Generalized Frame-Based Residual Neural Network Model for Drowsiness Detection

GASSER ELAZAB LIVIA MARIA WALLER

elazab | lmwaller @kth.se

February 12, 2022

Abstract

Drowsiness, the state of the transition between alertness to sleeping, can greatly reduce our ability to concentrate. Impaired concentration is especially a problem when driving a vehicle. Thus, drowsiness is one of the main leading causes of car accidents. Drowsiness detection is, therefore, a major concern when it comes to improving the safety of traffic. Current research of drowsiness detection explores different measurement approaches e.g. driving behaviour, physiological signals and body expressions. Frame- or video-based drowsiness detection using the driver's body expressions.

Most frame-based drowsiness detection approaches focus on certain drowsiness signs, which is an obstacle for implementing drowsiness detection in real-life. This project sets out to approach drowsiness detection from a generalized angle. Generalized drowsiness detection is not limited to specific body expression of the driver but can detect as many drowsiness signs as possible, which greatly improves the detection flexibility.

To solve this problem, a deep learning model based on the residual neural network (ResNet) architecture is built and trained on three different datasets and tuned for accuracy. Each dataset displays images of different drowsiness signs. Our model detects the drowsiness signs of each dataset as a multi-classification problem. This approach detects the various drowsiness signs with 94.7% accuracy, which is competitive in comparison to other comparable generalizability approaches for detecting drowsiness.

In this project a generalized drowsiness detection approach was presented for frame-based data with much potential to be expanded to more drowsiness signs and therefore enhanced generalizability.

.

Contents

List of Acronyms and Abbreviations

CNN convolutional neural network

ResNet residual neural network

1 Introduction

1.1 Theoretical Framework/Literature Study

First, the word ”drowsiness” should be defined to address the problem in detail. Drowsiness describes the state of the transition from awake to sleep, such as sleep deprivation or general fatigue, which leads to an inability to concentrate on driving. Discovering the criticality of this problem goes back to the 80s [1]. Driver drowsiness systems have been and are still a vital application to work on, as it is the cause of a lot of car accidents. As an example, a study was done in Great Britain in 1995 that 20% of serious car accidents happened because of driver fatigue [2]. Moreover, recent studies analyzing the driver's behaviour leading to accidents state the two main reasons as: drowsiness and distraction [3].

A lot of research focuses on the reasons behind drowsiness, such as the research done by the national heart et al. [4], they concluded some reasons behind the drowsiness of the drivers who had car accidents as driving at night (from 1:00 am to 7:00 am), driving alone, or being a young male. Despite the expansion of autonomous vehicles, a lot of different models to detect driver drowsiness have been developed in the last decade, these models address the problem in different ways and this is going to be illustrated in the next section.

Previous studies were mainly split into three main categories [5]:-

- Detecting the drowsiness of the driver based on his/her driving behaviour
- Studying the physiological signals of the driver using the electrical signals from the brain using various sensors
- Judging the drowsiness by the facial expressions of the driver, such as yawning or blinking frequency

Explaining the previous state-of-the-art methods to detect driver drowsiness is going to be discussed based on these three categories. Nonetheless, this is going to be beneficial to categorize our work, in the right research field, for a proper comparison in the future.

1.1.1 Drowsiness Detection Based on the Driving Behaviour

Forsman et al. made standardized driving experiments, driving in daylight on rural highways without obstacles, to be conducted by the participants [6]. Their procedure was to do a 10 minutes long simple reaction time task, a task to measure attention done by Lim and Dinges [7]. This task is to label the driver as drowsy or an awake driver, and all the driving behaviours were recorded, such as lane position, steering wheel angle, and driving speed [6]. They performed principal component analysis on the resulting data, and it shows that detecting drowsiness can be analyzed using two principal components only. These principal components were observed to be steering variability and lane variability. The pain point of their method is using standardized driving experiments, which may not provide a generalize-able model to be implemented. Also, in the era of autonomous driving, relying on the driving behaviour for judging the drowsiness of the driver is not always answering the original question of drowsiness. As an example, if the driver is operating the auto-pilot mode, this model will not be able to detect the drowsiness of the driver. However, it is still a valid model to be integrated with another method for better drowsiness detection.

1.1.2 Drowsiness Detection Based on the Driver's Physiological Signals

Another approach was done by measuring heart rate variability-rate with electroencephalography(EEG) by Fujiwara et al. [8], they proposed a discriminative model trained by drowsy and awake data. However, as drowsy data was not a trivial task, they formulated the same problem as anomaly detection from the awake data. They utilized a neural-network architecture, previously done by Patel et al. [9] to assess driver fatigue. Their results detected 12 out of 13 pre-N1, which is the pre-sleep stage [10]. However, there were also false-positive rate of 1.7 times per hour. The challenges for this research were the controlled environment of the experiments and the limited data collection from Japanese persons [8]. Although this model detected fatigue in accurate form of EEG signals, putting electrodes on driver's skin is not always applicable for commercial car companies.

Lee and Chung [11] proposed an innovative approach to overcome the barrier of putting electrodes during driver, while still using physiological signals at the same time. This approach was about judging based on monitoring the pulse rate and respiratory signals, these signals were derived from a smartwatch to detect driver's drowsiness. Although the methodology of implementing this model is quite similar to the approach of Fujiwara et al. [8], they were constrained by the computational capacity of the smart watch. They achieved accuracy of 99.2% using KPCA-SVM. However, it was too complicated to be practically possible to implemented on a smart watch. So, the most appropriate model was using M-SVM, and it achieves accuracy of 95.8%. Despite the fact that smart watch was a good solution for the mentioned problem of utilizing EEG signals, this solution can not be categorized as generalize-able one, as it depends on wearing a watch, which cannot be guaranteed.

1.1.3 Drowsiness Detection Based on Driver's Body Expressions

A lot of different approaches were taken to detect the driver's drowsiness. As an example, Kaplan et al. utilized conventional image processing techniques to detect the fatigue on the facial images of the driver [12]. They utilized a driving simulator in order to control the illumination, noise, and temperature. The main idea is to detect the face using Viola-Jones algorithm [13], and detect each time the eye blinks. After that, they measured the frequency of blinking as a signal, and these signals are going to be input for training a neural network. Although, they achieved sufficient accuracy, controlling the illumination and the condition of the driving simulator may hinder utilizing this model in real life.

There was a conducted research that yielded an algorithm for drowsiness detection by Abtahi, et al. [14]. This model was a seed for relying on detecting the yawning to be categorized as a sign of drowsiness. This model was simple and good at detecting the yawning phenomena. However, it was relying on the assumption that the driver's face is fixed on the camera's position. Subsequently, a deep learning model serving the save goal was conducted by Zhang et al. [15]. They utilized face, nose, and mouth detector in order to detect yawning efficiently. Their model achieved an accuracy of 92% in yawning detection.

One of the most recent publications in this field was the deep learning model for drowsiness detection by Rajkar, Kulkarni, and Raut [16]. They utilized two data sets for yawning [17] and closing eyes [18] to train a Convolutional Neural Network (CNN) to detect the drowsiness based on these two factors. They have achieved an accuracy of 96% of detecting the yawning and closing eyes. However, their drowsiness definition is based on these two factors only, while drowsiness can result from being drunk, general fatigue, and other forms of tiredness. Although their model accuracy is relatively high and their performance in real-time is sufficient, their model is well-suited for detecting these two features solely. That is why our motivation for this paper is to extend the applicability of drowsiness detection to be generalize-able for any other form of drowsiness.

1.2 Research Question

In this paper, we explore the possibility of generalization for the detection of drowsiness. As drowsiness has many forms of bodily and facial expression, our first question is if the drowsiness can be detected with a deep learning model from more generalized input data. Secondly, we will evaluate if the same precision as non-generalized models can be achieved with our generalized drowsiness detection model.

2 Method

2.1 Data Collection and Processing

For the generalization (see 2.2) aspect of the experiment, we collected three labeled datasets available from open-source data repositories or provided by research groups. The fatigue dataset and the eyes dataset are image-based, while the yawning Detection" dataset is video-based. These datasets were chosen because of their availability and their dissimilar focus on different facial features that can indicate drowsiness.

Datasets:

- Fatigue dataset [19]
- Yawning Detection dataset [17]
- Closed Eyes in the Wild dataset (Cew) [18]

We combined the datasets into one folder, omitting the split between training, test, and evaluation while keeping their respective classes. The yawning dataset was processed with a Python script to turn the video data into single frame data.

2.1.1 Data Preprocessing

Data preprocessing is always an essential step before training, especially in our implementation. The utilized data preprocessing steps are oriented to standardize the images between and within the different datasets. Normalizing the images ensures comparability between the images and minimizes the differences between the combined datasets. The preprocessing steps are the following:

- The images are turned into greyscale as colors are not indicators for drowsiness, and it greatly reduces computation time and effort.
- All the images are downsampled to 28x28 pixels to improve the computational time of training.
- The input images are normalized to adjust the pixel values into the range of -1 and 1.

In addition, the dataset is randomly split into training and test set, with 10% of the dataset assigned to the test set.

2.1.2 Quantitative Analysis

The quantitative analysis helps to explore the distribution of the frequency of the images among the categories and datasets. This will help us find where bias could be introduced because of the distribution among the classes and datasets as well as because of the size of the datasets.

Dataset	FatigueDD	YawningDD	CewDD
Total	3907	1450	2423
Training	3517	1300	2183
Test	390	150	240

Table 1: Comparison of dataset size

In the table 1 and figure 1, the distribution among the three dataset is shown. It becomes clear that FatigueDD and CewDD are over-proportionally represented. In 3 one can see the distribution among the classes and it shows that four classes are equally weighted whereas class "tired" over-represented and the classes of the YawningDD are under-represented.

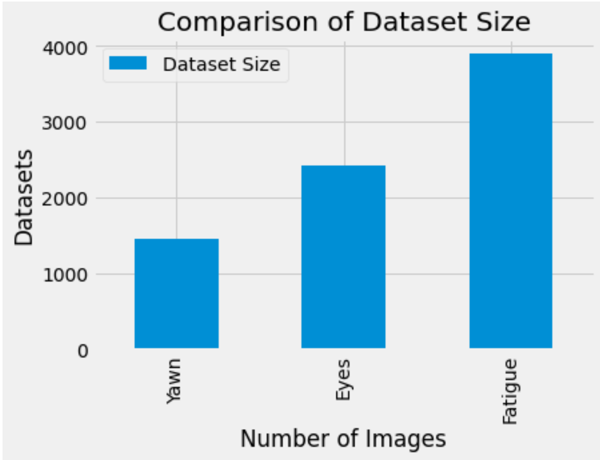


Figure 1: Partition among datasets

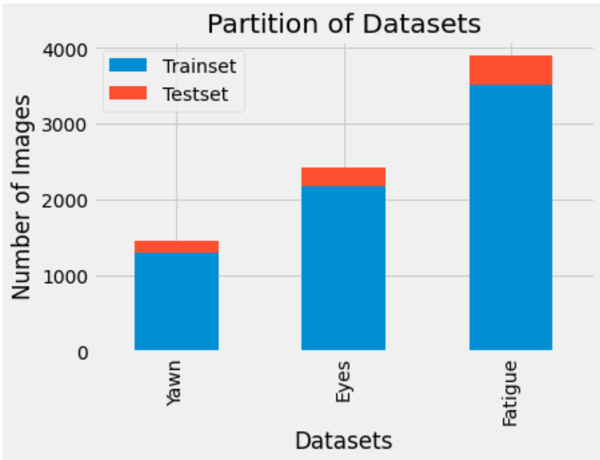


Figure 2: Partition among datasets

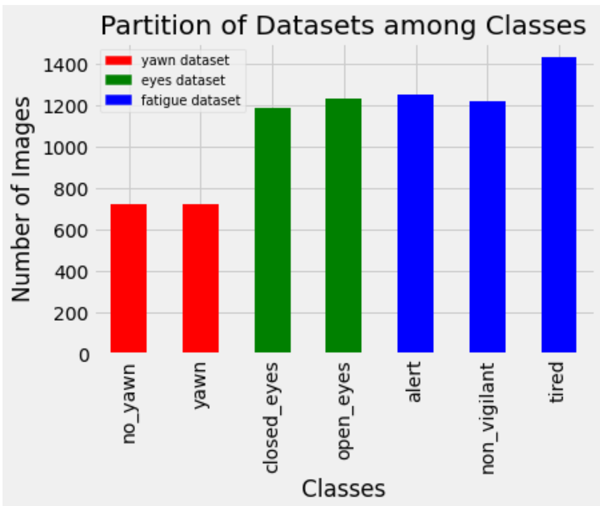


Figure 3: Partition among datasets

Concluded from this analysis, is that the data is mostly evenly distributed within the single dataset but not among all datasets. We have to be mindful of the bias that could be introduced when training with the combined dataset.

2.2 Concept of Generalized Drowsiness

The drowsiness of humans can have various symptoms and facial expressions. However, each feature alone is not a reliable sign of drowsiness. Yawning can be a sign of drowsiness but it does not mean that the driver is falling asleep. While having the eyes closed over a longer period is a sure way to detect that a driver is asleep but the eyes can also be closed while blinking which is not a sign of drowsiness. Tiredness in general can be detected from the facial expression as the facial expression tiredness differs from alertness or the non-vigilant state. But being tired is not a sure sign of drowsiness. One individual drowsiness expression therefore is not reliable enough to detect drowsiness. While many models in the literature like [16] [12] [14] [15] detect individual signs of drowsiness like yawning reliably, their approach is limited to just one sign of drowsiness.

For a more reliable way of detection drowsiness it is recommendable to consider as many different drowsiness symptoms as possible as described in 1.1.3 Our model will be able to detect three different drowsiness signs that it learns from the features of the three datasets that it will be trained on. With the output of our model, more informed decisions about the potential drowsiness of a driver can be made.

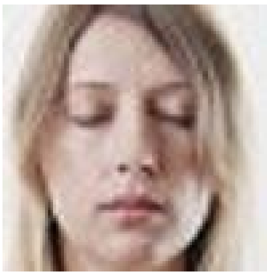
We call our model generalized because it is not limited to the detection of just one drowsiness sign rather it can detect a more generalized version of drowsiness.

2.2.1 Features

Each of our three datasets displays one unique sign of drowsiness as presented below. The dataset features complement each other since their features category is unique in the combined dataset. As to see below, the datasets show drowsiness features in different lighting, environment and humans. This variety in the datasets reinforce the fundamental concept of generalization, as the model will be trained to detect drowsiness features in various settings. Moreover, it is an important aspect to consider when automated drowsiness detection is used in real-life application.

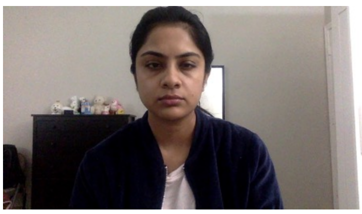
Eyes Dataset:
The eyes dataset shows faces with either closed or open eyes.

(a) Example of Open Eyes [18]



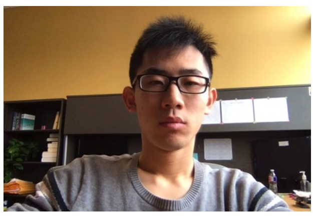
(b) Example of Closed Eyes [18]

Fatigue Dataset:
The fatigue dataset shows three possible types of wakefulness of humans: alert, non-vigilant and tired.



(a) Example of Tired [19]

(b) Example of Non-vigilant [19]



(c) Example of Alert [19]

Yawning Dataset:
The yawning dataset shows a person that does not yawn and a person that is in the middle of yawning.

(a) Example of No yawn [17] (b) Example of Yawn [17]

2.3 Deep Learning Model Architecture

The deep learning model architecture is a convolutional neural network based on residual nets (ResNets) as described in [20]. The decision of choosing a ResNet-based network is due to its efficiency for classification problems, especially in the biomedical problems [21, 22]. In addition, it solves the problem of vanishing gradient implicitly due to the skip connections. Hence, it enables us to train deeper neural network.

The ResNet architecture was also chosen for this project because it has proven to yield high-precision results for image classification, as a similar model was previously used to classify the FashionMNIST dataset [23]. The model structure used in our research is directly inspired by the convolutional neural network (CNN) model used for image classification from previous course work [23].

First, an illustration of the different types of ResNet blocks is as shown in figure 7. Each block consists of two convolutional layers with a 3x3 kernel and no bias term. The padding is just one pixel around the input, and at the end, there is a 2d batch normalization after each convolutional layer. A ReLU activation function is used after the first convolutional layer and at the end of the block.

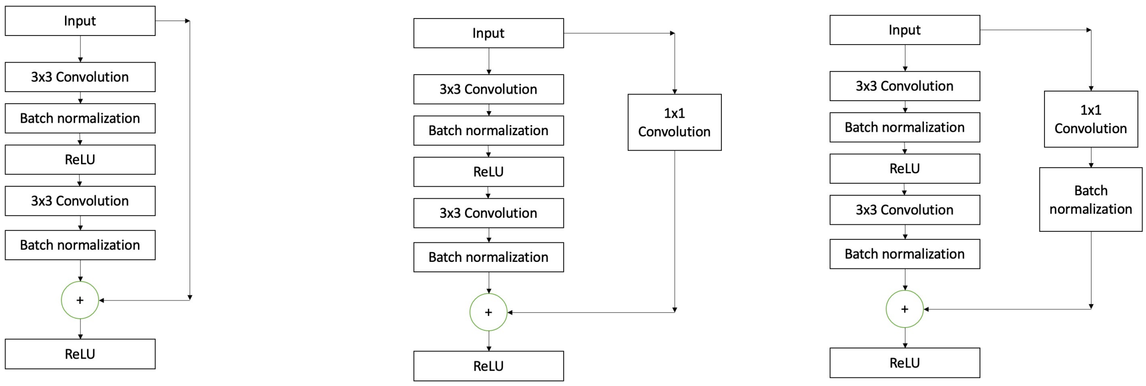


Figure 7: Different types of the skip connections in a ResNet block adapted from [24] and [23]

The unique property of the ResNet block is the skip connection, which just clones the input, in case the resolution and the channels are the same. However, if the number of channels is changed, but the resolution is fixed, the skip connection should have one 1x1 convolutional layer, without batch normalization (middle image of7). On the other hand, if the number of channels is fixed and the resolution changed, the skip connection will have a 1x1 convolutional layer with batch normalization, which is presented in the left image of figure 7. This is the basic structure of one ResNet block and we will use the first type of skip connection which just clones the input. We use varying number of sequential blocks to form one group.

The general model structure used in this experiment is illustrated in figure8 and consists of three ResNet-groups with varying numbers of blocks inside. Our residual nets architecture itself can be treated as a hyperparameter that is going to be tuned. The model is implemented with the PyTorch library, using the help of its pre-defined functions [25].

2.4 Training and Tuning

The training strategy was also adapted from the same model from previous course work [23] as it already has been proven effective for the FashionMNIST dataset.

The datasets are split into training and test data with the ratio of 9:1 and the training dataset is used only during the training. This strategy helps to detect overfitting and therefore increases the accuracy. As we are training with several datasets, the training is split in first, training of a single dataset, and second, training of the combined datasets.

In general, the training of the ResNet model is done in batches of 32 images and with the Adam optimizer. For the loss function, the Cross-Entropy Loss function is used. The number of epochs was defined by a mix of monitoring the training accuracy during the training and the limitations of our computation resources, and the number is currently cut off at 30 epochs for evaluation. However, the chosen best models are trained up to 50 epochs.

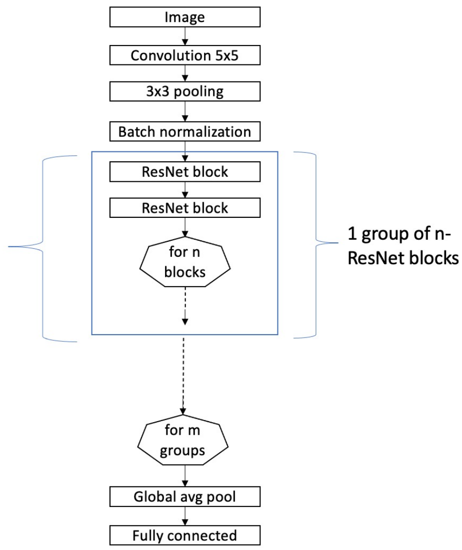


Figure 8: General model architecture of m groups, each group having n ResNet blocks inspired by [24][23]

The tuning of the model was done by using different learning rates and changing the numbers of blocks in each group of the ResNet architecture. The model was trained with the learning rate values of: 0.1, 0.01 and 0.001. The number of groups is going to be fixed for three groups only. However, the number of ResNet blocks in each group is going to vary between different combinations with a minimum of one block per group, and 4 blocks per group.

The proposed architectures for our model have been trained on each dataset separately, after that the generalized model is trained on all the datasets.

2.4.1 Single Dataset Model

To gain a deeper understanding of the datasets and our deep learning model, a single independent model was trained on each dataset separately using the architecture mentioned in section 2.3. This enabled us to judge the feasibility and potential accuracy precision that should and can be achieved with our model.

2.4.2 Generalized Model

In order to achieve generalizability for drowsiness detection, we combined the three datasets while keeping the respective classes. The number of training epochs is extended due to the increased data size.

2.5 Test Strategy

The tests for accuracy are done with the test dataset only. The split between training and test data ensures that the test accuracy is solely calculated from images that the model has not seen before. The trained model is used for predicting the image label, and the predicted label is compared to the real label. The test is done in batch size of 32 images and the results are combined.

3 Results and Analysis

The results are computed at the end of each training epoch on a separate validation set to gain more insight into our model behaviour. There are results for each independent model for each separate dataset, but these models were not tuned and were trained with learning rate of 0.01 and the architecture was fixed for three groups and each group has two ResNet blocks. These individual models were just preliminary steps to ensure that the proposed architecture is capable of classifying each dataset independently. Next, the results of the generalized model and its tuning is presented. Additionally, a detailed comparison and the outcomes are discussed at the end.

3.1 Accuracy

This section presents the result for each separate modol for each dataset as well as the evaluation of different architectures for the generralized model.

3.1.1 ResNet for Eyes Dataset

With regard to the Eyes Dataset, the model was trained to determine if the face. has a closing or opening eyes, the accuracy for inferring the labels on test data was 80% in the early epochs, and stabilized at the end at 91% at the end.

3.1.2 ResNet for Fatigue Dataset

For the fatigue dataset, the nature of this dataset with its three different labels makes it into a multi-classification problem. Although this problem should be more challenging, the accuracy for inferring the labels on the test data was 59% in the beginning, and stabilized at 97.7% at the end.

3.1.3 ResNet for Yawning Dataset

The accuracy for the yawning dataset stabilized at 83%, which is not a very good indicator. However, due to low computational resources, we trained it for only 5 minutes. In these sub-models, good behaviour during the training helped us to gain confidence to train the generalized full model.

3.1.4 ResNet for the Generalized Model

For the generalized model, a multi-classification problem based on the combined dataset, each test accuracy during the training epochs is shown in a plot. There is one plot for each learning rate value. The curve name is depicted with three numbers describes the model architecture used for the training. Each number indicates how many blocks are in each ResNet group. For example the indicator "[2,3,2]" specifies that in the first ResNet group there are two blocks, in the second group are three blocks and in the third group, there are two blocks again.

The ResNet models trained with learning rate = 0.1 reached an accuracy of around 80% after 30 epochs seen in figure 9 which is not very good in comparison the models trained with a different learning rate.

In comparison, the models trained with a learning rate of 0.01 and of 0.001 achieved accuracies in the 90-95% area as seen in figures 10 and 11. The best results were attained with the models depicted in figure 12.

The best model of an architecture with one block per group and learning rate of 0.01 achieved 94.7% accuracy after 30epochs of training.

This is a great result for the classification of being drowsy and also comparable with the latest research.

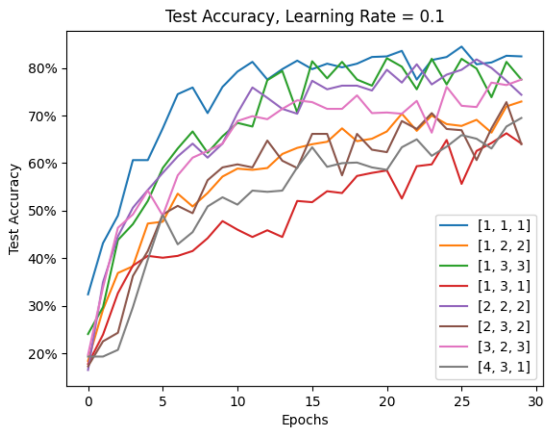


Figure 9: Test accuracy per training epoch with learning rate lr=0.1

Figure 10: Test accuracy per training epoch with learning rate lr=0.01

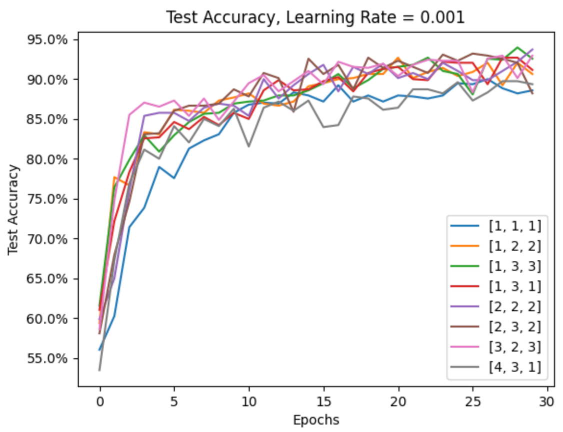


Figure 11: Test accuracy per training epoch with learning rate lr=0.001

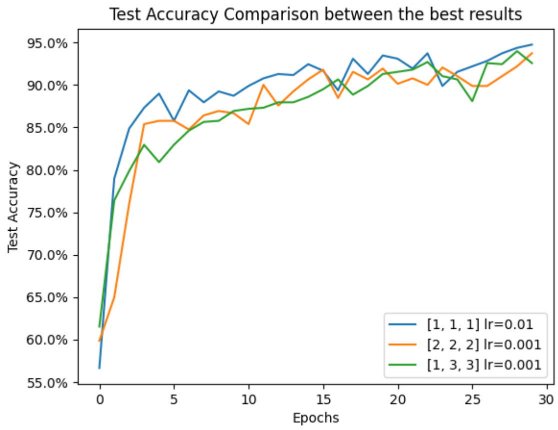


Figure 12: Test accuracy per training epoch for the three best training runs

3.2 Comparison

This research was primarily done to make improvements on the latest publications in drowsiness detection in the autonomous driving field. The improvements should not be only about the evaluation metric, but also about the objective of the model. That is why a proper comparison should be done to see what are the improvements as well as proposing future recommendations for the drowsiness detection.

	Rajkar et al. [16]	Fatigue detection [?]	Our generalized model
Model	3-layers CNN	ANN	ResNets
Datasets	2 Models of CewDD and YawnDD	FatigueDD	CewDD, YawnDD, and FatigueDD
Accuracy	Average accuracy of 96%	83.6%	94.7%

Table 2: Comparison among datasets and accuracy

These are the points of comparison with the models that utilized the same datasets. However, other models took other approaches, such as the model done by Nakamura, Maejima, and Morishima [26]. This model aims to detect the driver drowsiness by general facial expressions, and they achieved an accuracy of 82.8%. Although their model does not make a great improvement in accuracy, but their model was more generalized than the conventional drowsiness detection.

In addition, another model is targeting the generalization of drowsiness detection, which is the model implemented by Reddy et al. [27]. This model was trained on the DROZY dataset [28], which is a dataset of drowsy people using data provided from EEG sensors. Although this model can be categorized as drowsiness detection based on physiological signals, the model itself utilized convolutional neural network with a compression technique [27].It is also a good benchmark for comparison with our model, as it targets detecting drowsy and yawning as separate labels. Their accuracy reached 93.8%, and also they successfully detected yawning and drowsy in one single model.

4 Discussion

There are a vast majority of researches and publications tackling the problem of drowsiness detection in different ways. However, the proposed approach in this research is to build a single model using all drowsiness features. As a result, there are two aspects in the discussion, the accuracy and usability of the model.

Accuracy is a good measure of a deep learning model, when different models are doing the same job. However, in this case, a lot of drowsiness detection models are not doing the same job. Although the final goal is detecting drowsiness, each model achieves it differently. That is why accuracy cannot be a proper judgment for the efficiency of the model. The best accuracy achieved was the model done by RajKar et al. [16], but this accuracy was the average of two different models training on CewDD and YawDD. Although it detects drowsiness with accuracy of 96%, but it detects the drowsiness based on closing eyes and yawning only. As an example, it cannot detect if someone is drunk or sleep-deprived. Although this model achieved a slightly higher accuracy than our model, the capability of detecting diverse examples of drowsiness is needed, and its encapsulation in a single model is better in order to be implemented.

On the other hand, other models were more focused on generalizability as well, such as [27], their best accuracy was 93.8%. However, they did not choose specific activities of drowsiness, they trained their model on DROZY dataset [28] using a deep learning approach. They indeed succeeded in making a single model for drowsiness detection with a 93.8% accuracy. That is why this is also a good model to be considered as a benchmark for our evaluation.

The proposed model in this paper achieved an accuracy of 94.7%, which is a sufficient accuracy for drowsiness detection. With further training epochs, the accuracy is likely to improve above 95%. The model also succeeded in detecting different types of drowsiness in a single model, which is promising for future research and the one to be implemented. In addition, using the same approach, more datasets can be collected and trained in our model, hence increasing the generalizability in our model.

5 Conclusion

The main conclusion from this research is the objective of generalizing the drowsiness detection approach. As most of the researches put more focus into changing the architecture and getting better accuracy, this research shows that diversifying the concept of drowsiness is also as important as enhancing the accuracy. Moreover, the proposed collection of data and the proposed architecture successfully detects various types of drowsiness. The successful drowsiness types include, and not limited to, closing the eyes, yawning, and general fatigue. The general fatigue detection include various type of fatigue such as red eyes, drowsy faces, being drunk, and other features as discussed there. Also, further datasets should be added for enhancing the generalization. Last but not least, this research is an important milestone for a new direction in tackling the problem of drowsiness detection, which yields an accuracy of 94.7%, and also its implementation can be further improved.

References

[1] E. D. Sussman, H. Bishop, B. Madnick, R. Walter *et al.*, "Driver inattention and highway safety," 1985.

[2] J. A. Horne and L. A. Reyner, "Sleep related vehicle accidents," *Bmj*, vol. 310, no. 6979, pp. 565–567, 1995.

[3] S. Kaplan, M. A. Guvensan, A. G. Yavuz, and Y. Karalurt, "Driver behavior analysis for safe driving: A survey," *IEEE Transactions on Intelligent Transportation Systems*, vol. 16, no. 6, pp. 3017–3032, 2015.

[4] National Heart, Lung, and Blood Institute and others, "Drowsy driving and automobile crashes," 1998.

[5] A. Sahayadhas, K. Sundaraj, and M. Murugappan, "Detecting driver drowsiness based on sensors: a review," *Sensors*, vol. 12, no. 12, pp. 16 937–16 953, 2012.

[6] P. M. Forsman, B. J. Vila, R. A. Short, C. G. Mott, and H. P. Van Dongen, "Efficient driver drowsiness detection at moderate levels of drowsiness," *Accident Analysis & Prevention*, vol. 50, pp. 341–350, 2013.

[7] J. Lim and D. F. Dinges, "Sleep deprivation and vigilant attention." 2008.

[8] K. Fujiwara, E. Abe, K. Kamata, C. Nakayama, Y. Suzuki, T. Yamakawa, T. Hiraoka, M. Kano, Y. Sumi, F. Masuda *et al.*, "Heart rate variability-based driver drowsiness detection and its validation with eeg," *IEEE transactions on biomedical engineering*, vol. 66, no. 6, pp. 1769–1778, 2018.

[9] M. Patel, S. K. Lal, D. Kavanagh, and P. Rossiter, "Applying neural network analysis on heart rate variability data to assess driver fatigue," *Expert systems with Applications*, vol. 38, no. 6, pp. 7235–7242, 2011.

[10] A. Lemyre, F. Belzile, M. Landry, C. H. Bastien, and L. P. Beaudoin, "Pre-sleep cognitive activity in adults: A systematic review," *Sleep medicine reviews*, vol. 50, p. 101253, 2020.

[11] B.-G. Lee, B.-L. Lee, and W.-Y. Chung, "Smartwatch-based driver alertness monitoring with wearable motion and physiological sensor," in *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, 2015, pp. 6126–6129.

[12] M. Poursadeghiyan, A. Mazloumi, G. N. Saraji, M. M. Baneshi, A. Khammar, and M. H. Ebrahimi, "Using image processing in the proposed drowsiness detection system design," *Iranian journal of public health*, vol. 47, no. 9, p. 1371, 2018.

[13] P. Viola and M. J. Jones, "Robust real-time face detection," *International journal of computer vision*, vol. 57, no. 2, pp. 137–154, 2004.

[14] S. Abtahi, B. Hariri, and S. Shirmohammadi, "Driver drowsiness monitoring based on yawning detection," in *2011 IEEE International Instrumentation and Measurement Technology Conference*. IEEE, 2011, pp. 1–4.

[15] W. Zhang, Y. L. Murphey, T. Wang, and Q. Xu, "Driver yawning detection based on deep convolutional neural learning and robust nose tracking," in *2015 International Joint Conference on Neural Networks (IJCNN)*. IEEE, 2015, pp. 1–8.

[16] A. Rajkar, N. Kulkarni, and A. Raut, "Driver drowsiness detection using deep learning," in *Applied Information Processing Systems*. Springer, 2022, pp. 73–82.

[17] S. Abtahi, M. Omidyeganeh, S. Shirmohammadi, and B. Hariri, “Yawdd: A yawning detection dataset,” in *Proceedings of the 5th ACM Multimedia Systems Conference*, ser. MMSys ’14. New York, NY, USA: Association for Computing Machinery, 2014. doi: 10.1145/2557642.2563678. ISBN 9781450327053 p. 24–28. [Online]. Available: https://doi.org/10.1145/2557642.2563678

[18] F. Song, X. Tan, X. Liu, and S. Chen, “Eyes closeness detection from still images with multi-scale histograms of principal oriented gradients,” *Pattern Recognition*, vol. 47, no. 9, p. 2825–2838, 2014. doi: 10.1016/j.patcog.2014.03.024

[19] Timj, “Fatigue detection,” https://www.kaggle.com/timmjy/fatigue-detection, 8 2020, kaggle dataset.

[20] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” 2015.

[21] S. Murali, R. Deepu, R. Shivamurthy *et al.*, “Resnet-50 vs vgg-19 vs training from scratch: A comparative analysis of the segmentation and classification of pneumonia from chest x-ray images,” *Global Transitions Proceedings*, 2021.

[22] D. Sarwinda, R. H. Paradisa, A. Bustamam, and P. Anggia, “Deep learning in image classification using residual network (resnet) variants for detection of colorectal cancer,” *Procedia Computer Science*, vol. 179, pp. 423–431, 2021.

[23] A. Ilin, “Mycourses: Course: Cs-e4890 - deep learning d, 02.03.2021-26.05.2021.” [Online]. Available: https://mycourses.aalto.fi/course/view.php?id=28212

[24] A. Zhang, Z. C. Lipton, M. Li, and A. J. Smola, “Dive into deep learning,” *arXiv preprint arXiv:2106.11342*, 2021.

[25] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Kopf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai, and S. Chintala, “Pytorch: An imperative style, high-performance deep learning library,” in *Advances in Neural Information Processing Systems 32*. Curran Associates, Inc., 2019, pp. 8024–8035. [Online]. Available: http://papers.neurips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-library.pdf

[26] T. Nakamura, A. Maejima, and S. Morishima, “Driver drowsiness estimation from facial expression features computer vision feature investigation using a cg model,” in *2014 International Conference on Computer Vision Theory and Applications (VISAPP)*, vol. 2. IEEE, 2014, pp. 207–214.

[27] B. Reddy, Y.-H. Kim, S. Yun, C. Seo, and J. Jang, “Real-time driver drowsiness detection for embedded system using model compression of deep neural networks,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2017, pp. 121–128.

[28] Q. Massoz, T. Langohr, C. François, and J. G. Verly, “The ulg multimodality drowsiness database (called drozy) and examples of use,” in *2016 IEEE Winter Conference on Applications of Computer Vision (WACV)*. IEEE, 2016, pp. 1–7.