Drowsy State Detection System

Dendup Chophel[1], Manoj Chhetri[2]

Information Technology Department
College of Science and Technology: Rinchending, Chhukha, Bhutan
E-mail: [1]dendup1998@gmail.com, [2]manoj_chhetri.cst@rub.edu.bt*

Abstract

Drowsiness is defined as feeling abnormally sleepy during day time and may seem trivial but has been attributed to be one of the underlying causes of both small and big accidents. Making it critical and necessary for a drowsiness detection system that is both realistic, easy to deploy and effective in classifying the signs of drowsiness. This paper presents the drowsy state detection system developed using deep learning methods and the first ever Bhutanese Drowsiness Dataset containing 4000 images divided into two class and collected from different Bhutanese participants. The proposed CNN model achieved 99.9% training and 93.25% testing accuracy with a minimal loss and were evaluated using confusion matrix and F1-score.

Keywords: Drowsiness detection, Bhutanese Drowsiness dataset, Convolutional Neural Network, Deep Learning, Transfer Learning, Image preprocessing.

1. Introduction

Drowsiness is the feeling of being unusually exhausted or lethargic during the day time. It is the state between being awake and asleep (Blake, 2019). In the state of drowsiness, a person can cause and get into problems which otherwise can be easily avoided and in the worst-case scenario it may lead to fatal accidents as in the state of drowsiness a person's senses become dull, attentiveness decreases, reaction time becomes slower and information processing becomes less efficient.

Drowsiness is the state in which a person needs sleep but in this era of technologies and innovations, many people, especially younger generations, do not understand nor place importance on the need for a healthy rest. Many young people play games and watch movies for an extended period of time especially in the night when they are supposed to be sleeping and taking rest. There is a need for reminding oneself about taking rest from all these. Not having enough sleep at night leads to feeling sleepy or tired during the next day time which can hinder academic success in school and professional success at work, meaningful social interactions are undermined, personal relationships are put under stress and strain, and expose workers and drivers to severe dangers (Suni, 2020). Human beings need sleep as it's an inherent nature.

The solution is quite simple and that is to have a reminder that can remind oneself to sleep when drowsy at night so that it does not hamper the daytime state. People are bound to feel sleepy but many ignore the signs. A system like parental control is required to remind and limit one's indulgence in games, movies or excessive work and this is where a drowsy detection system can come in handy hence making it critical and necessary to design and develop real-world systems that can effectively detect drowsiness and is both realistic and convenient to use and deploy (Ghoddoosian, Galib, & Athitsos, 2019).

Many studies have been conducted related to drowsiness detection system using various methods such as deep learning, hardware based, behavioral based in many parts of the work except Bhutan. The key goal of this research is to employ methods of machine learning and deep learning to develop a drowsiness detection system to effectively recognize drowsiness in Bhutanese masses and it also presents the custom Bhutanese Drowsiness dataset developed from videos and images extracted from many different young Bhutanese individuals.

This paper's content is structured as follows: the first section provides some context information, followed by second section discussing related works and stating relevant literature reviews. Third section presents the methodology followed for this study, followed by the fourth section stating the findings from the experiments. Final section presents the final conclusion and possible future works.

2. Related work

For several years and in different parts of the world, the subject of drowsiness detection has been researched and investigated but it is mostly focused on driver drowsiness detection system.

Choudhary, et al. mentions and presents a study on different techniques used in the drowsiness detection and compares the advantages and disadvantages of all the techniques to address the problem caused by drowsiness in drivers as it was known that driving efficiency decreases with elevated drowsiness, culminating in injuries accounting for more than 20% of all automobile accidents. They have studied and research on Image Processing based techniques, Artificial Neural Network (ANN) based methods, vehicle/automobile structure-based methods and Electroencephalograph (EEG) techniques. From the comparison done showed that image processing-based techniques were most appropriate and non-intrusive method.

Poursadeghiyan, et al. describes and presents how image processing is applied in the drowsiness detection system. As fatigue and drowsiness leads to some apparent signs on face, and amongst various facial features, eyes are relatively of more importance. There are several methods and techniques employed in detecting drowsiness and from all the numerous techniques, image processing methods were found to be simple, fast and more precise. Hence, they employed this method to study the varying level of driver drowsiness in a simulated driving environment. The percentage of accuracy of detecting system was found to be 93% which was very high and the system was efficient.

Khan, et al. aimed to evaluate specific activities of the driver to determine the drowsiness level. They stated that most of the accidents occur because of the drowsiness and it is something that can be brought under control and addressed. The Drowsiness detection system that they proposed focused on the behavioral factors of the person driving the car which included eye closure, eye blinking, yawning and also the head pose. Applying image processing techniques, they developed a small model that achieved 81% accuracy.

Ghoddoosian, et al. developed the publicly available UTA-RLDD dataset and trained a baseline temporal model using the network known as Hierarchical Multiscale Long Short-Term Memory (HM-LSTM) on the developed dataset. They mainly focused on the relationship between varying levels of drowsiness and the blink rate of the person under the study achieving an accuracy of 65.2% that was higher than the human judgement benchmark of 57.8%. They could effectively detect multistage drowsiness and their custom dataset was significantly larger than existing datasets and their temporal end-to-end model had low computational and storage demands.

Park et al. designed and devised a new model based on three deep neural networks, AlexNet, VGG-FaceNet and FlowImageNet and achieved 73.06% accuracy on the benchmark NTHU drowsiness dataset. They were able to classify the images extracted from the videos in the dataset into four different classes: yawning, alert, drowsy with eye blinking and nodding of head. Jabbar et al. applied the traditional method of MLP (Multilayer Perceptron) classifier and trained it on NTHU dataset and achieved 81% accuracy on the testing portion of the dataset.

Dwivedi et al. adapted and trained three-layered convolutional neural network on a private custom dataset of 30 participants and were able to achieve 78% detection accuracy of drowsy drivers. Vanjani & Varyani employed a CNN and LSTM based deep learning network on a custom dataset that had videos of eight subjects simulating drowsiness and obtained 88.5% accuracy score. Hashemi et al. used the concept of transfer learning and pretrained CNN models; TL-VGG19, TL-VGG16, FD-NN on ZJU eye blink dataset achieving 94.96%, 95.45%, 98.15% drowsiness detection accuracy respectively.

3. Methodology

This section presents the methodology undertaken for designing and implementing the supervised learning drowsy state detection system.

3.1 System Overview

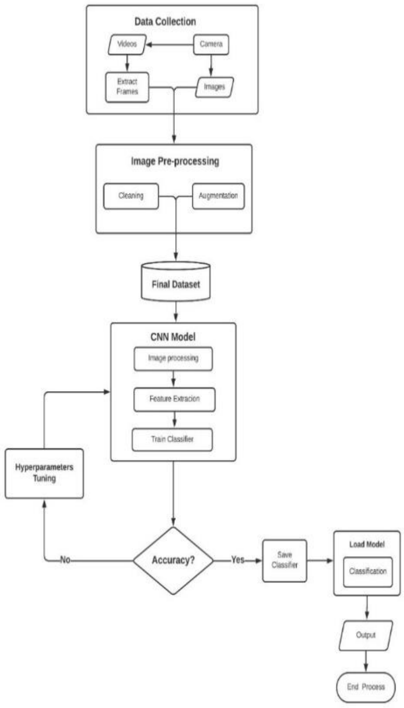

Figure 1. System overview.

Figure 1 presents the overview of the proposed system. The system has four main stages and multiple sub stages. The main stages are: data collection or acquisition, image pre-processing, CNN model, and classification. Videos and images were collected from participants which were pre-processed and feed into the CNN model that extracted the relevant features and trained the classifier. The finalized model is saved after achieving the desired accuracy and loaded in to make prediction on test set.

3.2 Dataset Collection

For every machine and deep learning project, the most important asset is the dataset so the deep learning model developed here is trained and tested on the Bhutanese Drowsiness dataset.

The Bhutanese Drowsiness dataset is the custom dataset developed specifically for this particular study and it has 4000 images of Bhutanese people labelled into two states of alert and drowsy. The videos and images for the dataset were collected from 50 young Bhutanese people. The raw data mostly consisted of videos recorded into two states of awake and drowsy, only some were images.

As video data cannot be used as a training data for neural network, frames were extracted from the videos that were of 30 frames per second at a gap of 10 frames using OpenCV. From the extracted frames, duplicates were removed and the dataset had 200 raw images which were reproduced to make the final dataset of 4000 images, divided equally between the two class.

3.3 Image Preprocessing

This is done mainly to prepare the raw images acquired into a standardized format to be fed into the neural network. The dataset was split before preprocessing. The training, validation and testing split were done manually and the train set had 70%, validation had 20% and test set had 10% of the total images in the dataset.

Then the images in each class and set were preprocessed using the augmentation techniques. Data augmentation is done mainly to get different variations of a single image which helps in solving the problem of overfitting. The image processing applied were adding saturation, brightness manipulation, flipping, rotation, blurring effect, contrast change and grayscale conversion as shown in Figure2. For every single raw image, the augmentation process generated 20 variations resulting in a total of 4000 images in Bhutanese Drowsiness dataset.

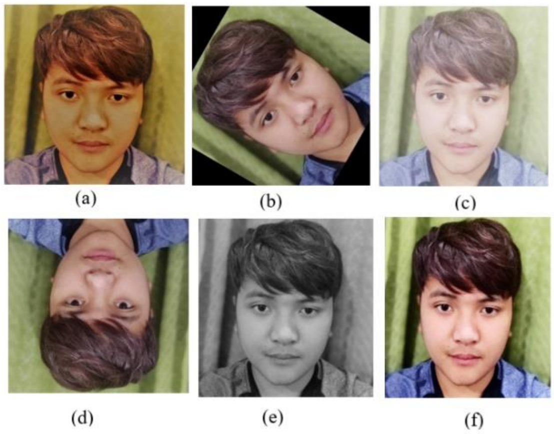

Figure 2. Samples of augmented images: (a) saturated, (b) rotated image, (c) high brightness image, (d) flipped image, (e) grayscale image, (f) high contrast image**.**

3.4 CNN Model

Before the CNN could extract the features and train the classifier, the dataset needs to loaded into the neural network. This was done using the ImageDataGenerator class and flow from the directory iterator. During the loading of the data, additional image processing was applied such as zooming, shearing and shift. The images were normalized (1/255) before feeding it into the network in order to make the number of parameters less so that the network requires less computation power and have reduced training time. This process is known as normalization.

As images of varying sizes cannot be fed into the CNN as it requires the images to be in standard size, preferably squared. The images in the dataset were of different sizes hence all the images were rescaled into 244 x 244 x 3 pixels when it was loaded into the network. The batch size was set to 32 because it was considered a good default value and gave optimum result (Brownlee, 2017).

Following the VGGNet general architectural principle, the CNN model was developed. This network was used to extract the features from the input images and train the classifier. It was configured similar to VGGNet but only till four blocks whereas VGGNet had a minimum of 13 convolutional blocks.

The proposed network had four convolutional blocks with the kernel size of 3x3 followed by batch normalization layer that enables the network to converge faster by standardizing each mini-batch inputs (Ioffe & Szegedy, 2015). Then had a max pooling layer of pool size 2x2 to compute the maximum value from the input feature maps followed by dropout layer with a rate of 25% to help the model fight overfitting. Every layer used the ReLU activation function given in Equation (1), except for the final fully connected dense layer that used sigmoid activation given in Equation (2) to get two class output probabilities. The loss was monitored using binary cross-entropy function provided in the Equation (3) and the Adam optimizer that estimates the adaptive learning rate for all parameters that are involved in the training process and considered the best optimizer for CNN (Yaqub, et al., 2020) with the default learning rate of 1e-3 was used to optimize the network.

Equation (1):

$$f(x) = \max(0, x)$$

Equation (2):

$$f(x) = \frac{1}{1 + e^{-(x)}}$$

Equation (3):

$$H_P(q) = -\frac{1}{N}\sum_{I=1}^{N} y_i \,.\log\big(p(y_i)\big) + (1 - y_i)\,.\,\log\big(1 - p(y_i)\big)$$

The callback function ReduceLROnPlateau was used to monitor the validation loss and optimize the learning rate when it hit a plateau. Additionally, early stopping mechanism that stops the training when the overfitting starts was implemented with patience of 10 epochs and the model stopped training after 21 epochs.

3.5 Classification

After trying very possible hyperparameter tuning and experiments, the model was finalized after achieving the desired accuracy. Then the model was saved in .h5 format onto the local disk and later loaded in to make prediction and evaluate the hold out dataset that wasn't shown to the model. The output images were plotted with their corresponding class and the evaluation metrics used were the confusion matrix, precision, recall and F1-score.

Table 1. Dataset division and image category

Dataset	Class	Number of raw images	Number of images
Training	Awake	70	1400
	Drowsy	70	1400
Validation	Awake	20	400
	Drowsy	20	400
Testing	Awake	10	200
	Drowsy	10	200
Total		**200**	**4000**

4. Result Analysis

As stated in Section 3, the model was trained on Bhutanese Drowsiness dataset which had a total of 4000 images but the training portion was only 70% of the total images. The cross validation was done on 20% and the final evaluation was carried out on the remaining 10% of the total dataset with the batch size of 32. Each set had two subfolders awake and drowsy containing their respective images. The number of images in each class under each set is shown in Table 1.

The drowsiness detection model was trained on Lenovo workstation having Intel Core i5 (10^{th} Generation) CPU with 2.5 GHz and up to 5GHz clock speed, 8 MB cache memory, 8GB DDR4 2933MHz RAM and Nvidia GeForce 1650Ti GPU with 4GB VRAM and 7.5 compute capability. The accuracy results of various models trained using various hyperparameters before finalizing the model is given in Table 2.

Table 2. Model variations and associated accuracies

Model configuration	Accuracy (%)		
	Train	Validation	Test
Baseline model (11 epoch)	97.66	81.25	85.75
Baseline model (100 epoch)	99.93	81.88	91.25
Baseline model + Dropout	99.93	80.26	78.25
Baseline model +Early stopping	97.5	67.87	79.75
Finalized model	**99.90**	**92.75**	**93.25**

The overall accuracy of the final model with four convolutional blocks on the training and validation set are 99.9% and 92.75% respectively. Their associated loss in the training set went down till 0.0073 whereas the validation loss was recorded as 0.1224 when the training stopped at 21^{th} epoch using the early stopping mechanism as presented in Figure3.

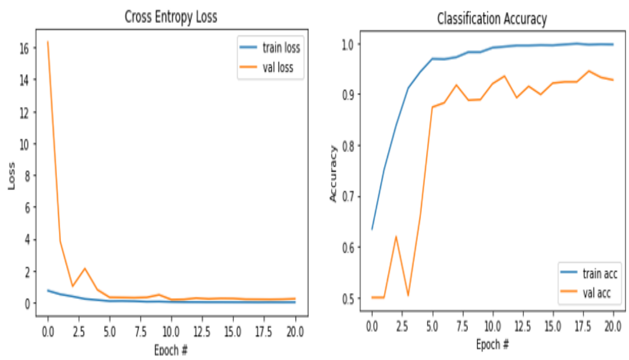

Figure 3. Train and validation accuracy (right) and loss (left).

The confusion matrix in Figure 4 which shows the number of correct predictions and incorrect predictions presents the performance measure of the drowsiness detection system on the test dataset of 400 images. From the matrix, it is concluded that awake state is predicted correctly most of the time but the classifier has difficult time in detecting the drowsy state as it is observed that drowsy state was incorrectly predicted as awake state for 27 times. The reason could be due to the network being not able to locate the face region, especially eyes of the input images.

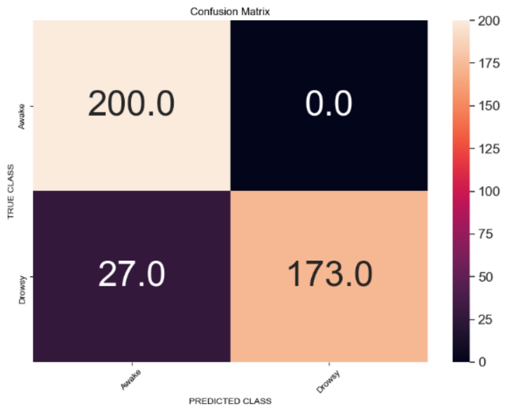

Figure 4. Confusion matrix on test dataset.

Classification report was also generated for the test dataset evaluation as seen in Table 3. This report presents the individual classes precision, recall, F1-score and the overall weighted average of the model on test dataset. The harmonic mean of precision and recall gives the 1-score. The awake class had precision of 88%, recall of 100% and higher F1-score of 94%. The drowsy class had 100% precision, 86% recall and 93% F1-score and the final weighted average of the system was observed at 93% for this particular test dataset.

Table 3. Tabulation of classification report

Class	Precision	Recall	F1-score
Awake	0.88	1.00	0.94
Drowsy	1.00	0.86	0.93
Weighted Avg	**0.94**	**0.93**	**0.93**

5. Conclusion

The main aim of this study was to develop an image processing and deep learning-based drowsiness detection system that can effectively detect drowsiness using the custom dataset. For that purpose, a Bhutanese Drowsiness dataset was created and the proposed CNN model was trained, validated and tested on the dataset. The model development used many hyperparameter tuning and approaches. The final training, validation and testing accuracies of the proposed CNN model was 99.9%, 92.75% and 93.25% respectively. The system showed a good performance measure and this can be improved further by adding more images into the dataset, applying more hyperparameter tuning such as L1 and L2 regularization. The concept of transfer learning can be applied on this dataset to help the model reduce errors and increase the accuracy on unseen data.

In the future, the system can be integrated with a web interface or with a mobile application to make drowsiness detection in real time. Further improvement on the custom dataset can be done by incorporating more signs of drowsiness such as yawning, heading bopping and body posture.

References

Blake, K. (2019, July 31). *Everything You Need to Know About Drowsiness*. Retrieved from healthline: https://www.healthline.com/health/drowsiness#:~:text=Feeling%20abnormally%20sleepy%20or%20tired,falling%20asleep%20at%20inappropriate%20times.

Brownlee, J. (2017, July 21). *A Gentle Introduction to Mini-Batch Gradient Descent and How to Configure Batch Size.* Retrieved from Machine Learning Mastery: https://machinelearningmastery.com/gentle-introduction-mini-batch-gradient-descent-configure-batch-size/

Choudhary, P., Sharma, R., Singh, G., & Das, S. (2016). A Survey Paper On Drowsiness Detection & Alarm System for Drivers. *International Research Journal of Engineering and Technology (IRJET)*, pp.1433 - 1437.

Dwivedi, K., Biswaranjan, K., & Sethi, A. (2014). Drowsy Driver Detection using Representation. *Advance Computing Conference.*

Ghoddoosian, R., Galib, M., & Athitsos, V. (2019). A Realistic Dataset and Baseline Temporal Model for Early Drowsiness Detection. *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 178-187.

Hashemi, M., Mirrashid, A., & Shirazi, A. B. (2020). CNN-based Driver Drowsiness Detection. *arXiv:2001.05137v2*, pp 1 - 8.

Ioffe, S., & Szegedy, C. (2015, February 11). *Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift.* Retrieved from arXiv: https://arxiv.org/abs/1502.03167

Jabbar, R., Al-Khalifa, K., Kharbeche, M., Alhajyaseen, W., Jafari, M., & Jiang, S. (2018). Real-time Driver Drowsiness Detection for Android Application Using Deep Neural Networks Techniques . *The 9th International Conference on Ambient Systems, Networks, and Technologies*, (pp. pp.400-407).

Khan, R., Menon, S., Patil, S., Anchan, S., & R, S. L. (2019). Human Drowsiness Detection System.

International Journal of Engineering and Advanced Technology , pp.316 - 319.

Park, S., Pan, F., & Yoo, C. (2017). Driver Drowsiness Detection System Based on Feature Representation Learning Using Various Deep Networks. *Asian Conference on Computer Vision.*

Poursadeghiyan, M., Mazloumi, A., G. N., Baneshi, M. M., Khammar, A., & Ebrahimi, M. H. (2018). Using Image Processing in the Proposed Drowsiness Detection System Design. *Iran J Public Health*, pp.1370-1377.

Suni, E. (2020, July 31). *Medical and Brain Conditions that causes Excessive Sleepiness*. Retrieved from Sleep Foundation: https://www.sleepfoundation.org/physical-health/medical-and-brain-conditions-cause-excessive-sleepiness

Vanjani, H. B., & Varyani, U. (2019). Identify Dozyness of Person Using Deep Learning. *International Journal of Applied Engineering Research*, pp. 845-848.

Yaqub, M., Feng, J., Zia, M. S., Arshid, K., Jia, K., Rehman, Z. U., & Atif Mehmood. (2020, July 3). State-of-the-Art Optimizer for Brain Tumor Segmentation in Magnetic Resonance Images. *Brain sciences*, 427. Retrieved May 12, 2020, from https://doi.org/10.3390/brainsci10070426