



Yawn Based Driver Fatigue Level Prediction

Haider A. Kassem¹, Morshed U. Chowdhury¹, Jemal Abawajy¹, Senior member IEEE, and Ahmed Raad Al-sudani¹

¹School of Information Technology, Deakin University Geelong, Australia

¹Email: {hakas; morshed.chowdhury; Jemal.abawajy; aralsuda}@deakin.edu.au

Abstract: The fatigue-related accident is increasing due to long work hours, medical reasons, and age that decrease response time in a moment of hazard. One of drowsiness and fatigue visual indicators is excessive yawning. In this paper, a non-optical sensor presented as a car dashcam that is used to record driving scenarios and imitates real-life driving situations such as being distracted or talking to a passenger next to the driver. We built a deep CNN model as the classifier to classify each frame as a yawning or non- yawning driver. We can classify the drivers' fatigue into three levels, alert, early fatigue and fatigue based on the judgement of the number of yawns. Alert level means when the driver is not yawning, while, early fatigue is when the driver yawns once in a minute. Fatigued is when the driver yawns more than once in a minute. An overall decision is made by analyzing the source score and the condition of the driver's fatigue state. The robustness of the proposed method was tested under various illumination contexts and a variety of head motion modes. Experiments are conducted using YAWDD dataset that contains 322 subjects to show that our model presents a promising framework to accurately detect drowsiness level in a less complex way.

Keyword: Drivers Fatigue, Driver Yawning, Fatigue prediction , machine learning fatigue prediction.

1. Introduction

The number of traffic accidents is increasing due to a diminished vigilance level of the driver has become a serious problem for our society. Accidents related to driver hypo-vigilance are more serious than other types of accidents since sleep deprived drivers often do not take evasive action before a collision [1], [2]. Monitoring the driver's level of vigilance and alerting them when he/she is not paying adequate attention while driving is essential to prevent driving accidents. According to the U.S. National Highway Traffic Safety Administration, for a 5-year period, 100,000 motor vehicle collisions per year (1.6% of 6.3 million) were reported by the police and drowsiness identified as the leading cause [3].

Data from the UK Department of Transport show that fatigue contributed to 20% of the total number of road accidents [4]. Therefore, prevention of accidents is a Primary focus of effort in the field of active road safety research.

In this paper, we are proposing a yawning based driver fatigue prediction model, yawning is considered an early symptom of fatigue [5]. The person is fatigued when frequent yawning occurs [6], Therefore detecting the accurate yawn is essential for predicting drivers fatigue level or early drivers fatigue detection. Which gives enough time for the vehicle operators to respond in driving crisis.

In our opinion every vehicle needs fatigue prediction systems for drivers to avoid collision/accident. But this system still in the development stages due to the need of attaching sensors to the drivers which is invasive and would be treated as an obstacle to a sensitive process such as driving. Furthermore, the fatigue prediction has been applied in modern cars require expensive sensors to be attached to the car such as pressure sensors on the throttle accelerator and the steering wheel.

In this study, we are presenting an inexpensive model that can detect drivers yawn with high accuracy using visual data from a camera, which can be applied to any vehicle.

We have used a dataset of males and females' drivers who yawned while behaving naturally such as talking to someone sitting next to them and cover their mouth during yawn that gives extra challenge to detect the yawns.

Our main contribution in this work is drivers fatigue level prediction model based on yawn detection. We have used different training strategies to obtain maximum results. This system can help minimizing the danger of fatigue related accidents apart from the driving fatigue accidents or time management such as: to check office/chemical lab employees fatigue state while working in front of a computer. Where fatigue can play a big role in their work achievements.

Our system can help understand the time and the required efforts. By analysing the results of our model, we can mark the number of employees that goes into fatigue state while working on a specific project or the time that takes employees to get fatigued in a certain environment which can help managing the working hours and taking important measures to get the best of the employees time.

The fatigue levels we use in this paper are three, (Level 1) is the alert state, and this is considered when we do not detect any yawn. (Level 2) is the early fatigue level, and it happened when a driver starts yawning one in minute time. (Level 3) is the fatigued level, and it occurs when the driver starts to yawn frequently (more than one yawn at one-minute time). We use a CNN model which consumes the time

and effort compared to other methods that use manual feature extraction. The model's performance is measured using Precision, Recall and F1 score metrics.

The rest of the paper will be structured as follows: section 2 illustrates the related work to our research, section 3 illustrates our outlined methodology, Section 4 discussing our experiment and results, while Section 5 draw the conclusion.

2. Related work

The phenomena of yawning are commonly associated with tiredness and fatigue [7]. It is signified by the unintentional opening of the mouth. Much work has gone into devising a system that can detect a natural yawn and these systems generally work by calculating the area of mouth opening and lip parameter. Measurement of mouth opening using the height to width formula was used in the research of [8]. Moreover, this method is not uncommon as it was also adopted during the research of [9]. In this work, the yawn is determined when the ratio is above 0.5 in 20 capture frames by employing the use of a 30 fps video capture camera. S.Abati et.al [10] used a slightly different method as they depended on the identification of the isolated mouth area, then they used conventional face detection to validate the determination of the yawn. In [11], authors used the color analysis method in conjunction with connected component analysis to gather visual pixel data from the lip area. Then they used a geometric calculation technique to monitor the changing shape of the mouth to determine and distinguish different mouth activities such as yawning and talking. Y.Lue et.al [12] was used face detection by employing two images as points of reference. Then they pinpointed the area between the nostrils using an integral projection technique. Yawn state is determined here by measuring the midpoint between the nose and the center of the chin. Given that geometry was the primary form of analysis here, the algorithms can be hard to distinguish between various mouth activities. The geometric approach has many weaknesses; one of the major setbacks of this method is that it will depend heavily on the geometric makeup of the face of the subject. Another weakness of this method is that it focuses heavily on the shape of the mouth, this is weak because readings will vary based on lighting, facial orientation; head size, facial hair, and sharp lip movement. Other major issue with this system is that if a subject lifts their hand to cover the yawn, this action will hinder the ability of the system to correctly detect the yawn [13]. A gravity center template use is a new and innovative method to detect the face, this method was proposed by [14], and the yawn was to be detected through the use of linear discriminant analysis. Authors T. Azim et.al. [15] demonstrated a technique where the yawn determination style employed Viola-Jones face detection for pinpointing the facial region; isolated the mouth opening, and then found

the lips. M.M. Ibrahim et.al [16] in their work had put forward a proposition to detect yawning using open mouth, covering of mouth region and face distortion. They proposed local binary patterns (LBP) to address the mouth covering issue.

These physical characteristics-based algorithms are complex; hence, these algorithms may not be realistically applicable in a real-time situation. One of the main reasons of impracticality is that these algorithms requires large amounts of data, robust processing systems, and sophisticated internal Real-time captured camera.

3. Approach

We formulate this problem as a binary classification, where the task is to classify each video frame as either: yawning or non-yawning.

We build a deep CNN model as the classifier. This CNN model follows the Resnet-50 architecture, having 50 layers with residual units. This model is among state-of-the-art models for visual recognition tasks in the literature.

$$S = \begin{cases} 1, & \text{when yawning} \\ 0, & \text{otherwise} \end{cases} \dots \dots eq(1)$$

Where S is the input frame that has been constructed from the input video.

$$\sum_{s=1}^n D(Y) \dots \dots \dots eq (2)$$

Where n is max no of yawning that determine the person has fatigue state, $D(Y)$ function that determine the person is in yawning state.

The Figure 1 illustrates the structure of the model. Note that this figure depicts a Resnet model with 34 layers, instead of the 50-layer variant that we used; however, it gives an idea of the structure of the model.

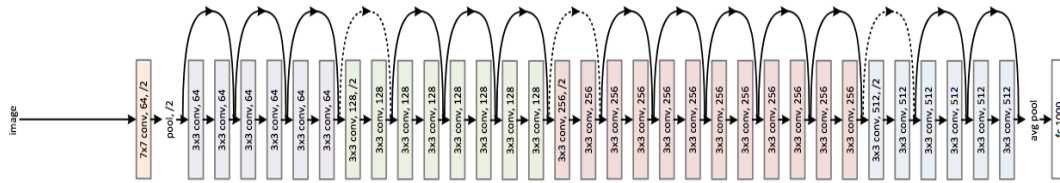


Figure 1: The Resnet deep architecture.

The basic difference between the Resnet architecture and other CNN architectures (e.g., AlexNet, VGGNet, GoogLeNet etc.) is that instead of having a block of hidden layers to implement a non-linear function $H(x)$ of an input x ; needs to be fit, we use that block of layers to implement a non-linear residual function $F(x)$ such that: $H(x) = F(x) + x$.

The Resnet architecture has been proved to perform very well in various computer vision problems, as we will observe in our work compared to previous works.

3.1 Training the model

- We initialize the model’s parameters with the parameters from a model trained on the ImageNet dataset.
- The training is done by using the Stochastic gradient descent algorithm to optimize the cross-entropy loss function:

$$L = \frac{1}{n} \sum_{i=1}^n -(y_i \log(p_i) + (1 - y_i) \log(1 - p_i)) \dots\dots\dots \text{eq (3)}$$

with n the number of training samples, y_i the label of the i -th training sample ($y_i = 1$ if the sample has label ‘yawning’ and $y_i = 0$ if the sample has label ‘non-yawning’), p_i is the predicted probability that the sample has label ‘yawning’.

3.2 Testing the model

At the testing stage, 10 random crops of a test image are generated, and the final prediction result is recorded by performing majority voting from the 10 predictions for those crops.

4. Experiments

In this section, we describe our training and testing data, and then, we explain our experimental setup and experimental results.

We use the YawDD dataset [17] for the initial experiments. As the dataset does not include yawning labels, Therefore, we manually labelled it using the VOTT annotation tool [18].

After this annotation step, we now have a labelled dataset from 16 videos in which each frame is associated with a label, with the characteristics in Table 1:

Table 1: The number of total training/testing data that includes yawning and non-yawning labels.

	Images	Yawning	Non-yawning
Training	754	38	716
Test	252	18	234

It may be noted that the collected dataset is highly imbalanced.

4.1 Experimental Results

As the dataset is highly imbalanced, therefore model has been trained using all the training images and the results is in highly biased. The model only predicts non-yawning. We use two approaches: (a) multiplying the losses associated with yawning images by a large factor to influence the model to emphasis more on those cases, and (b) sub-sampling the training dataset to get a smaller but balanced training set.

- In approach (a), the loss function is modified according to eq (2)

$$L = \frac{1}{n} \sum_{i=1}^n -\alpha(y_i \log(p_i) + (1 - y_i) \log(1 - p_i)) \dots\dots\dots \text{eq (4)}$$

Where α set to 10. In other words, a yawning image contributes are 10 times more to the loss function than a non-yawning image. In approach (b), after the sub-sampling step, we completed the training sets

having 38 yawning and 38 non-yawning images. To train our propose model with this small training set, we use data augmentation and heavy regularization (dropout and weight decay) to deal with overfitting.

4.2. Yawning detection results

We evaluate the proposed model using the YawDD datasets [17]. Our aim is to investigate the performance of the proposed model's architectures and quantifying the yawning classification by implementing Resnet architecture in CNN.

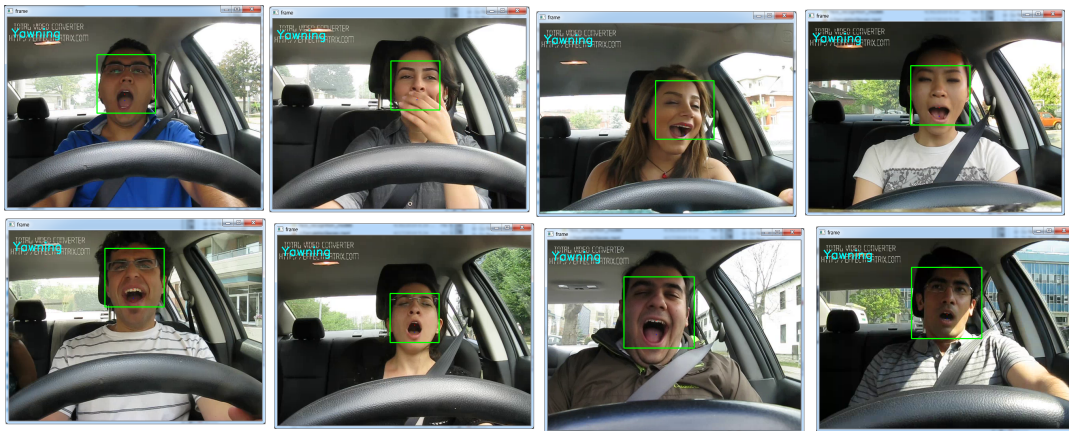


Figure 2: A sample of our yawning detection results on YawDD dataset.

4.2.1 Dataset

YawDD dataset contains 322 males and females' drivers' videos with different facial characteristics that mainly used for models and algorithms and specifically yawning detection area. Different ethnicities are presented, and each person has three/four video clips with different mouth conditions such as normal, talking/singing, and yawning, as well as drivers wear glasses /sunglasses.

Each driver has a single video clip containing scenes with driving, driving while talking, and driving while yawning.

The experimental results are summarized in the Table 2. The model's performance is measured using three metrics, such as precision, recall and F1 score.

Table 2: illustrates the various types of training strategies and how different matrices results when yawning or not yawning.

Training strategy	Yawning			Non-yawning		
	Precision	Recall	F1 score	Precision	Recall	F1 score
Normal training	0	0	0	0.93	1.0	0.96
Strategy (a)	0.2	0.22	0.21	0.94	0.93	0.93
Strategy (b)	0.39	0.78	0.52	0.98	0.91	0.94

By analysing the results, the model is highly biased toward non-yawning when training in a normal way with the imbalanced training set. Adding high weights to yawning samples during training (strategy (a)) improves the results, making the prediction more balanced. Training the model using the strategy (b) gives the best performance overall, with more balanced predictions compared to those produced using strategy (a).

Table 3 shows a significant performance improvement of our method that reaches a rate an average of accuracy equal to 96.2 %. According to the exposed results, we note some failures. These failures are primarily due to Lack of illumination in which the detected edges are not clear enough. In addition, our model some time cannot identify the difference between a long-time open mouth and a yawn. But compared to [19], [20], [21] and [22], it shows that these errors happened because of some drivers unconsciously hide their mouth by putting the hand in the mouth. This suggests that Resnet is good at learning temporal relations on high level features. It also supports the idea that incorporating information across video sequences will enable better prediction.

Table 3: Yawning Detection Accuracy rate of different methods including ours.

Reference	Accuracy	Dataset
[19]	92%	YawDD
[20]	75%	YawDD
[21]	88.6%	YawDD

[5]	99.80 %	Self-collect data
[22]	94.63%	YawDD
Our method	96.2	YawDD

According to literature, excessive yawning is an indicator that the person is fatigued, and while an alert person would experience no yawning during the trip. The driver may yawn once in minute time and still in control of the vehicle. As seen in Figure 3 our proposed framework contains various steps that uses yawns as a main fatigue indicator. The input of our framework is a constricted image from a video. The second stage (the binary yawn detector), in this step the frame is being classified into yawn to continue to the next step or non-yawn, to navigate back to the previous step. The third step is to count the yawns in one-minute time to output the level of: Alert, when the driver does not yawn and then goes back to step one. Early fatigue, when the driver yawns once. Fatigued, when the driver yawns more than once.

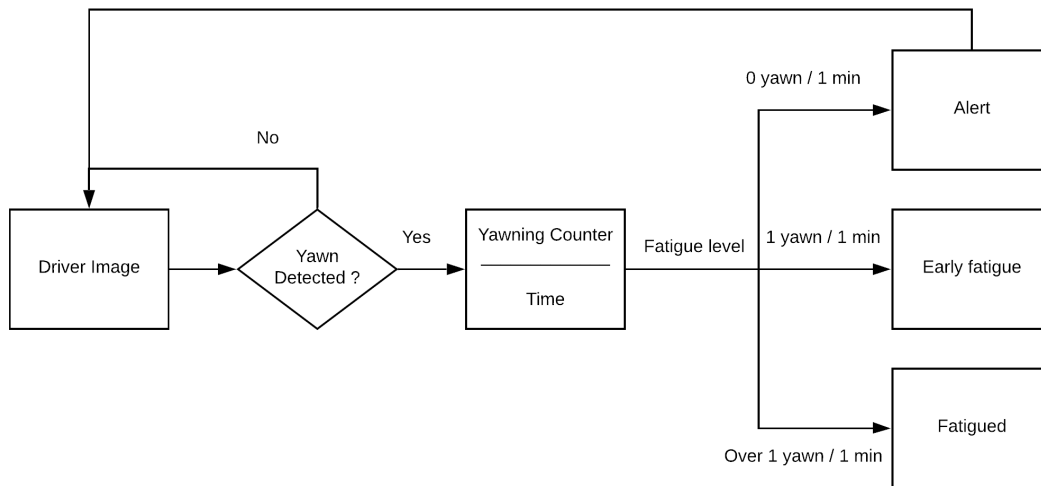


Figure 3: an overview of our proposed yawn-based fatigue level prediction framework.

5. CONCLUSION

We presented a yawning based fatigue prediction method that monitors driver drowsiness levels and is also capable of aggregating frame-level CNN outputs into different fatigue levels. Comparative to the efforts of previous researchers that relied on manual features such as hog and contours that triangulate mouth and nose, our most advanced approach involves the use of video which has the capability to adapt and can learn to store, analyse and integrate different information over prolonged periods of time to enhance and produce ideal performance. The outcome of this work indicates the accuracy of the hypothesis that detection of yawning is ought to be conducted over prolonged sequential video capturing rather than single imagery. It is also just to include within using this method which involves the video capturing of space-temporal elements can have some drawbacks in the case of low levels of illumination. In the work we will conduct in the future within this field of research we have noted to include different approaches to produce different samples, such as exploring the effect of Optical Flow of information which are computed over adjacent frames.

References

- [1] L. M. Bergasa, J. Nuevo, M. A. Sotelo, R. Barea, and M. E. Lopez, "Real-time system for monitoring driver vigilance," *IEEE Transactions on Intelligent Transportation Systems*, vol. 7, no. 1, pp. 63-77, 2006.
- [2] V. B. Hemadri and U. P. Kulkarni, "Detection of drowsiness using fusion of yawning and eyelid movements," in *Advances in computing, communication, and control*: Springer, 2013, pp. 583-594.
- [3] P. S. Rau, "Drowsy driver detection and warning system for commercial vehicle drivers: field operational test design, data analyses, and progress," in *19th International Conference on Enhanced Safety of Vehicles*, 2005: Citeseer, pp. 6-9.
- [4] P. Jackson, C. Hilditch, A. Holmes, N. Reed, N. Merat, and L. Smith, "Fatigue and road safety: a critical analysis of recent evidence," 2011.
- [5] W. Tipprasert, T. Charoenpong, C. Chianrabutra, and C. Sukjamsri, "A Method of Driver's Eyes Closure and Yawning Detection for Drowsiness Analysis by Infrared Camera," in *2019 First International Symposium on Instrumentation, Control, Artificial Intelligence, and Robotics (ICA-SYMP)*, 2019: IEEE, pp. 61-64.
- [6] B. G. Pratama, I. Ardiyanto, and T. B. Adj, "A review on driver drowsiness based on image, bio-signal, and driver behavior," in *2017 3rd International Conference on Science and Technology-Computer (ICST)*, 2017: IEEE, pp. 70-75.
- [7] Z. Jie, M. Mahmoud, Q. Stafford-Fraser, P. Robinson, E. Dias, and L. Skrypchuk, "Analysis of yawning behaviour in spontaneous expressions of drowsy drivers," in *Automatic Face & Gesture Recognition (FG 2018)*, 2018 13th IEEE International Conference on, 2018: IEEE, pp. 571-576.
- [8] Q. Ji, Z. Zhu, and P. Lan, "Real-time nonintrusive monitoring and prediction of driver fatigue," *IEEE transactions on vehicular technology*, vol. 53, no. 4, pp. 1052-1068, 2004.

- [9] T. Wang and P. Shi, "Yawning detection for determining driver drowsiness," in *VLSI Design and Video Technology, 2005. Proceedings of 2005 IEEE International Workshop on*, 2005: IEEE, pp. 373-376.
- [10] S. Abtahi, B. Hariri, and S. Shirmohammadi, "Driver drowsiness monitoring based on yawning detection," in *Instrumentation and Measurement Technology Conference (I2MTC), 2011 IEEE, 2011: IEEE*, pp. 1-4.
- [11] W. Rongben, G. Lie, T. Bingliang, and J. Lisheng, "Monitoring mouth movement for driver fatigue or distraction with one camera," in *Intelligent Transportation Systems, 2004. Proceedings. The 7th International IEEE Conference on*, 2004: IEEE, pp. 314-319.
- [12] Y. Lu and Z. Wang, "Detecting driver yawning in successive images," in *Bioinformatics and Biomedical Engineering, 2007. ICBBE 2007. The 1st International Conference on*, 2007: IEEE, pp. 581-583.
- [13] M. Saradadevi and P. Bajaj, "Driver fatigue detection using mouth and yawning analysis," *International journal of Computer science and network security*, vol. 8, no. 6, pp. 183-188, 2008.
- [14] X. Fan, B.-C. Yin, and Y.-F. Sun, "Yawning detection for monitoring driver fatigue," in *Machine Learning and Cybernetics, 2007 International Conference on*, 2007, vol. 2: IEEE, pp. 664-668.
- [15] T. Azim, M. A. Jaffar, and A. M. Mirza, "Automatic fatigue detection of drivers through pupil detection and yawning analysis," in *Innovative Computing, Information and Control (ICICIC), 2009 Fourth International Conference on*, 2009: IEEE, pp. 441-445.
- [16] M. M. Ibrahim, J. J. Soraghan, L. Petropoulakis, and G. Di Caterina, "Yawn analysis with mouth occlusion detection," *Biomedical Signal Processing and Control*, vol. 18, pp. 360-369, 2015.
- [17] S. Abtahi, M. Omidyeganeh, S. Shirmohammadi, and B. Hariri, "YawDD: A yawning detection dataset," in *Proceedings of the 5th ACM Multimedia Systems Conference*, 2014: ACM, pp. 24-28.
- [18] Microsoft. (2019). VoTT [Online]. Available: <https://github.com/Microsoft/VoTT>.
- [19] 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)*, 2015: IEEE, pp. 1-8.
- [20] M. Omidyeganeh et al., "Yawning detection using embedded smart cameras," *IEEE Transactions on Instrumentation and Measurement*, vol. 65, no. 3, pp. 570-582, 2016.
- [21] W. Zhang and J. Su, "Driver yawning detection based on long short term memory networks," in *2017 IEEE Symposium Series on Computational Intelligence (SSCI)*, 2017: IEEE, pp. 1-5.
- [22] Z. Jie, M. Mahmoud, Q. Stafford-Fraser, P. Robinson, E. Dias, and L. Skrypchuk, "Analysis of yawning behaviour in spontaneous expressions of drowsy drivers," in *2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018)*, 2018: IEEE, pp. 571-576.