Indonesian Journal of Electrical Engineering and Computer Science
Vol. 37, No. 3, March 2025, pp. 1692~1701
ISSN: 2502-4752, DOI: 10.11591/ijeecs.v37.i3.pp1692-1701

A new deep learning model based on convolutional neural network and residual blocks for driver drowsiness detection

Abdelfettah Soultana, Faouzia Benabbou, Nawal Sael, Soukaina Bohsissin
Laboratory of Modeling and Information Technology, Faculty of Sciences Ben M'SIK, University Hassan II of Casablanca, Casablanca, Morocco

Article Info

Article history:

Received Jun 29, 2024
Revised Sep 19, 2024
Accepted Oct 7, 2024

Keywords:

Computer vision
Convolution neural network
Deep-learning
Driver drowsiness
Residual neural network

ABSTRACT

Recognizing the pivotal importance of monitoring driver inattention in the quest to minimize accidents and enhance safety and security, it is essential to highlight the inherent danger posed by drowsiness-a specific form of inattention that can significantly contribute to accidents. To address this issue, several propositions involving artificial intelligence have been put forth to effectively monitor and identify instances of driver drowsiness. However, challenges persist in the form of real-time processing constraints, the intricate nature of model parameters, and the response time of the model. The proposed methodology focuses on using convolutional neural networks (CNNs) and Residual blocks as robust and effective deep-learning models for the real-time detection of driver drowsiness. The integration of CNNs and Residual blocks enhances the model's precision, striking a well-balanced synergy between computational efficiency and performance. Notably, this approach demonstrated an impressive accuracy of 96.09% along with a recall, f1-score, and precision all at 96% when evaluated on the publicly available eye_dataset. Furthermore, the runtime of the developed model is a mere 70 ms. To further validate the efficiency of our proposed model, we conducted a comparative analysis with various pre-trained residual neural networks, including ResNet152, ResNet50, ResNet50v2, ResNet101, and ResNet101v2.

Corresponding Author:

Abdelfettah Soultana
Laboratory of Modeling and Information Technology, Faculty of Sciences Ben M'SIK
University Hassan II of Casablanca
Casablanca, Morocco
Email: abdelfettah.soultana-etu@etu.univh2c.ma

1. INTRODUCTION

Every year, a rising rate of accidents is observed globally. Several factors can contribute to this observation, including non-compliance with traffic regulations, traffic congestion, environmental conditions, and notably, the driver's state. Indeed, the driver's context can be a significant source of road accidents. Factors such as intoxication, negative emotions, and states of inattention, like distraction and drowsiness, lead to dangerous and disastrous situations.

Intelligent transportation systems (ITS) play a pivotal role in modernizing and enhancing road safety. These integrated technologies leverage data and communication systems to optimize traffic flow, improve driver behavior, and ultimately reduce accidents. One notable application involves the implementation of advanced driver monitoring systems (DMS), which utilize cutting-edge sensors and algorithms to track and analyze a driver's behavior and alertness. Additionally, ITS incorporates sophisticated alert systems, such as lane departure warning (LDW) [1] and forward collision warning (FCW) [2], which provide real-time notifications to drivers about potential hazards. Drowsiness is often understood to mean sleepiness in inappropriate contexts. Even if drowsiness lasts only a few minutes, the consequences can be devastating. Tired drivers are as dangerous as drunk drivers because they have slower reaction times and suffer from reduced attention, awareness, and ability to control their vehicles [3]. According to the US National Highway Traffic Safety Administration [4], 50,000 injuries and nearly 800 deaths have been reported, with 91,000 traffic accidents related to drowsiness. A study by the AAA Foundation for Traffic Safety estimated that 328,000 drowsy driving crashes occur annually [5]. The aforementioned alarming statistics have shown the necessity of a system for driver drowsiness monitoring and alerting, thereby preventing unfortunate traffic accidents from happening. Recently, many models have been developed for automatic drowsiness detection systems. Often, the inputs of the detection systems are images captured from a camera placed in front of the driver face. The eyes play a central role as the primary factor in detecting signs of fatigue, including drooping eyelids or extended periods of closure. Deep convolutional neural networks (CNNs) have recently performed exceptionally well on different tasks in computer vision, including object detection, image classification, image segmentation, object tracking, etc. In particular, several proposed drowsiness detection systems employed different convolutional network architectures and achieved effective and reliable results. The power of CNN models excel is the extraction of intricate features from images and videos, contributing to enhanced accuracy in identifying signs of drowsiness. Among the myriad features analyzed, the eyes emerge as a focal point in drowsiness detection. The identification and mitigation of drowsiness are addressed through diverse approaches, encompassing various aspects of both physiological and behavioral indicators. Physiologically, the activity of the brain is scrutinized using EEG, providing insights into the individual's cognitive state. Simultaneously, behavioral cues from the vehicle itself, such as lane departure patterns, contribute valuable information regarding the driver's attentiveness. Additionally, facial analysis serves as a key dimension in assessing drowsiness, capturing nuanced expressions and subtle signs that may indicate a decline in alertness...

Neural networks have become much more profound, with state-of-the-art networks [6], evolving from having just a few layers (e.g., AlexNet) to over a hundred layers. Detecting drowsiness with CNN introduces challenges that directly impact the model and its response time. Real-time constraints are a significant factor, demanding algorithms that can swiftly process video streams in a timely manner. The architecture's complexity plays a crucial role, as intricate models may hinder quick response times. Additionally, the number of trained parameters in CNNs directly influences the model's responsiveness, affecting its ability to promptly detect and respond to signs of drowsiness. Balancing the intricacy of the model with the need for rapid responses is essential for developing effective drowsiness detection systems and ensuring their practical utility in real-world scenarios. ResNet has accomplished exceptional execution in picture acknowledgment assignments, outperforming the capabilities of past CNN designs [7], [8]. During each iteration of training a neural network, all weights receive an update proportional to the partial derivative of the error function concerning the current weight. If the gradient is minimal, the weights will not be changed effectively and may stop the neural network from further training. The phenomenon is called vanishing gradients. More specifically, the data disappears through the layers of the deep neural network due to a very slow gradient descent. The integration of Residual blocks is motivated by the challenge of training deeper networks while avoiding the vanishing gradient problem [9]. Residual blocks, with their skip connections, facilitate the flow of gradients through the network, enabling the training of deeper architectures without sacrificing performance. Residual networks, like ResNet, have consistently achieved state-of-the-art results in significant computer vision benchmarks and competitions [10]. Their performance on tasks like ImageNet classification demonstrates their efficacy in learning intricate features from diverse image datasets. This strong track record reinforces their reliability and makes them a compelling choice for various computer vision tasks.The aim objective of this study is to effectively tackle the crucial challenge of driver drowsiness detection, a matter of utmost importance for road safety that necessitates real-time inference capabilities. Through the integration of CNNs with Residual blocks, we not only enhance the model's accuracy but also strike a delicate balance between computational efficiency and performance. This balance is especially crucial when considering the deployment of the model in resource-constrained environments, such as embedded systems within vehicles.

Many solutions have been presented to tackle the challenge of detecting driver drowsiness. In this context, we highlight a selection of recent works that contribute to this field. Chirra *et al.* [11] presented a system leveraging deep learning techniques for detecting driver drowsiness, focusing on assessing eye state. The initial steps of the approach involved employing the Viola-Jones algorithm for tasks encompassing face detection and localization of the eye regions. If the classifier detects ten consecutive instances of closed eyes, it indicates that the individual is exhibiting signs of drowsiness. A CNN model was trained using a dataset comprised of 2850 images, encompassing two distinct classes: open eyes and closed eyes, achieving an accuracy of 96.42%. CNN model was also used by Rafid *et al.* [12] to develop a drowsiness detection system. This approach employed the Viola-Jones algorithm for precise localization of the eyes. The drowsiness state if confirmed when the classifier identifies 30 consecutive instances of closed eyes. The CNN model achieved an accuracy rate of 94.5% using the MRL Eye Dataset. Shakeel *et al.* [13] introduced a MobileNet model to detect drowsiness cases. The methodology incorporated TensorFlow detection model zoo for both face detection and eyes localization. The model was trained on a dataset comprising 6000 images and achieved an accuracy of 76.9%. The detection of signs of drowsiness is confirmed if the classifier identifies ten consecutive instances of closed eyes. Rajamohana *et al.* [14] introduced an hybrid approach that combines a CNN model with Bidirectional long short-term memory (BiLSTM) algorithm to detect driver drowsiness based on eye state. The eyes region extraction process was facilitated by employing the DLIB library. The dataset, comprising 2208 images, served as the foundation for their research, leading to an accuracy rate of 94%. The methodology initiates a trigger action when eyes remain closed for an extended period. Jose *et al.* [15] used EfficientNetV2 lightweight convolutional network model for detecting drowsiness by employing a concurrently with facial landmark detection. Their approach involved the MediaPipe Face Mesh algorithm to extract eye regions. The model was trained using a dataset comprising 10,076 images, achieving an accuracy of 96.92%. A pivotal aspect of their strategy involves the decision-making process. The number of frames considered for confirming drowsiness detection is 60 frames. In their research, Bekhouche *et al.* [16] explored video-based drowsiness detection by leveraging the NTHU drowsy driver detection dataset (NTHU-DDD). Their investigation focused on identifying the most effective configurations concerning frame count and inter-frame skips. The experiments led to an accuracy of 86% when combining a ResNet-50 with support vector machine (SVM) classifiers. For precise eye region localization, Tiny-YOLO model was employed. The findings highlighted that the best configuration, considering computation time, entailed utilizing nine frames and skipping two inter-frames. Civik and Yuzgec [17] presented a CNN-based model for drowsiness detection, specifically targeting the eyes and mouth regions. The DLIB library was used to extract the facial regions, and the YawDD datasetto train the model exhibiting an accuracy of 94.5%. However, the specific threshold for defining drowsiness was not provided in this work. Florez *et al.* [18] focused specifically on eye closure duration as a crucial indicator to detect drowsiness. For precise eye region delineation, the MediaPipe Face Mesh algorithm was used, and for classification the InceptionV3, VGG16, and ResNet50V2 models, were employed. The best accuracy of 98.15% was obtained using the ResNet50V2 model on the NITYMED dataset, witha response time of 106.5 milliseconds. The tracking duration of eye closure is 300 milliseconds, and when exeeded it is a confirmation of drowsiness state. Phan *et al.* [19] compared multiple deep-learning models for drowsiness detection, including LSTM, VGG-16, DenseNet, and Inception-V3. The SSD-ResNet-10 model was employed for efficient eye region localization. Using a large dataset of 13,729 images, the fusion approach achieved 98% accuracy using DenseNet. However, the study did not provide the response time neither the threshold used to determine the drowsy state.

According to the analysis of the state of the art, various algorithms are deployed for drowsiness detection, including models like CNN, ResNet50V2, MobileNet, EfficientNetV2, and others. In terms of performance, the ResNet50V2 algorithm, used by Florez *et al.* [18], has demonstrated outstanding accuracy at 98.15%, showcasing its effectiveness. However, it's crucial to note that the minimum accuracy level may vary across studies, underscoring the importance of choosing an algorithm based on specific needs. While the performances are promising, considering the response time of the trained model is imperative. Some algorithms, despite their high accuracy, may have longer response times, a critical factor for real-time applications such as drowsiness detection while driving. Therefore, the selection of an algorithm should be guided not only by its intrinsic performance but also by the relevance of its response time in the intended usage context. ResNet addresses the challenges of deep network training by incorporating skip connections or residual blocks. This innovative design, seen in variants like ResNet50 and ResNet101, enables the training of much deeper networks, effectively mitigating issues such as vanishing gradients. The architecture not only enhances accuracy but also improves training efficiency, making it particularly suitable for applications where processing time is a critical factor. ResNet's ability to balance depth and efficiency renders it a compelling choice, especially in scenarios where processing time is a crucial consideration. Therefore, when considering algorithms for drowsiness detection, the unique features and advantages of ResNet should be taken into account alongside their respective performances.

The remainder of this paper is structured as follows. Section 2 presents the proposed methodology based on a new CNN architecture. Section 3 presents the results and discussion. Finally, section 4 provides a conclusion and future works.

2. RESEARCH METHOD

The objective of this research is to propose a CNN based model for drowsiness detection that produce better response time and that have less hyperparameters. The approach starts by capturing successive video frames from a camera and processing each following a set of steps. The initial steps involve the application of Haar cascades to detect facial and ocular features accurately. Specifically, pre-trained cascade classifiers are employed for recognizing facial contours as well as the distinctive characteristics of the eyes. A ResNet18 neural network model is introduced to determine the eyes' state. This model analyzes the extracted eye regions and predicts whether they are open or closed. As the video frames are processed, the code maintains a score that is incremented each time an instances of closed-eye is detected. When the score exceeds a predefined threshold of 10 consecutive images with the eyes closed, the system provides a signal to the driver about his state of drowsiness in real time. The Figure 1 presents two steps : the offline model construction which is the construction phase of the pretrained model and the online drowsiness detection that capture in real time driver video and assess the drowsiness cases.

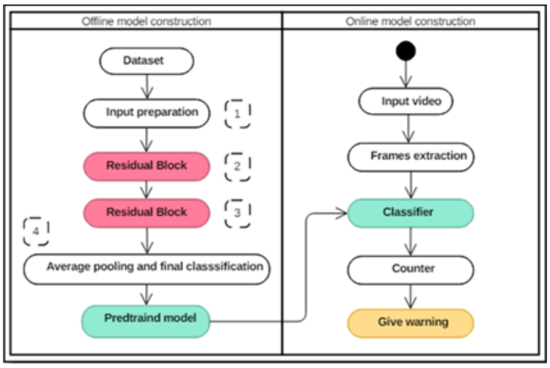


Figure 1. Block diagram of the proposed drowsiness detection system

2.1. Dataset

The deep learning model developed here is trained on images from the Kaggle eyes dataset [20]. The eyes dataset comprises a total of 4,850 grayscale images. Among these, 3,790 images are allocated for training, distributed across two classes: "closed" and "open" eyes. Additionally, 1,060 images are designated for testing, covering both "open" and "closed" eye classes. These images exhibit diversity in lighting conditions and subject variability and even include instances of individuals wearing glasses. Each image has dimensions of 24×24 pixels. Given these characteristics, this dataset proves to be well-suited for effectively training and constructing a resilient model capable of accurately tracking eye closure. The Figure 2 presents some examples of images from the dataset, showcasing both cases of open and closed eyes.

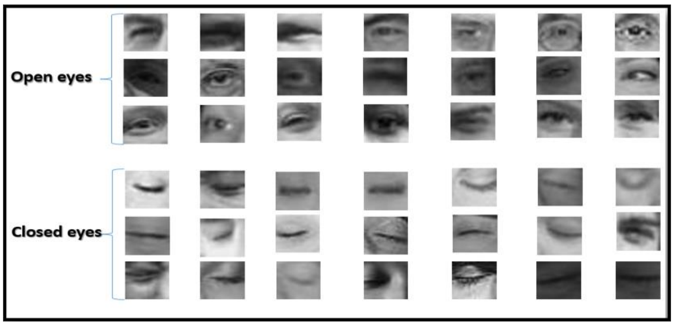


Figure 2. Representation of annotated images of the eyes_dataset

2.2. Offline model construction

The offline training diagram shows how the model is built. Indeed, the proposed model incorporates residual blocks to facilitate deeper representations learning and improve the model performance. The architecture consists of the following stages:

– Input preparation and initial convolutional layer: the input images have dimensions of (145, 145, 3) and a zero-padding is used to maintain size. Subsequently, a convolutional layer with a 7×7 kernel and a stride

of 2 is introduced to reduce the dimensions. Batch normalization and ReLU activation are applied, followed by max-pooling to downsize the feature maps further.
- Residual block (stage 2): the input passes through a residual block with two convolutional layers and a shortcut connection. The block enhances the model's ability to learn intricate features. Batch normalization and ReLU activation follow each convolution, and the result is added element-wise to the original input before a final ReLU activation.
- Residual block (stage 3): similar to stage 2, the input goes through another residual block with increased filters, creating a deeper understanding of features. The same pattern of batch normalization, ReLU activation, and element-wise addition is maintained.
- Average pooling and final classification: the average pooling layer reduces feature map dimensions, preparing them for the final classification. The flattened features are fed into a fully connected layer with a softmax activation function to output class probabilities.

2.3. Online drowsiness detection

The online drowsiness detection process initiates with input from a video stream, followed by frame-by-frame extraction. The Haar cascade algorithm is applied for face detection, and if a face is detected, the process proceeds to eye detection; otherwise, it moves to the next frame. After extracting the eye region using the Haar cascade, it undergoes classification using our pre-trained model, detailed in the first section, to determine the state of the eyes as open or closed. A counter is incremented when closed eyes are detected, and when it reaches a predetermined threshold over 10 consecutive frames, the system concludes that the driver is in a state of drowsiness. At this point, an alert is sent to the driver, notifying them of their drowsy state and prompting appropriate actions to ensure safety while driving. This continuous and real-time process ensures constant monitoring of the driver's condition during their journey.

2.4. Proposed model details

ResNet revolutionized the CNN architectural race by introducing the concept of residual learning in CNNs and devised an efficient methodology for the training of deep networks [21]. They are characterized by their deep architectures, which can produce low error rates. The basic building blocks of a residual neural network are the so-called residual blocks. The basic idea here is that so-called “skip connections” are built into the network. These ensure that the activation of a layer is added together with the output of a later layer. Limonova *et al.* [22] has conclusively demonstrated that the implementation of the ResNet model demands notably fewer logic gates, reducing the requirement by a factor of 2.1-2.9 times in comparison to alternative models. Residual network (ResNet) is a network model with better performance and smaller model size and parameters. It occupies less memory space while maintaining accuracy and has strong modularity and portability. This architecture allows the network to skip specific layers significantly if they do not contribute to a better result. A residual neural network is composed of several of these so-called residual blocks [23], [24]. In ResNets, two primary block types come into play based on whether the input and output dimensions align or differ. Both block types, namely the "identity block" and the "convolutional block," are integral components of the implementation. The identity block is the standard choice for ResNets, employed when the input and output activations share the same dimensions. Below, we illustrate an example of an identity block and convolutional block in our model featuring the "shortcut path" and the "main path." These two blocks together form the foundational architecture of the proposed model [25]. The convolutional block captures intricate features, while the Identity Block ensures efficient gradient flow, allowing for the adequate training and optimization of deep networks. The RES block incorporates both the convolution block and the identity Block, as depicted in Figure 3, to form an integral part of the final model architecture.

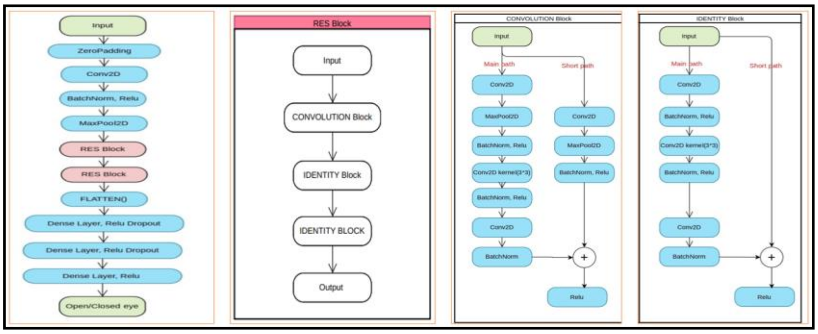


Figure 3. Final model with ResNet block, identity block, and convolution block

3. RESULTS AND DISCUSSION

In this section of results and discussion, we have presented the various outcomes of the models trained on the eyes_dataset, emphasizing four metrics: accuracy, precision, recall, and F1-score. Additionally, we have displayed the confusion matrix for the most proficient models along with their respective model loss and model accuracy. Additionally, a discussion of these results is provided.

3.1. Performance metrics

Several performances were used to evaluate the proposed model. The confusion matrix measures the quality of a classification system based on scores of true positive (TP), true negative (TN), false positive (FP), and false positive (FP). Additionally, accuracy, precision, recall and F1-score were captured to have a global vison of model performance:

$$\text{Accuracy}(\%) = \frac{TP+TN}{TP+FP+FN+TN} \times 100 \quad (1)$$

$$\text{Recall}(\%) = \frac{TP}{TP+FN} \times 100 \quad (2)$$

$$\text{Precision}(\%) = \frac{TP}{TP+FP} \times 100 \quad (3)$$

$$\text{F1-score}(\%) = \frac{2\times\text{Precision}\times\text{Recall}}{\text{Precision}+\text{Recall}} \times 100 \quad (4)$$

3.2. Experiment results

In addition of the proposed model evaluation using key metrics such as accuracy, precision, recall, and F1-score, a comparison with the performance of other pre-trained ResNet models is conducted including ResNet152, ResNet50, ResNet50v2, ResNet101, and ResNet101v2. The Table 1 provides the performance comparison of the proposed model and the ResNet pre-trained models. Additionally, the table includes information on response time in milliseconds and total parameters, providing a comprehensive overview of computational efficiency and complexity of each model.

Table 1. Report classification

Algorithms	Accuracy	Precision	recall	F1-score	Response time (ms)	Total params
ResNet152	51.95%	74%	51%	35%	150	58 633 345
ResNet50	95.31%	96%	96%	96%	110	23 850 113
ResNet50v2	96.48%	96%	96%	96%	100	23 827 201
ResNet101	94.14%	95%	95%	95%	120	42 920 577
ResNet101v2	94.92%	95%	95%	95%	120	42 888 961
Proposed model	96.09%	96%	96%	96%	70	1 194 114

The proposed model, ResNet50, ResNet50v2, ResNet101, and ResNet101v2, exhibit strong and balanced performance across accuracy, precision, recall, and F1-score, making them reliable choices for drowsiness detection task. Our model achieved a high accuracy of 96.09%, indicating its capability to classify instances correctly. Additionally, the precision, recall, and F1-score are all equally impressive with a score of 96%, reflecting a well-rounded ability to correctly classify positive instances in the dataset. ResNet152 performed relatively high precision of 74%, but it suffers from a lower recall and F1-score of 51% and 35%, respectively, resulting in an overall accuracy of 51.95%. This suggests that although the algorithm is precise in classifying positive instances, it needs help to capture a substantial portion of relevant instances from the dataset. In contrast, ResNet50, ResNet50v2, and ResNet101, demonstrate similarly high levels of accuracy ranging from 94.14% to 96% and ResNet50v2 outperformed with an accuracy of 96.48%. All precision, recall, and F1-scores are all consistent at 96%, indicating balanced performance between accurately classifying positive instances and comprehensively capturing relevant instances from the dataset. Considering execution time, the proposed model obtained a remarkable response time of just 70 milliseconds, a significant leap in speed compared to all other models in the comparison. This translates into a level of processing swiftness that is crucial in real-time or time-sensitive applications, setting it apart as an exceptional choice. In contrast, while other models like ResNet50 and ResNet50v2 exhibit commendable performance, their response times of 110 and 100 milliseconds respectively, while still respectable, lag behind the rapidity achieved by our approach. In addition, our model has the smallest number of parameters highlighting its efficiency and resource optimization. This underscores the high equilibrium struck by the proposed model, seamlessly marrying high performance with optimal response time, rendering it a highly promising candidate for applications where speed is critical. The Figure 4 illustrates the model loss and accuracy for the two models that exhibit the most significant metric values.

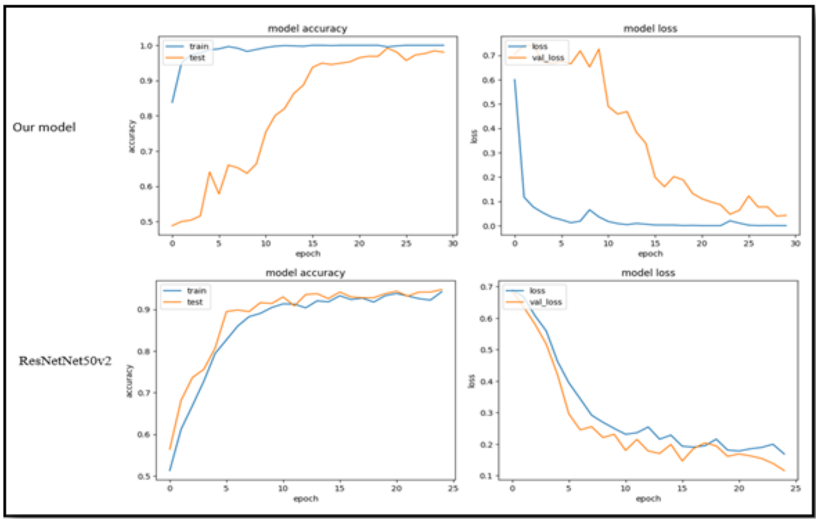


Figure 4. Model loss and model accuracy

The graph shows that the precision and loss of our model have a learning progression over 30 epochs. The model's training accuracy starts at around 79.8% and steadily improves, reaching close to 100% accuracy by the 13th epoch. The validation accuracy, which begins at approximately 49.2%, undergoes a substantial improvement and stabilizes at a high level of 96% accuracy after a few epochs. This trend indicates that the model effectively learns from the training data and generalizes well to the validation data. However, the relatively small gap between the training and validation accuracy suggests the model could be well-regularized and continue to perform well on new, unseen data. The provided model accuracy and model loss for ResNetNet50v2 illustrate the learning trajectory of a neural network across 25 epochs. Initially, the model achieves a training accuracy of 86.5%, which subsequently advances consistently, culminating in an accuracy of around 96.48% by the 25th epoch. Meanwhile, the validation accuracy starts at 58.6% and displays some oscillations in the early epochs. However, the fluctuations stabilize, leading to a convergence of around 96.48% as the training progresses. This convergence signifies the model's robust capacity to generalize effectively to new and unseen data. The confusion matrices for the two models are also displayed in the Figure 5.

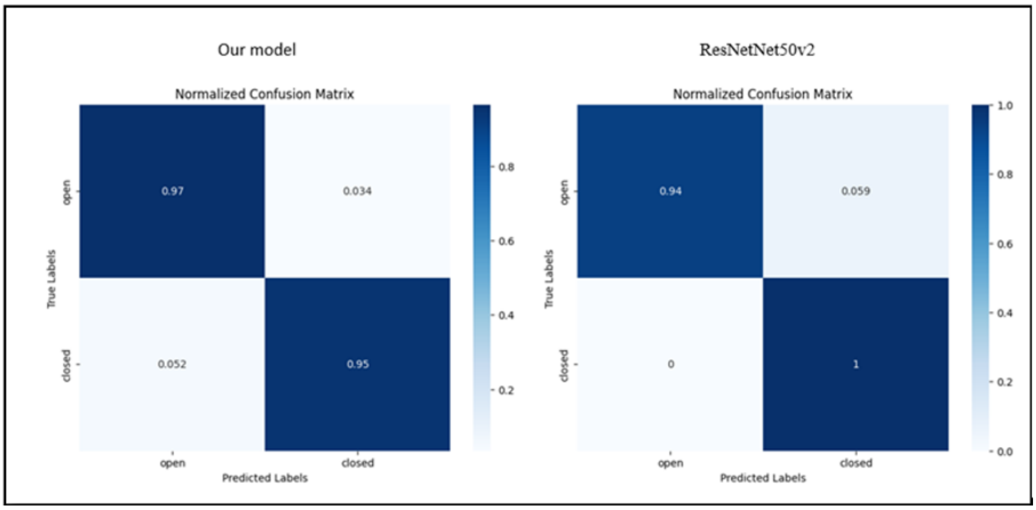


Figure 5. Confusion matrix

This matrix shows that the proposed model performs quite well overall, with high values on the diagonal indicating accurate predictions. The model seems slightly more likely to misclassify "closed" instances as "open" than the reverse. The "ResNet50v2" model also performs well, particularly in the "closed" class, where it achieves perfect accuracy. However, there are some instances where it misclassifies "open" instances as "closed," leading to a non-zero value in the top right corner. In conclusion, both models perform reasonably well in the binary classification task but have slightly different misclassification patterns. The proposed model has a balanced misclassification pattern, while ResNet50v2 excels in the closed class but has some misclassifications in the open category. In conclusion, both models, the proposed model and ResNet50v2, demonstrate promising performance for drowsiness detection. The proposed model stands out with its faster response time and reduced number of parameters, making it a preferred choice for real-time drowsiness detection in drivers. The proposed model demonstrates exceptional performance on the primary dataset, achieving an accuracy of 96.09%. This high level of accuracy is supported by various evaluation metrics, including precision, recall, F1 score, and a detailed confusion matrix. The stability of the training process, as evidenced by consistent loss and accuracy profiles, further validates the model's effectiveness.

In addition to these metrics, a comparative analysis was conducted with several pre-trained CNN models based on the ResNet architecture. The results of this comparison reveal that our proposed 18-layer model significantly outperforms these alternatives, achieving a compelling level of accuracy. This establishes our model's superior efficacy in the domain of driver drowsiness detection. To assess the model's robustness and generalization capabilities, we conducted a secondary evaluation using an alternative Kaggle dataset comprising 1,870 images depicting both open and closed eyes. Remarkably, the model achieved an accuracy of 98.28% on this independent dataset. This impressive performance on a different dataset underscores the model's robustness and broad applicability, making it a reliable solution for various real-world scenarios.

4. CONCLUSION

In conclusion, addressing the critical issue of drowsy driving in the realm of driver safety demands innovative solutions, particularly those grounded in real-time monitoring systems. This research introduces a groundbreaking approach utilizing convolutional neural networks (CNNs) and Residual blocks to promptly detect driver drowsiness. The fusion of CNNs with Residual units, incorporating crucial analytical components, establishes a robust methodology for addressing this critical challenge. Our devised strategy exhibits remarkable efficacy, boasting an impressive 96.09% accuracy when tested on the widely accessible eye_dataset. Notably, our proposed approach demonstrates an exceptional response time of 70 ms, underscoring its efficiency in real-time applications. Furthermore, our investigation extends to the exploration of diverse pre-trained residual neural networks, including ResNet152 and ResNet50, emphasizing the adaptability of our methodology. Looking forward, our envisioned trajectory involves the seamless integration of our drowsiness detection system into a comprehensive global driver inattention detection system. This expanded system transcends the singular focus on drowsiness, encompassing factors such as fatigue, negative emotions, and driver distraction. By broadening its scope, we aspire to establish a more resilient and proactive approach to bolstering driver safety and attentiveness on the roads. This holistic strategy, coupled with an impressive response time, reflects our commitment to advancing not only the field of drowsy driving detection but also contributing to a broader framework for comprehensive driver safety solutions. Future research can build upon these findings by exploring real-time implementation, enhancing dataset diversity, integrating multimodal data, investigating transfer learning, and evaluating long-term performance. This holistic strategy, coupled with an impressive response time, reflects our commitment to advancing not only the field of drowsy driving detection but also contributing to a broader framework for comprehensive driver safety solutions.

REFERENCES

[1] S. Sternlund, J. Strandroth, M. Rizzi, A. Lie, and C. Tingvall, "The effectiveness of lane departure warning systems-a reduction in real-world passenger car injury crashes," *Traffic Injury Prevention*, vol. 18, no. 2, pp. 225–229, Feb. 2017, doi: 10.1080/15389588.2016.1230672.

[2] L. Yue, M. Abdel-Aty, Y. Wu, J. Ugan, and C. Yuan, "Effects of forward collision warning technology in different pre-crash scenarios," *Transportation Research Part F: Traffic Psychology and Behaviour*, vol. 76, pp. 336–352, Jan. 2021, doi: 10.1016/j.trf.2020.12.004.

[3] L. R. Williams, D. R. Davies, K. Thiele, J. R. Davidson, and A. W. MacLean, "Young drivers' perceptions of culpability of sleep-deprived versus drinking drivers," *Journal of Safety Research*, vol. 43, no. 2, pp. 115–122, 2012, doi: 10.1016/j.jsr.2012.02.002.

[4] "Drowsy driving," NHTSA. Accessed: Aug. 13, 2023. [Online]. Available: https://www.nhtsa.gov/risky-driving/drowsy-driving

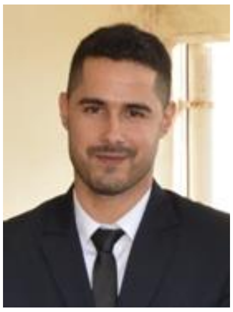
[5] B. C. Tefft, "Prevalence of motor vehicle crashes involving drowsy drivers, United States, 1999-2008," *Accident Analysis and Prevention*, vol. 45, pp. 180–186, Mar. 2012, doi: 10.1016/j.aap.2011.05.028.

[6] Y. Lecun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015, doi: 10.1038/nature14539.

[7] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, Jun. 2016, pp. 770–778. doi: 10.1109/CVPR.2016.90.

[8] M. Shafiq and Z. Gu, "Deep residual learning for image recognition: a survey," *Applied Sciences (Switzerland)*, vol. 12, no. 18, p. 8972, Sep. 2022, doi: 10.3390/app12188972.
[9] F. Huang, J. T. Ash, J. Langford, and R. E. Schapire, "Learning deep ResNet blocks sequentially using boosting theory," in *35th International Conference on Machine Learning, ICML 2018*, 2018, pp. 3272–3290.
[10] Z. Wu, C. Shen, and A. van den Hengel, "Wider or deeper: revisiting the ResNet model for visual recognition," *Pattern Recognition*, vol. 90, pp. 119–133, Jun. 2019, doi: 10.1016/j.patcog.2019.01.006.
[11] V. R. Reddy Chirra, S. R. Uyyala, and V. K. Kishore Kolli, "Deep CNN: a machine learning approach for driver drowsiness detection based on eye state," *Revue d'Intelligence Artificielle*, vol. 33, no. 6, pp. 461–466, Dec. 2019, doi: 10.18280/ria.330609.
[12] A. U. I. Rafid, A. I. Chowdhury, A. R. Niloy, and N. Sharmin, "A deep learning based approach for real-time driver drowsiness detection," in *2021 5th International Conference on Electrical Engineering and Information and Communication Technology, ICEEICT 2021*, IEEE, Nov. 2021, pp. 1–5. doi: 10.1109/ICEEICT53905.2021.9667944.
[13] M. F. Shakeel, N. A. Bajwa, A. M. Anwaar, A. Sohail, A. Khan, and Haroon-ur-Rashid, "Detecting driver drowsiness in real time through deep learning based object detection," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 11506 LNCS, 2019, pp. 283–296. doi: 10.1007/978-3-030-20521-8_24.
[14] S. P. Rajamohana, E. G. Radhika, S. Priya, and S. Sangeetha, "Driver drowsiness detection system using hybrid approach of convolutional neural network and bidirectional long short term memory (CNN_BILSTM)," *Materials Today: Proceedings*, vol. 45, pp. 2897–2901, 2021, doi: 10.1016/j.matpr.2020.11.898.
[15] J. Jose, A. J, K. Raimond, S. Vincent, and J. Eunice R, "SleepyWheels: an ensemble model for drowsiness detection leading to accident prevention," *Journal of Autonomous Intelligence*, vol. 7, no. 3, Jan. 2024, doi: 10.32629/jai.v7i3.1117.
[16] S. E. Bekhouche, Y. Ruichek, and F. Dornaika, "Driver drowsiness detection in video sequences using hybrid selection of deep features," *Knowledge-Based Systems*, vol. 252, p. 109436, Sep. 2022, doi: 10.1016/j.knosys.2022.109436.
[17] E. Civik and U. Yuzgec, "Real-time driver fatigue detection system with deep learning on a low-cost embedded system," *Microprocessors and Microsystems*, vol. 99, p. 104851, Jun. 2023, doi: 10.1016/j.micpro.2023.104851.
[18] R. Florez, F. Palomino-Quispe, R. J. Coaquira-Castillo, J. C. Herrera-Levano, T. Paixão, and A. B. Alvarez, "A CNN-based approach for driver drowsiness detection by real-time eye state identification," *Applied Sciences (Switzerland)*, vol. 13, no. 13, p. 7849, Jul. 2023, doi: 10.3390/app13137849.
[19] A. C. Phan, T. N. Trieu, and T. C. Phan, "Driver drowsiness detection and smart alerting using deep learning and IoT," *Internet of Things (Netherlands)*, vol. 22, p. 100705, Jul. 2023, doi: 10.1016/j.iot.2023.100705.
[20] S. Raju, "yawn_eye_dataset_new." Accessed: Aug. 06, 2023. [Online]. Available: https://www.kaggle.com/datasets/serenaraju/yawn-eye-dataset-new
[21] A. Khan, A. Sohail, U. Zahoora, and A. S. Qureshi, "A survey of the recent architectures of deep convolutional neural networks," *Artificial Intelligence Review*, vol. 53, no. 8, pp. 5455–5516, Dec. 2020, doi: 10.1007/s10462-020-09825-6.
[22] E. Limonova, D. Alfonso, D. Nikolaev, and V. V. Arlazarov, "ResNet-like architecture with low hardware requirements," in *Proceedings - International Conference on Pattern Recognition*, IEEE, Jan. 2020, pp. 6204–6211. doi: 10.1109/ICPR48806.2021.9413186.
[23] Y. Zhong, S. Qiu, X. Luo, Z. Meng, and J. Liu, "Facial expression recognition based on optimized ResNet," in *2020 2nd World Symposium on Artificial Intelligence, WSAI 2020*, IEEE, Jun. 2020, pp. 84–91. doi: 10.1109/WSAI49636.2020.9143287.
[24] J. Zhang, Y. Zhu, W. Li, W. Fu, and L. Cao, "DRNet: a deep neural network with multi-layer residual blocks improves image denoising," *IEEE Access*, vol. 9, pp. 79936–79946, 2021, doi: 10.1109/ACCESS.2021.3084951.
[25] F. He, T. Liu, and D. Tao, "Why ResNet works? residuals generalize," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 31, no. 12, pp. 5349–5362, Dec. 2020, doi: 10.1109/TNNLS.2020.2966319.

BIOGRAPHIES OF AUTHORS



Abdelfettah Soultana recently obtained his doctorate in computer science. He completed his master's degree in software quality at Hassan II University, Casablanca, Morocco, in 2015. His thesis, titled "Towards a contextual system for smart car management," marked a significant milestone in his academic journey. Currently, he is deeply engaged in his doctoral studies at the Laboratory of Information Processing and Modeling (LTIM) within the Ben M'sik Faculty of Science. His research focuses on machine learning, deep learning, and the application of Internet of Things technologies for driver context monitoring. He can be contacted at email: soultana.abdelfettah@gmail.com.

Faouzia Benabbou teacher-researcher since 1994, Authorized Professor since 2008 and Professor of Higher Education in the Department of Mathematics and Computer Science at the Ben M'Sick Faculty of Sciences in Casablanca since 2015. She is a member of the Information Technology and Modeling Laboratory and leader of the ICCNSE (Cloud Computing, Network and Systems Engineering) team. His research areas include cloud Computing, data mining, machine learning, and natural language processing. She can be contacted at email: faouzia.benabbou@univh2c.ma.

Nawal Sael teacher-researcher since 2012, Authorized Professor since 2014 and Professor of Higher Education in the Department of Mathematics and Computer Science at the Ben M'Sick Faculty of Sciences in Casablanca since 2020 and her engineer degree in software engineering from ENSIAS, Morocco, in 2002. Here research interests include data mining, educational data mining, machine learning, deep learning and internet of things. She can be contacted at email: saelnawal@hotmail.com.

Soukaina Bouhsissin received a B.Sc.degree in mathematical sciences and computer science from the Faculty of Sciences Ben M'Sick, Hassan II University of Casablanca, Morocco, in 2018, and an M.Sc. degree in data science andbig data from Hassan II University of Casablanca, in 2020, where she is currently pursuing the Ph.D.degree in computer science. Her research interests include driver behavior classification, intelligent transport systems, machine learning, deep learning, image classification, and temporal series. She can be contacted at email: bouhsissin.soukaina@gmail.com.