CLASSIFICATION OF DRIVER DROWSINESS DETECTION USING A LIGHTWEIGHT NEURAL NETWORK MODEL

Habib Olaniyi Aliu & Joy Oluwabukola Olayiwola
Computer Engineering Department, Federal Polytechnic, Ilaro, Ogun State.
Olaniyi.aliu@federalpolyilaro.edu.ng; Oluwabukola.olayiwola@federalpolyilaro.edu.ng

Abstract

One of the main contributing factors to traffic accidents is driver drowsiness, which is mostly brought on by exhaustion. Even though drowsiness only lasts a short while, it has such a detrimental effect on everyone—not just the driver. In this study, a CNN lightweight model called MnasNet is employed to identify driver drowsiness. MnasNet has the capacity to be both quick and precise. The study used a unique dataset with 6,000 photos that was then divided into four categories: yawn, no yawn, eye closed, and eye closed, respectively. A ratio of 80:10:10 was used to split the data samples for each class into training, validation, and testing groups accordingly. The model's implementation accuracy was 83.16%.

Keywords: CNN, MnasNet, driver drowsiness, Machine Learning

Introduction

The word "drowsiness" refers to excessive daytime sleepiness or exhaustion. Even though drowsiness may only last a short while, negative effects could still occur. In most cases, fatigue is the main cause of drowsiness since it reduces consciousness and focus (Cavuoto and Megahed, 2016). Even though falling asleep while driving is harmful, being exhausted makes it tough to stay alert when driving. One in twenty drivers is predicted to have slept off behind the wheel (Mahajan et al., 2019). According to Ahmed et al. (2022), a traffic accident occurs when the driving system of a road vehicle is unable to fulfil one or more tasks that are necessary for the journey to be completed safely.

In the fourth quarter of 2022, 22,852 people were involved in road traffic collisions (RTC), of which 1,600 died and 10,232 were injured, according to the Federal Road Safety Corps (FRSC), which focuses mainly in Nigeria (FRSC, 2023a). The World Bank has ranked Nigeria as the 54th most accident-prone country in the world. The nation boasts one of the longest road networks in sub-Saharan Africa (SSA), stretching over 194,394 km in total (Yakubu et al., 2023). The stunning rise in road traffic accidents in emerging nations has made them a significant contributor to the world mortality rate (Adedotun et al., 2022).

The research on how to gauge a driver's level of sleepiness is divided into three categories: 1) Vehicle behaviour depends on mechanical and sensor input. 2) Sensor-based biological signals; 3) Image processing with computer vision that concentrates on changes in facial features, is one of the linked symptoms of sleepiness (Sunagawa et al., 2019). There are now two types of solutions available to deal with driver drowsiness: intrusive systems and nonintrusive systems. The most accurate system is intrusive because it monitors physiological signals, but drivers find it annoying because it requires wearing sensors (Zandi et al., 2019). The observation of eye closure in drowsiness detection is believed to be the first and most significant sign that one should look out for. In contrast, non-intrusive systems focus on facial features. feature adjusts under various face angles, expression and lighting (Lotfy and Saparon, 2020).

A key component of artificial intelligence, machine learning (ML) is the process of educating a computer system to generate precise predictions from input data. By providing the computer with specific data to assist in differentiating appropriately and obtaining outcomes at the same time, it enables the computer to handle important tasks (Von Rueden et al., 2021). In areas including fraud detection, character recognition, object detection, picture identification, and object segmentation, machine learning has been employed. The CNN architecture was chosen primarily because it can analyse massive amounts of data and generate extremely accurate predictions, making it ideal for image recognition and image classification as well as other computer vision applications (Ramprasath, Anand, and Hariharan, 2018).

There are various ways to categorise a drowsy driver. Cui et al. (2021) employed a convolutional neural network to predict driver fatigue across participants using just one channel of EEG data. It included 27 participants (aged 22 to

28) who provided 62 EEG data sets between 2005 and 2012. A virtual reality driving simulator offered a sustained-attention driving challenge. The drivers were instructed to keep the vehicle in the centre of the lane and adjust their manoeuvres accordingly. In order to detect intoxication while driving, a convolutional neural network (CNN) will be utilised to recognise shared EEG data among numerous individuals. Because the model structure incorporates a Global Average Pooling (GAP) layer, it is possible to identify which input signal regions contribute more to classification using the Class Activation Map (CAM) technique. The suggested model surpasses other cutting-edge deep learning techniques and traditional machine learning techniques in categorising 2-class cross-subject EEG data, having an overall accuracy of 73.22% on 11 participants, according to the results.

Four measures have been identified for the detection of drowsiness, according to a study by Albadawi et al. (2022). One of the measures is based on the vehicle, taking into account the angle of the steering wheel and any deviation from the highway lane; another is based on bio-signals like electrooculogram (EOG), electrocardiography (ECG), electroencephalography (EEG), etc. Although highly accurate, biosignal measurements are intrusive for drivers. The other metric is based on image analysis and focuses on the position of the driver's eyes, mouth, and head. It is popular since it is non-invasive and uncomfortable for the driver.

Furthermore, the majority of drowsiness signals are expressed in facial features, making it simpler to spot drowsy driving signs in a driver. The last approach combines the previous three strategies. Singh et al. (2023) developed a system to recognise drowsiness using PERCLOS, which calculates the percentage of time the eyelids are closed, along with the eye aspect ratio (EAR). Python was used to develop the system, which only addresses the face as a unique physical element. A digital camera (webcam) is placed directly in front of the driver's face to record the input video. After several frames, if no face is found, the system concludes that the drivers are nodding off. 68 facial markers and OpenCV are used to identify the face and eye. It is possible to determine if an eye is open or closed using the Euclidean eye aspect ratio. Dlib was used to do eye point extraction, which is necessary for EAR. Their system's accuracy for the presented method was 80%.

Ghourabi et al. (2020) developed a trustworthy approach for tiredness detection by concentrating on the driver's nodding, yawning, and blinking while evaluating photographs of the driver's face. The five processes that were considered when determining driver drowsiness were preprocessing, face detection, face combination, face extraction, and classification utilising the Multilayer Perceptron Classifier (MLP) and the K-Nearest Neighbours (K-NN). The benchmark NTHU-DDD video dataset was used to evaluate the proposed method. There are 22 themes covered in it, including drivers who are men or women. Each participant acts out one of five scenarios: "Sunglasses," "No glasses," "Glasses," "Night glasses, and "Night no glasses." Data can be divided into two categories: training data and evaluation data. The training data set has 18 participants. A total of 180 films involving 18 subjects were used during the training phase. The evaluation data includes 20 films on four different topics. The average scores for K-NN and MLP were 72.58% and 74.91%, respectively.

Methodology

The suggested approach makes use of a collection of yawns and human eyes as well as a CNN model called MnasNet for categorization. MnasNet was used for mobile neural architecture search, which specifically factors in model latency into its primary goal in order to construct a model that provides a good trade-off between accuracy and latency. An autonomous neural architecture search methodology was utilised to build mobile models with reinforcement learning (Li, 2017).

Data collection

The data used in the study was compiled from publicly accessible datasets, pre-processed, and set up according to a predefined architecture, enabling easy CNN training procedures. The dataset was gathered from two sources. Eye open and eye closed was gotten from (Nabil, 2018) while yawn and no yawn was gotten from (Vazquez, 2021). Then, the pre-trained CNN architecture was used to train the prepared datasets, and the results of the detection model's efficacy were evaluated. 3,000 eyes images and 3,000 yawn images taken from Kaggle make up the dataset.

Table 1: Driver Drowsiness Dataset Description

Dataset Type	No of images
Eye open	1500
Eye closed	1500
Yawn	1500
No yawn	1500

Data Preparation

There were 6,000 datasets altogether in the collection, which were split into four labels. The dataset's labels each contain 1500 data points. To prevent overfitting, the data samples for each label were divided into 80:10:10 training, validation, and testing groups, respectively. The dataset of driver drowsiness photos has a variety of sizes, so they must be resized to meet the specifications of the pre-trained CNN architectures. To prepare the datasets for input into the MnasNet model, they were scaled to 224x224x3 pixels. The input to the model, which consists of photos of the driver with their eyes open, closed, yawning, and not yawning, was shrunk to a size of 224x224x3 pixels.

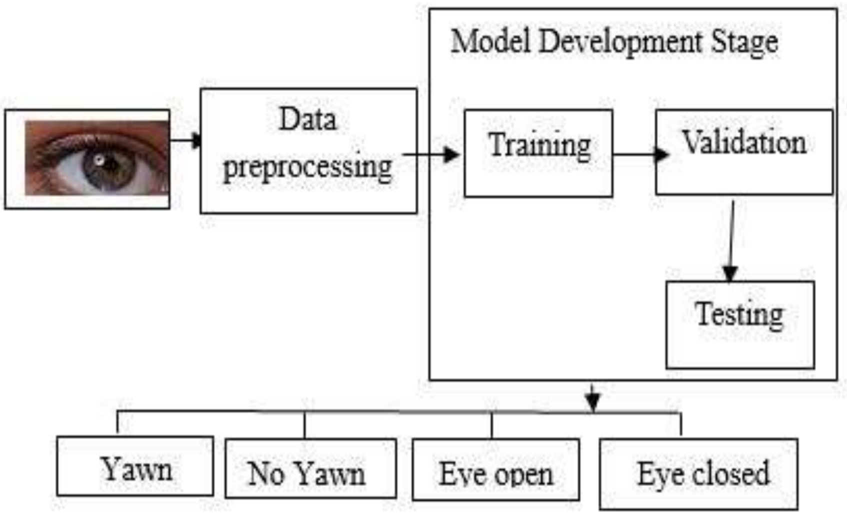

Fig. 2.1: Block diagram of driver drowsiness classification model

Implementation Details

A dense layer containing four (4) output layers, which is the desired output, and global average pooling were used to fine-tune the model. The global pooling layer is used to convert a two-dimensional feature vector into a one-dimensional feature vector. The next set of four layers, which are either completely interconnected or dense, follow the global pooling layer. The ReLU (rectified linear unit) activation function was employed to activate the completely linked layers, each of which comprised 128 nodes. The output layer, which comes after, classifies using feature maps it has learned from the models. These maps are used to classify the output layer. The output layer has a SoftMax activation algorithm.

The dataset was normalised using the normalise function from the OpenCV libraries, which is part of the Python pre-processing package. During the training phase, an optimizer, a scheduler for the learning rate, and a quantity of epochs were all defined. The Adam optimizer was used for its hyperparameters and tuning. The MnasNet model has an epoch of 10. The early termination callback function was employed since the learning rate scheduler was designed to cut the training rate by a factor of 0.01 if the validation loss was not reduced after 10 epochs. The Python model was developed using the TensorFlow library (Vasilec et al., 2019). The testing was performed using a 3.90 GHz Intel(R) Core(TM) i3-7100 processor and 4GB of 2400 MHz RAM with Python Juypter Notebook software.

These four criteria will be used to assess the models. They fit the criteria listed in Equations 2.1 to 2.4:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \qquad (2.1)$$

The ratio of true positive and true negative measurements to the total measurements is used to calculate accuracy.

$$\mathrm{Pr}ecision = \frac{TP}{TP + FP} \qquad (2.2)$$

$$\mathrm{Re}\,call = \frac{TP}{TP + FP} \qquad (2.3)$$

The percentage of all correctly mentioned categories is known as recall. This assesses the potency of the particular classification scheme in use.

$$F1 - Score = \frac{2 * precision * \mathrm{Re}\,call}{\mathrm{Pr}ecision + \mathrm{Re}\,call} \qquad (2.4)$$

The weighted average of precision and recall equal to twice the sum of precision and recall is used to get the F1 score.

And, respectively, "true positive," "false positive," "true negative," and "false negative" are designated by the acronyms TP, FP, TN, and FN.

Results

The datasets were correctly classified using the MnasNet neural network model, yielding the results shown below.

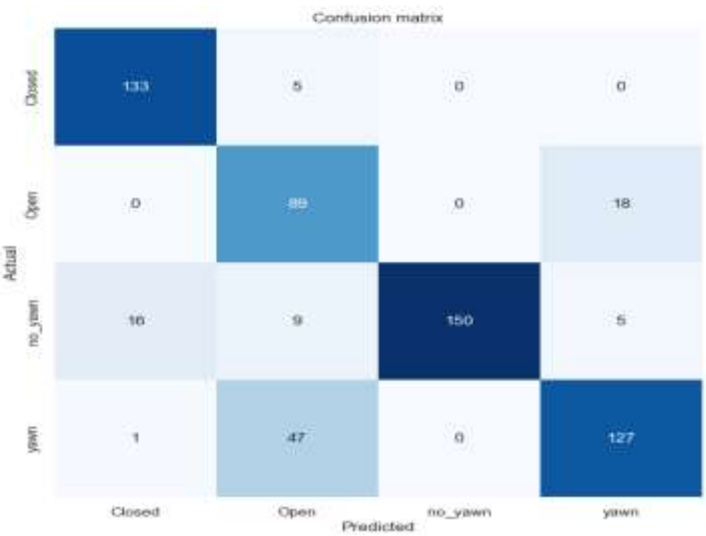

Fig. 2: Confusion matrix for MnasNet model

Table 2: MnasNet loss and accuracy function for classification of drowsiness

Epoch	**Loss**	**accuracy**	**val loss**	**val accuracy**
1	1.7745	0.5210	1.0997	0.6433
2	0.7343	0.7588	0.8911	0.7267
3	0.6079	0.8021	0.8326	0.7350
4	0.5562	0.8117	0.6662	0.8167
5	0.5034	0.8335	0.6897	0.7933
6	0.4729	0.8454	0.6789	0.8333
7	0.4590	0.8481	0.5802	0.8300
8	0.4459	0.8487	0.5870	0.8217
9	0.4319	0.8569	0.5293	0.8383
10	0.4058	0.8621	0.5503	0.8233

Table 3: Result of training with MnasNet

Dataset	Precision	recall	f1-score	support
Closed	0.89	0.96	0.92	138
Open	0.59	0.83	0.69	107
No yawn	1.0	0.83	0.91	180
Yawn	0.85	0.73	0.78	175

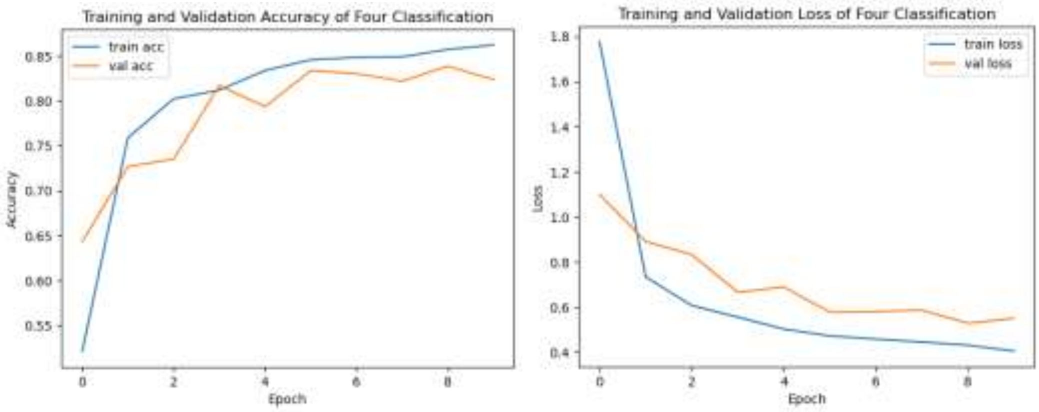

Fig. 3: Plots of MnasNet for the training and validation accuracy and loss

Discussion

After training, the MnasNet model has an accuracy of 83.16%. For each metric, an average value greater than 0.41 is shown. This suggests that there is a very low likelihood of it correctly forecasting the result which is not positive. This will make it easier to recognise the real positive and real negative. The early stopping function terminates the training procedure when the epoch approaches 10 and the Loss is not further reduced. Figure 3 displays both the training and validation accuracy graph, along with the training and validation loss graph of the MnasNet model for the classification of driver drowsiness. Additionally, Table 3 displays the recall, precision and accuracy values for MnasNet.

Conclusion

According to the classification model's output, MnasNet had an accuracy rate of 83.16%. For each metric, an average value greater than 0.41 is shown. This suggests that there is a very low likelihood of it correctly forecasting an outcome that is not positive. This will make it easier to recognise the true negative and the true positive. The early stopping function terminates the training procedure when the epoch approaches 10 and the loss is not further reduced. The MnasNet model has proven to be successful at spotting driver drowsiness. The system can still be improved, and future work may concentrate on incorporating more advanced techniques and combining video and image datasets to give the model access to more representation and additional signs of drowsiness. This would further improve the system's accuracy and reliability.

References

Adedotun, A. F., Adesina, O. S., Onasanya, O. K., Onos, E. S., & Onuche, O. G. (2022). Count Models Analysis of Factors Associated with Road Accidents in Nigeria. International Journal of Safety and Security Engineering, 12(4), 533–542. https://doi.org/10.18280/ijsse.120415

Albadawi, Y., AlRedhaei, A., & Takruri, M. (2023). Real-time machine learning-based driver drowsiness detection using visual features. Journal of imaging, 9(5), 91.

Cavuoto, L., & Megahed, F. (2016, June). Understanding fatigue and the implications for worker safety. In ASSE Professional Development Conference and Exposition (pp. ASSE-16). ASSE.

Cui, J., Lan, Z., Liu, Y., Li, R., Li, F., Sourina, O., & Müller-Wittig, W. (2022). A compact and interpretable convolutional neural network for cross-subject driver drowsiness detection from single-channel EEG. Methods, 202, 173-184.

Federal Road Service Corp. (2023). FRSC statistical digest [online]. Available from: https://frsc.gov.ng/wp-content/uploads/2023/05/FRSC-STATISTICS-DIGEST-FOURTH-QUARTER-2022-CORRECTED.pdf pg. 13 (Accessed: 24 May, 2023)

Ghourabi, A., Ghazouani, H., & Barhoumi, W. (2020). Driver drowsiness detection based on joint monitoring of yawning, blinking and nodding. In 2020 IEEE 16th International Conference on Intelligent Computer Communication and Processing (ICCP) (pp. 407-414). IEEE.

Li, Y. (2017). Deep reinforcement learning: An overview. arXiv preprint arXiv:1701.07274.

Lotfy, N. E. B. M., & Saparon, A. (2020, December). Detecting drowsy driver using photoplethysmography sensor. In AIP Conference Proceedings (Vol. 2306, No. 1). AIP Publishing.

Mahajan, K., Velaga, N. R., Kumar, A., Choudhary, A., & Choudhary, P. (2019). Effects of driver work-rest patterns, lifestyle and payment incentives on long-haul truck driver sleepiness. Transportation research part F: traffic psychology and behaviour, 60, 366-382.

Nabil, A. (2018). *The Detection Of Eyes and Classification*. (Technical University of Ostrava, Media Research Lab), ISVC. https://www.kaggle.com/code/abdelaziznabil/the-detection-of-eyes-and-classification

Ramprasath, M., Anand, M. V., & Hariharan, S. (2018). Image classification using convolutional neural networks. International Journal of Pure and Applied Mathematics, 119(17), 1307-1319.

Singh, J., Kanojia, R., Singh, R., Bansal, R., & Bansal, S. (2023). Driver Drowsiness Detection System: An Approach By Machine Learning Application. arXiv preprint arXiv:2303.06310.

Sunagawa, M., Shikii, S. I., Nakai, W., Mochizuki, M., Kusukame, K., & Kitajima, H. (2019). Comprehensive drowsiness level detection model combining multimodal information. IEEE Sensors Journal, 20(7), 3709-3717.

Vasilev, I., Slater, D., Spacagna, G., Roelants, P., & Zocca, V. (2019). Python Deep Learning: Exploring deep learning techniques and neural network architectures with Pytorch, Keras, and TensorFlow. Packt Publishing Ltd.

Vazquez, D. (2021). *Yawn Dataset*. https://www.kaggle.com/datasets/davidvazquezcic/yawn-dataset

Von Rueden, L., Mayer, S., Beckh, K., Georgiev, B., Giesselbach, S., Heese, R., & Schuecker, J. (2021). Informed machine learning–a taxonomy and survey of integrating prior knowledge into learning systems. IEEE Transactions on Knowledge and Data Engineering, 35(1), 614-633.

Yakubu, A.T., Ajide, F.M. & Abdulrahman, I.A.,(2023). Income Inequality and Road Transport Accidents in Nigeria. African Journal of Economic Review, 11(2), pp.43-54.

Zandi, A. S., Quddus, A., Prest, L., & Comeau, F. J. (2019). Non-intrusive detection of drowsy driving based on eye tracking data. Transportation research record, 2673(6), 247-257.