factor that causes road accidents each year and considerably affects road safety. According to the World Health Organization (WHO), drowsy driving may contribute to approximately 6% of fatal and severe road accidents. To overcome this problem, we present a state-of-the-art, real-time drowsiness detection system, which exploits innovative deep-learning techniques to evaluate facial expressions. Our system analyzes not just the driver's eyes, mouth, and head rotation pose with front angles but also left and right yaw angles up to 90° to ensure the driver's safety. We gathered a dataset from public stock image websites, and manual image captures to develop the system. After processing the dataset, we extracted a wide range of features, which we fed into a deep convolutional neural network (CNN) algorithm. Specifically, we employed three different CNN algorithms which are EfficientDet D0, SSD MobileNet V2, and SSD ResNet50 V1, to classify the driver's drowsiness status using the facial key attributes in real time. Our results show that the SSD ResNet50 V1 model exhibited the highest accuracy and consistency in detecting driver drowsiness, underscoring the potential of our innovative system in promoting road safety. Our future work will focus on fine-tuning the approach to enhance its accuracy and performance.

Keywords: *Convolutional Neural Network (CNN), Deep Learning (DL), Driver Drowsiness, Facial Expression, Fatigue.*

Received for review: 21-12-2022; Accepted: 28-03-2023; Published: 10-4-2023
DOI: 10.24191/mjoc.v8i1.20286

1. Introduction

Drowsy driving is a critical contributor to the alarming rise of road accidents and requires immediate attention. A World Health Organization (WHO) survey conducted in 2021 shows that over 1.3 million people die in road accidents yearly. Reports on road accidents connect "drift-out-of-lane" crashes to drowsiness. The US National Sleep Foundation (2009) published that around 54% of drivers have driven vehicles while feeling drowsy, whereas 28% fall asleep at the wheel (Can, 2010). Various drowsiness detection systems have been proposed and developed to alert drivers when tired.

Traditional methods of detecting drowsiness, such as monitoring eye closure (Kumar *et al.*, 2021; Vesselenyi *et al.*, 2017), eye blinking (Maior *et al.*, 2018; Vijayan & Sherly, 2019), yawning (Jie *et al.*, 2018; Vijayan & Sherly, 2019), head movements (Yang *et al.*, 2020; Vijayan & Sherly, 2019), facial expression (Joshi *et al.*, 2020), and facial landmark detection (Mounika *et al.*, 2021) have limitations with high rates of false alarms, low precision, and complicated designs. This study presents a unique approach to detecting drowsy drivers by analyzing awake or sleepy facial expression status and using convolutional neural networks (CNN or ConvNets). The proposed system capitalizes on the recent advancements in deep learning and computer vision to detect tired drivers in real-time.

This research evaluates three different CNN algorithms, including EfficientDet D0, SSD MobileNet V2, and SSD ResNet50 V1, to determine driver fatigue through facial expression analysis. The performance of the proposed system is tested on a publicly available and manually captured image dataset and compared with the current state-of-the-art methods. The results show that the proposed system surpasses existing accuracy methods.

This study begins by discussing the existing driver drowsiness detection systems. It describes the proposed approach and the datasets used for evaluation. It then describes the implementation details and results of the experiments. Subsequently, it concludes the research and outlines future work.

In summary, this research paper presents a novel and effective solution for detecting drowsy drivers through facial expression analysis using CNN. Hence, this study aims to offer a system to detect drowsy driver facial expressions by focusing on the drivers' eyes, mouth contour area, and head movements with real-time recall and precision. It embarks on the following objectives:

- To propose an object detection-based driver drowsiness facial expression detection system using CNN algorithms.
- To enhance the accuracy of facial expression object detection by analyzing the driver's eyes, mouth, and head rotation pose with front angles or left and right yaw angles up to 90° simultaneously.

2. Related Works

Deep learning is based on the idea that a network should recognize patterns in a given data set. As such, it is becoming increasingly popular in many fields because of its ability to process large amounts of data quickly and accurately (Fan *et al.*, 2020). Deep learning's strength lies in its ability to decipher hierarchical features from raw data without human help. Deep learning (DL) is the basis for many artificial intelligence (AI) applications and services, such as automatic speech recognition (Deng & Yu, 2014; Han *et al.*, 2023), credit card fraud detection (Longe *et al.*, 2010; Sarmadi *et al.*, 2022), credit risk assessment (Halim & Shuhidan, 2022), OCR and spelling correction (Leblond *et al.*, 2018), traffic signs detection (Cireşan *et al.*, 2012), as well as drowsiness detection (Gulhane & Mohod, 2014; Joshi *et al.*, 2020), showcasing its incredible versatility.

Joshi *et al.* (2020) developed a drowsiness detection system to predict fatigue status by using a set of pose, expression, and emotion-based representations of the input video on four types of classes (alert, slightly drowsy, moderately drowsy, and extremely drowsy). They found that the 2D CNN model is the best technique to predict fatigue, with a 78% macro-averaged AUC-ROC, followed by LSTM (77%), 2D CNN and SMOTE (75%), 1D CNN (75%), and the random forest (RF) model (72%). On the other hand, Vijayan and Sherly (2019) compared the accuracy of three CNN techniques, namely ResNet50, VGG16, InceptionV3, and feature fused architecture (FFA). They found that, for the facial movements, InceptionV3 outperformed the other techniques with 78.57% accuracy, followed by ResNet50 (76.19%), FFA (75.71%), and VGG16 (71.42%).

However, we have found no study to predict drowsiness using the eyes, mouth, and head rotation pose with front angles or left and right yaw angles up to 90° simultaneously of the sample images. In deep learning, SSD MobileNet V2 and EfficientDet D0 are the most popular CNN models and share common roots in image classification and object detection (Yu *et al.*, 2019; Zocco *et al.*, 2022). We also include the SSD ResNet50 V1 model in the study to observe the performance of the supervised algorithm in deep learning, as done by other studies (Garcia-Venegas *et al.*, 2021; Yu *et al.*, 2019). So, this study looks at how well the EfficientDet D0, SSD MobileNet V2, and SSD ResNet50 V1 models can predict fatigue using the facial expressions with front angles or left and right yaw angles up to 90° in the sample images from the dataset gathered from Kaggle and taken manually.

2.1 Dataset

The dataset from Kaggle, called Drowsiness Prediction Dataset, comprised 9,120 photos of drivers. It included images of both sleepy and alert drivers. Since the interest of the study was to accomplish the experimental system objectives, we collected 24,000 images using a Rapoo C280 2K web camera with OpenCV in Python. We then manually labeled each image as "awake" and "drowsy" by looking at the face of the driver using the LabelImg tool. A total of 33,120 annotated images were stored in the training (29,440 images) and testing (3,680 photos) folders of the proposed system, with a ratio of 9:1. We then used this dataset for training and testing the deep learning model.

2.2 Deep Learning Technology

The CNN have become the cornerstone of deep learning techniques for computer vision tasks, such as image classification and object detection. CNNs have a hierarchical structure of layers that process input image data to produce meaningful information about objects and features in the image. In this study, we chose three popular CNN object detection models from the Tensorflow detection model zoo based on their inference speed and accuracy, represented by their mean average precisions (mAP) on the COCO 2017 dataset (Yu *et al.*, 2020). The models were the EfficientDet D0, SSD MobileNet V2, and SSD ResNet50 V1. We compared how well these models worked to find the best way to figure out if a driver is tired.

2.2.1 SSD MobileNet V2

SSD MobileNet V2 is a single-shot object detection model that combines the MobileNet V2 architecture with the Single Shot MultiBox Detector (SSD) architecture, as represented in Figure 1.

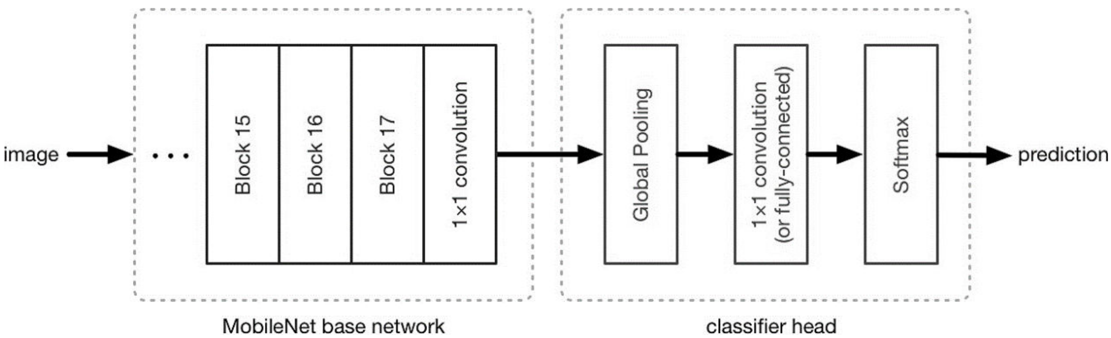

Figure 1. SSD MobileNet V2 (Hollemans, 2018).

MobileNet V2 is a lightweight CNN architecture optimized for speed and efficiency and designed to work well on mobile and edge devices with limited computational resources (Sandler *et al.*, 2019). The SSD architecture is a fast object detection framework that uses a single feed-forward network to predict object classes and bounding boxes (Liu *et al.*, 2016).

The combination of MobileNet V2 and the SSD architecture results in an efficient model capable of running on mobile devices with limited computational resources while still providing state-of-the-art accuracy. The model uses MobileNet V2's lightweight architecture to extract features, while the SSD layer precisely provides the accuracy needed to detect objects in an image (Sandler *et al.*, 2019).

2.2.2 SSD ResNet50 V1

The SSD ResNet50 V1, a sophisticated object detection model, synergizes the ResNet50 architecture with the SSD Feature Pyramid Network (FPN) framework (Lu *et al.*, 2019), as shown in Figure 2. ResNet50, a deep residual network honed on the ImageNet corpus, boasts exceptional accuracy and computational efficiency in image classification tasks (He *et al.*, 2016).

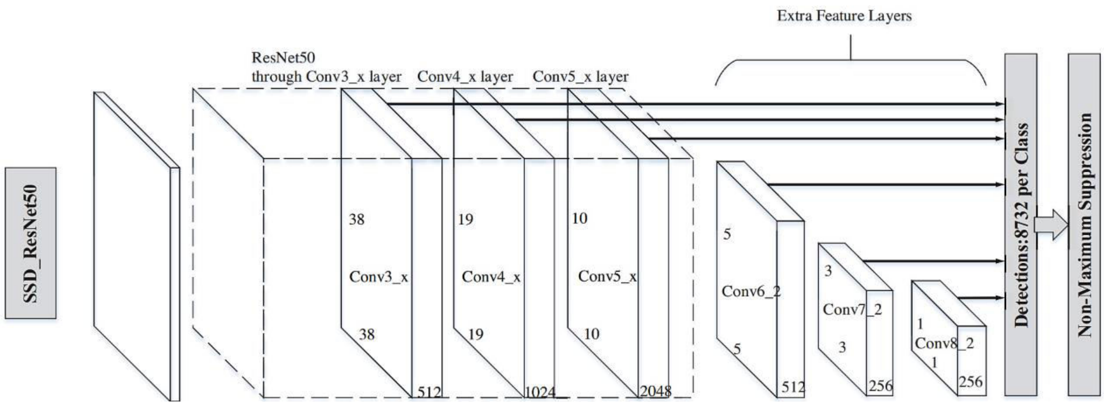

Figure 2. SSD ResNet50 V1 (Hu & Huang, 2020).

By combining the ability of ResNet50 to efficiently recognize objects with the SSD FPN's capacity to detect objects at various scales, SSD ResNet50 V1 can achieve high accuracy in object detection. FPN further augments ResNet50, allowing it to detect and accurately localize objects of varying sizes within a single image. The model's object localization lets it put bounding boxes around small objects at each place where they are found (Fathabadi *et al.*, 2022). This model is beneficial for applications such as fatigue detection, where the ability to detect and localize small objects accurately is crucial for driving safely.

2.2.3 EfficientDet D0

The EfficientDet D0 is a well-established one-stage detector paradigm, as highlighted in several seminal works (Liu *et al.*, 2016; Lin *et al.*, 2017). The models combine EfficientDets as the backbone network and BiFPN as the feature network, as shown in Figure 3. The BiFPN takes the level 3–7 features, represented by P3, P4, P5, P6, and P7, from the backbone network and executes a series of top-down and bottom-up bidirectional fusions, repeating the process multiple times.

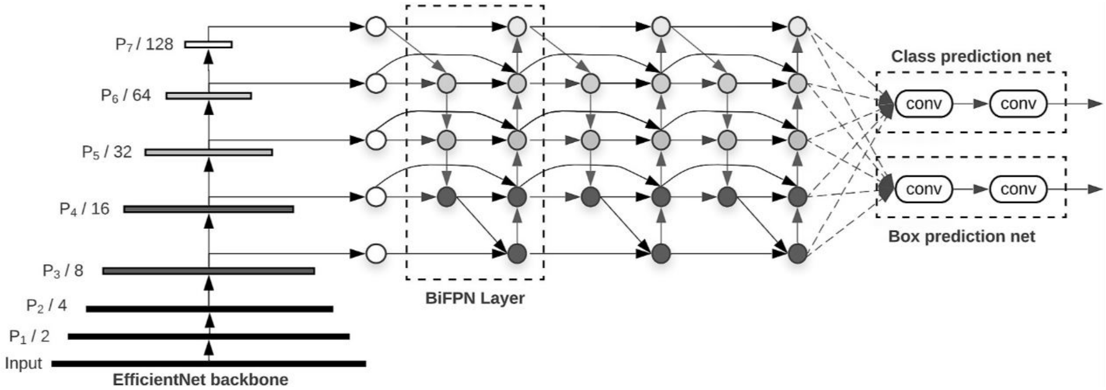

Figure 3. EfficientDet D0 (Tan *et al.*, 2020).

These multiple fusions help the model learn a feature representation suited for object detection. The result of these fusions is fed into a class and box network tasked with making predictions for object classes and bounding boxes (Tan *et al.*, 2020). The combination of EfficientDet and BiFPN provides superior accuracy compared to other state-of-the-art models.

2.3 Evaluation

Every time the training creates a new checkpoint, the evaluation tool will use the video frames in a given directory to make predictions about the model. The evaluation is done both during and after training. The predictions made by the evaluation tool during training can adjust the model and ensure that it is learning effectively.

2.3.1 Evaluation Metric

In this study, we used various metrics to analyze the performance of these three classification and object detection algorithms from the Tensorflow detection model zoo. These metrics depend on the confusion matrix (Reddi & Eswar, 2021). The confusion matrix shows the number of correct and incorrect predictions made by the system models by comparing the actual and predicted outcomes, either drowsy or awake, from the drivers' facial expressions. A confusion matrix is a helpful tool for diagnosing the performance of the system models.

Table 1. The confusion matrix.

	Awake	**Drowsy**
Awake	TP	FP
Drowsy	FN	TN

As represented in Table 1, it comprises four parameters: true positives (TP), false positives (FP), true negatives (TN), and false negatives (FN). These parameters are used to calculate our work's most critical performance metrics: precision, recall, F1-score, and mean average precision (mAP) (Al-Azzoa *et al.*, 2018). Precision (1) shows how good the predictions are, and recall (2) shows the percentages of objects detected, including the classifier.

$$Precision = \frac{TP}{TP+FP} \tag{1}$$

$$Recall = \frac{TP}{TP+FN} \tag{2}$$

2.3.2 Intersection over Union

The intersection over union (IoU) metric is a measure of accuracy used to evaluate the performance of object detectors. It assesses the precision of the system's prediction of the object's location within an image, as shown in Figure 4. The analysis is achieved by drawing a bounding box around the object and measuring the IoU threshold. The TensorFlow Object Detection API employs the "PASCAL VOC 2017 metrics" and considers a location prediction true positive (TP) if the IoU is above 50%. For example, if an object is correctly located in an image and the IoU between the predicted bounding box and the actual box is 0.6, it will be labeled as a TP.

Only one bounding box is assigned to each object. A TP implies that the model correctly identified the positive class's presence. Conversely, true negatives (TN) demonstrate the model's accuracy in recognizing the absence of the positive class. False positives (FP) show an erroneous prediction of the positive class, and false negatives (FN) represent an incorrect rejection of the positive class (Taqi *et al.*, 2019). To understand a model's performance, one must analyze the true positives (TPs), true negatives (TNs), false positives (FPs), and false negatives (FNs).

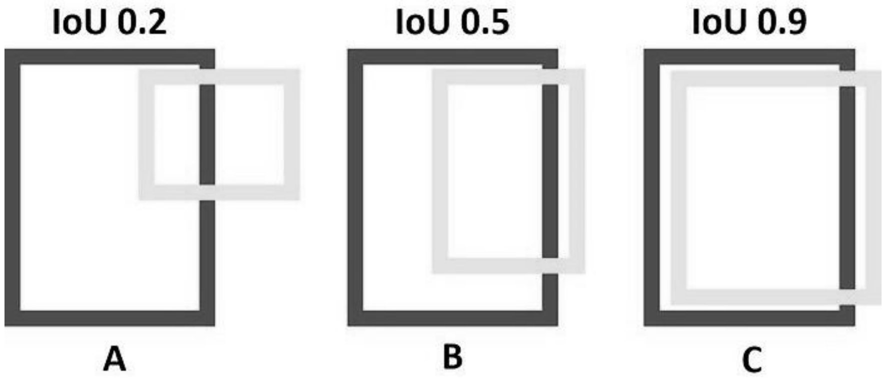

Figure 4. An instance of calculating IoU for different bounding boxes (Gad, 2021).

2.3.3 Mean Average Precision (mAP)

The mAP results in detecting bounding boxes to evaluate the network's ability to identify objects of interest while disregarding irrelevant information. A high mAP score signifies better network performance. The mAP, derived from the precision-recall curve, provides a single-value representation of the detector's characteristics. It is an important performance metric used to measure the accuracy of object detectors.

The Average Precision (AP) calculation estimates the recall and precision rate and then averages over multiple thresholds. In PASCAL metrics, these thresholds range from 0 to 1. It is calculated by calculating the area under the precision-recall curve generated by comparing the number of true positives, false positives, and false negatives (Al-Azzoa *et al.*, 2018).

2.3.4 Average Recall

Like AP, the Average Recall (AR) calculation estimates the recall rate of the model and then the average overlap of multiple thresholds. In PASCAL metrics, these thresholds range from 0.5 to 1. A higher AR score means the model can successfully identify more objects from a given image (Hosang *et al.*, 2016).

3 Implementation

The implementation of the proposed system embodies a sophisticated approach to object detection through innovative technology. Drawing upon the powerful capabilities of Python 3 and leveraging the TensorFlow Object Detection API, this implementation trains CNN with an unwavering focus on accuracy and efficiency. OpenCV is used skillfully to preprocess images, making it possible to add visual data to the training process with no problems.

The training is based on a robust and varied dataset, allowing models to learn and improve. We chose three specific models for this implementation: EfficientDet D0, SSD MobileNet V2, and SSD ResNet50 V1. Each model was fine-tuned to work at its best through a carefully calculated total of 5000 (5K) training steps.

At the inception of training, the models were immediately tested, presenting valuable loss metrics, such as classification, localization, regularisation, and total loss matrices. These metrics revealed a great deal about the performance of the models. For example, they showed where the training steps needed to be tweaked or where the dataset needed to be relabeled.

The Tensorboard tool was employed to evaluate the models, providing a comprehensive analysis of the training and evaluation metrics. This data-driven approach to evaluation ensured that we thoroughly vetted the models, with any areas of weakness swiftly identified and addressed. For this implementation, the system requirements were high, with an Intel Core i7-11800H CPU running at 2.30 GHz, 16 GB of RAM, and an Nvidia GeForce RTX 3060 GPU at the top of the list. With this kind of hardware, the deep learning models could be trained with a high level of accuracy, and the model weights could be optimized for the most efficiency.

4 Results and Discussion

This study examined the details of loss metrics, learning rate, and evaluation metrics in EfficientDet D0, SSD MobileNet V2, and SSD ResNet50 V1 CNN classifiers.

The loss matrices, as illustrated in Figure 5 via Tensorboard, represent a crucial aspect of understanding the performance of the trained deep learning models. The loss metrics encompass multiple crucial elements, including classification loss, regularisation loss, localization loss, and the overall total loss. Looking at these loss metrics, one can determine if the model converges and can be used in other situations.

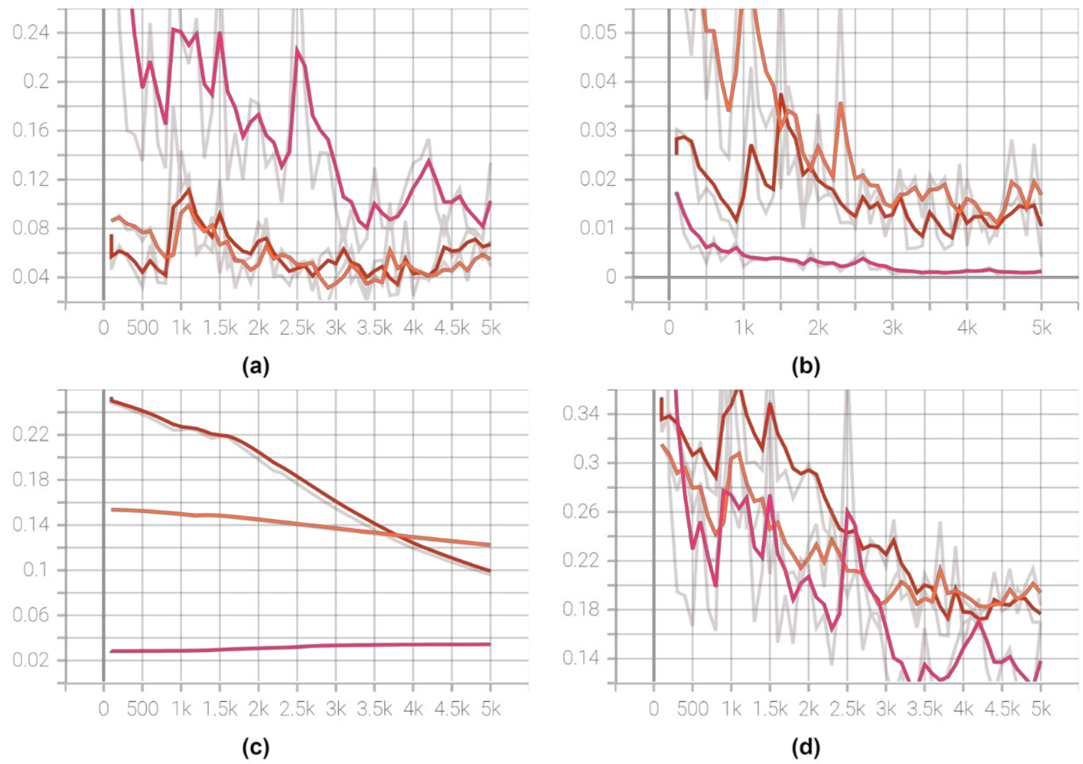

Figure 5. (a) Classification loss, (b) localization loss, (c) regularisation loss, and (d) total loss metrics. The loss incurred at distinct steps for each model.

Our examination revealed that all models incurred losses over 5K training steps. The classification loss of EfficientDet D0 was the lowest, at 0.1337, followed by SSD MobileNet V2 (0.04909) and SSD ResNet50 V1 (0.06986). EfficientDet D0 also had the lowest localization loss value, at 1.5159e-3, followed by SSD MobileNet V2 (0.01692) and SSD ResNet50 V1 (4.2058e-3).

Regarding regularisation loss, EfficientDet D0 had the lowest value, at 0.03431, followed by SSD MobileNet V2 (0.1214) and SSD ResNet50 V1 (0.09586). EfficientDet D0 also had the lowest total loss value, at 0.1696, followed by SSD MobileNet V2 (0.1835) and SSD ResNet50 V1 (0.1699). Thus, results show that the loss metrics decrease and become smoother for all drowsiness detection system models during training, leading to well-moderated metrics for the drowsiness detection system models. This result was further evident when comparing the accuracy scores of the drowsiness detection system models.

Table 2. The learning rate with the number of steps for each CNN model.

Model	**Learning Rate**	**Steps**
EfficientDet D0	0.0385	5K
SSD MobileNet V2	0.07878	5K
SSD ResNet50 V1	0.07999	5K

Table 2 displays the learning rate of the evaluated models, EfficientDet D0, SSD MobileNet V2, and SSD ResNet50 V1, evaluated over 5,000 training steps. EfficientDet D0 had the lowest learning rate, at 0.0385, suggesting the model had the most stable and efficient learning process. On the other hand, SSD MobileNet V2 and SSD ResNet50 V1 had higher learning rates of 0.07878 and 0.07999, respectively. Results meant these models needed more fine-tuning to work at their best. Despite the higher learning rates, SSD MobileNet V2 and SSD ResNet50 V1 provided superior performance over EfficientDet D0.

Table 3. The performance comparison of best techniques.

Measures	EfficientDet D0	SSD MobileNet V2	SSD ResNet50 V1
AP@ 0.5 IoU	1.000	1.000	1.000
AR@ 0.5:0.95 IoU	0.824	0.897	0.919
mAP@ 0.5 IoU	1.000	1.000	1.000

Table 3 presents the results of the three CNN models based on three performance metrics: AP, AR, and mAP, all calculated at an intersection over a union (IoU) threshold of 0.5. All models get a perfect score of 1.000 when the maximum AP value is 0.5 IoU, which means they are very good at detecting drowsiness.

The maximum AR values at 0.5:0.95 IoU displays the AR rate, with EfficientDet D0 having a recall rate of 0.824, SSD MobileNet V2 0.897, and SSD ResNet50 V1 0.919. The highest value, 0.919, shows that SSD ResNet50 V1 can find signs of drowsiness in images more accurately.

The maximum mAP values at 0.5 IoU depict the mAP across all classes, with all models achieving the maximum score of 1.000. This result indicates consistent accuracy in detecting drowsiness across all classes for all models.

In conclusion, the results show that all three CNN models do an excellent job of detecting objects. SSD ResNet50 V1 is the best model for this study because it is the most accurate and consistent. Nevertheless, SSD ResNet50 V1 could not find all the objects ideally and had some false positives, which suggests that more research is needed to make better models for detecting drowsiness.

5 Conclusion

This study aimed to introduce an innovative solution for detecting drowsy drivers via analyzing facial expressions using CNN. The study aimed to determine the superior accuracy of the proposed system compared to existing state-of-the-art methods. The system leverages the analysis of drivers' eyes, mouth contour area, and head movements to predict drowsiness accurately. A dataset of 33,120 annotated images was employed to train and evaluate three different CNN algorithms: EfficientDet D0, SSD MobileNet V2, and SSD ResNet50 V1. We implemented the proposed system using Python 3 and the TensorFlow Object Detection API, with the training and evaluation metrics analyzed through Tensorboard. The evaluation results showed that the SSD ResNet50 V1 algorithm performed well and exhibited the highest accuracy and consistency in predicting drowsiness. This research makes up a promising development in the ongoing efforts to mitigate road accidents caused by drowsy driving. Future work could involve integrating this system into vehicles and continuously improving the algorithms for enhanced performance.

Acknowledgment

The author wishes to extend heartfelt thanks to Nilai University for their invaluable contributions and guidance in facilitating the successful completion of this scholarly endeavor.

Funding

The author (s) received no specific funding for this work.

Author Contribution

Author1 prepared the literature review and research method, implemented the experimental system to conduct data analysis, and interpreted the results. Author2 supervised the writing of the article. Author3 proofread the paper.

Conflict of Interest

The authors have no conflicts of interest to declare.

References

Al-Azzoa, F., Mohammed, A., & Milanovab, M. (2018). Human Related-Health Actions Detection using Android Camera based on TensorFlow Object Detection API. International Journal of Advanced Computer Science and Applications, 9(10). https://doi.org/10.14569/ijacsa.2018.091002.

Can, D. B. G. E. S. (2010). Save your life this Memorial Day. National Sleep Foundation (NSF): Arlington, VA, USA.

Cireşan, D., Meier, U., Masci, J., & Schmidhuber, J. (2012). Multi-column deep neural network for traffic sign classification. Neural Networks, 32, 333–338. https://doi.org/10.1016/j.neunet.2012.02.023.

Deng, L. & Yu, D. (2014). Deep Learning: Methods and Applications. Foundations and Trends® in Signal Processing, 7(3-4), 197–387. https://doi.org/10.1561/2000000039

Fan, J., Ma, C., & Zhong, Y. (2020, April 19). A selective overview of deep learning. PubMed Central (PMC). Retrieved February 12, 2023, from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8300482/.

Fathabadi, F. R., Grantner, J. L., Abdel-Qader, I., & Shebrain, S. A. (2022). Box-Trainer Assessment System with Real-Time Multi-Class Detection and Tracking of Laparoscopic Instruments, using CNN. Acta Polytechnica Hungarica, 19(2), 7–27. https://doi.org/10.12700/aph.19.2.2022.2.1.

Gad, A. (2021). Evaluating Object Detection Models Using Mean Average Precision - KDnuggets. KDnuggets. Retrieved February 9, 2023, from https://www.kdnuggets.com/evaluating-object-detection-models-using-mean-average-precision.html.

Garcia-Venegas, M., Mercado-Ravell, D. A., & Carballo-Monsivais, C. A. (2021). On the safety of vulnerable road users by cyclist orientation detection using Deep Learning. Machine Vision and Applications, 32(5), 109. https://doi.org/10.1007/s00138-021-01231-4.

Gulhane, M., & Mohod, P. S. (2014). Intelligent fatigue detection and automatic vehicle control system. arXiv. https://doi.org/10.48550/arXiv.1407.2412.

Halim, Z., & Shuhidan, S. M. (2022). Towards development of robust machine learning model for Malaysian corporation: A systematic review of essential aspects for corporation credit risk assessment. Malaysian Journal of Computing (MJoC), 7(1), 1011-1026.

Han, M., Wang, Q., Zhang, T., Wang, Y., Zhang, D., & Xu, B. (2023). Complex dynamic neurons improved spiking transformer network for efficient automatic speech recognition. arXiv. https://doi.org/10.48550/arXiv.2302.01194.

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep Residual Learning for Image Recognition. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 770–778. https://doi.org/10.1109/cvpr.2016.90.

Hosang, J., Benenson, R., Dollar, P., & Schiele, B. (2016). What Makes for Effective Detection Proposals? IEEE Transactions on Pattern Analysis and Machine Intelligence, 38(4), 814–830. https://doi.org/10.1109/tpami.2015.2465908.

Hu, X., & Huang, B. (2020). Face Detection based on SSD and CamShift. 2020 IEEE 9th Joint International Information Technology and Artificial Intelligence Conference (ITAIC). https://doi.org/10.1109/itaic49862.2020.9339094.

Jie, Z., Mahmoud, M., Stafford-Fraser, Q., Robinson, P., Dias, E. and Skrypchuk, L. (2018) Analysis of Yawning Behaviour in Spontaneous Expressions of Drowsy Drivers. 13th IEEE International Conference on Automatic Face & Gesture Recognition (pp. 571-576). IEEE.

Joshi, A., Kyal, S., Banerjee, S., and Mishra, T. (2020). In-the-wild Drowsiness Detection from Facial Expressions. 2020 IEEE Intelligent Vehicles Symposium (IV), 2020 IEEE Conference on (pp. 207-212). IEEE.

Kumar, R., Rathore, H., Agrawal, P., & Gupta, P. (2021). Drowsiness Detection Using Viola-Jones Object Detection Algorithm for Real-Time Data. Advances in Intelligent Systems and Computing, 369–380. https://doi.org/10.1007/978-981-16-0171-2_35.

Leblond, R., Alayrac, J.-B., Osokin, A., & Lacoste-Julien, S. (2018). Searnn: Training runs with global-local losses. arXiv. https://doi.org/10.48550/arXiv.1706.04499.

Lin, T.-Y., Dollár, P., Girshick, R., He, K., Hariharan, B., & Belongie, S. (2017). Feature Pyramid Networks for Object Detection. ArXiv:1612.03144 [Cs]. https://arxiv.org/abs/1612.03144v2.

Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.-Y., & Berg, A. C. (2016). SSD: Single Shot MultiBox Detector. Computer Vision – ECCV 2016, 21–37. https://doi.org/10.1007/978-3-319-46448-0_2.

Longe, O. B., Mbarika, V., Kourouma, M., Wada, F., & Isabalija, R. (2010). Seeing beyond the surface, understanding and tracking fraudulent cyber activities. arXiv. https://doi.org/10.48550/arXiv.1001.1993.

Lu, X., Kang, X., Nishide, S., & Ren, F. (2019). Object detection based on SSD-ResNet. 2019 IEEE 6th International Conference on Cloud Computing and Intelligence Systems (CCIS). https://doi.org/10.1109/ccis48116.2019.9073753.

Maior, C.B., Moura, M.D., Santana, J.M., do, L.M., Nascimento, Macedo, J.B., Lins, I.D., & Droguett, E.L. (2018). Real-time SVM Classification for Drowsiness Detection Using Eye Aspect.

Mounika, D., Deepika, K. D., Varma, A.R.S.R., P., & Kesava, M. (2021). A driver drowsiness detection framework using deep learning heuristic. In Anil Neerukonda Institute of Technology and Sciences (UGC Autonomous). https://cse.anits.edu.in/projects/projects2021A5.pdf.

Reddi, S., & Eswar, G. V. (2021). Fake news in social media recognition using Modified Long Short-Term Memory network. Security in IoT Social Networks, 205–227. https://doi.org/10.1016/b978-0-12-821599-9.00009-1.

Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., & Chen, L.-C. (2018). MobileNetV2: Inverted Residuals and Linear Bottlenecks. ArXiv.org. https://arxiv.org/abs/1801.04381.

Sarmadi, A., Fu, H., Krishnamurthy, P., Garg, S., & Khorrami, F. (2022). Privacy-preserving collaborative learning through feature extraction. arXiv. https://doi.org/10.48550/arXiv.2212.06322.

Tan, M., Pang, R., & Le, Q. V. (2020). EfficientDet: Scalable and Efficient Object Detection. ArXiv:1911.09070 [Cs, Eess]. https://arxiv.org/abs/1911.09070v7.

Taqi, A. M., Al-Azzo, F., Awad, A. E., & Milanova, M. (2019). Skin Lesion Detection by Android Camera based on SSD- MobileNet and TensorFlow Object Detection API. Www.semanticscholar.org. https://www.semanticscholar.org/paper/Skin-Lesion-Detection-by-Android-Camera-based-on-Taqi-Al-Azzo/776c0b3620d795bec5e940bff2d846bcc0c8f815.

Vesselenyi, T., Moca, S., Rus, A., Mitran, T., & Tătaru, B. (2017). Driver drowsiness detection using ANN image processing. IOP Conference Series: Materials Science and Engineering, 252, 012097. https://doi.org/10.1088/1757-899x/252/1/012097.

Vijayan, V. & Sherly, E. (2019). Real-time detection system of driver drowsiness based on representation learning using deep neural networks. Journal of Intelligent & Fuzzy Systems, 36(3), 1977–1985.

Yang, C., Wang, X., & Mao, S. (2020). Unsupervised Drowsy Driving Detection With RFID. IEEE Transactions on Vehicular Technology, 69(8), 8151–8163. https://doi.org/10.1109/tvt.2020.2995835.

Yu, H., Chen, C., Du, X., Li, Y., Rashwan, A., Hou, L, Jin, P., Yang, F., Liu, F., Kim, J., and Li, J. (2020). TensorFlow Model Garden. Available at: https://github.com/tensorflow/models.

Yu, J., Seidel, R., & Hirtz, G. (2019). OmniPD: One-Step Person Detection in Top-View Omnidirectional Indoor Scenes. Current Directions in Biomedical Engineering, 5(1), 239–244. https://doi.org/10.1515/cdbme-2019-0061.

Zocco, F., Huang, Ching-I., Wang, H.-C., Khyam, M. O., & Van, M. (2022). Towards More Efficient EfficientDets and Low-Light Real-Time Marine Debris Detection. ArXiv:2203.07155 [Cs]. https://arxiv.org/abs/2203.07155.