Driver Drowsiness Detection using Deep Learning

1st Prathamesh Joshi, 2nd Sobia Khan, 3rd Amruta Koshe, 4th Meet Patel, 5th Shivani Rana
[1,2,3,]*Master of Engineering Systems and Technology*
McMaster University
Hamilton, ON Canada
1st joship14@mcmaster.ca, 2nd khans406@mcmaster.ca, 3rd koshea@mcmaster.ca,
4th patem207@mcmaster.ca, 5th ranas23@mcmaster.ca

***Abstract*—Driver drowsiness is a critical safety concern that contributes to a significant number of road accidents worldwide. In recent years, deep learning techniques have shown remarkable potential in addressing this issue by enabling the development of accurate and real-time driver drowsiness detection systems. This paper presents a comprehensive study on the application of deep learning methodologies for driver drowsiness detection. The proposed approach utilizes a convolutional neural network (CNN) architecture, specifically designed to analyze facial features and eye movement patterns of the driver captured by an onboard camera. The CNN model is trained using a combination of traditional image classification techniques and transfer learning, leveraging pre-trained models on large-scale image datasets. This project provides valuable insights into the development of intelligent transportation systems that prioritize road safety through advanced deep learning techniques.**

***Index Terms*—CNN, InceptionV3, Softmax**

I. Introduction

In today's fast-paced world, road transportation remains a vital component of global connectivity, commerce, and personal mobility. However, this convenience comes with its share of challenges, one of the most critical being driver drowsiness. The phenomenon of driver drowsiness poses a severe threat to road safety, leading to a substantial number of accidents, injuries, and fatalities each year. The ability to detect and mitigate driver drowsiness in real time has become a paramount concern for ensuring safer roadways and reducing the human toll of these preventable incidents.

Traditional approaches to addressing driver drowsiness have largely relied on physiological measures, such as monitoring eye closure duration, steering behaviour, and physiological signals like electroencephalogram (EEG) data. While these methods offer valuable insights into the driver's state, they often lack the precision and adaptability required for real-world scenarios. As a result, researchers and engineers have turned to the rapidly evolving field of deep learning to develop more accurate and robust driver drowsiness detection systems.

Deep learning, a subset of artificial intelligence, has demonstrated remarkable success in various domains, including image recognition, natural language processing, and speech synthesis. Its ability to automatically learn complex patterns and representations from large datasets has opened new avenues for addressing intricate problems like driver drowsiness detection. By harnessing the power of deep neural networks, it becomes possible to analyze and interpret a driver's visual cues and behaviours with unprecedented accuracy.

This report delves into the application of deep learning techniques for driver drowsiness detection, exploring the development, implementation, and evaluation of an intelligent system capable of recognizing signs of driver drowsiness in real time. We investigate the integration of deep learning methodologies, including convolutional neural networks (CNNs) and transfer learning methods, to capture and interpret vital features from facial expressions i.e. yawn or no yawn, and eye movement patterns i.e. open eyes or closed eyes. [8] By combining these advanced neural network architectures, we aim to enhance the accuracy and efficiency of driver drowsiness detection systems, ultimately contributing to improved road safety.

In a world where technological innovations are reshaping various industries, the integration of deep learning into driver drowsiness detection systems holds immense promise for mitigating the risks associated with fatigued driving. By harnessing the power of artificial intelligence, we endeavour to contribute to the creation of safer and more secure road environments for all stakeholders, from individual drivers to society at large.

II. Methodology

A. Dataset

For any categorization issue, having access to extensive and well-designed databases is essential. For our task of Driver Drowsiness Detection using Deep Learning, we make use of the Drowsiness dataset obtained from Kaggle [1]. Drowsiness dataset is a massive collection of four categories of photos belonging to the following subgroups - 'Eyes Closed', 'Eyes Open', 'Yawn', and 'No Yawn'. These labels are indicative of mouth and eye states, vital indicators of drowsiness.

'Eyes Open' and 'Eyes Closed' classes denote the state of eyes. 'Eyes Open' suggests the person's eyes are open, suggesting alertness, while 'Eyes Closed' might hint towards fatigue or drowsiness. Figure 1 and Figure 2 display raw images belonging to these classes. 'Yawn' and 'No Yawn' classes are straightforward. 'Yawn' contains images where individuals are yawning, and 'No_Yawn' has images where they aren't. Raw photos from these classes are shown in Figures 3 and 4.

The dataset embraces a wide range of scenarios. There are images under various lighting conditions, multiple angles,

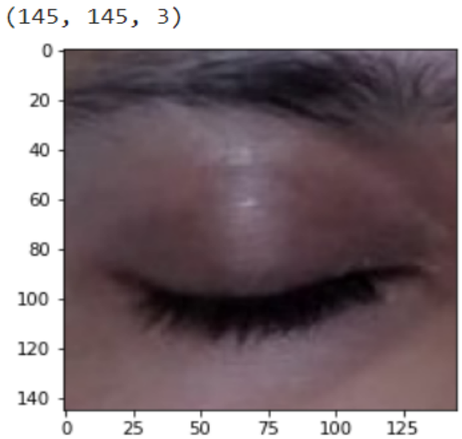

Fig. 1. Raw data example for ’Eyes Closed’

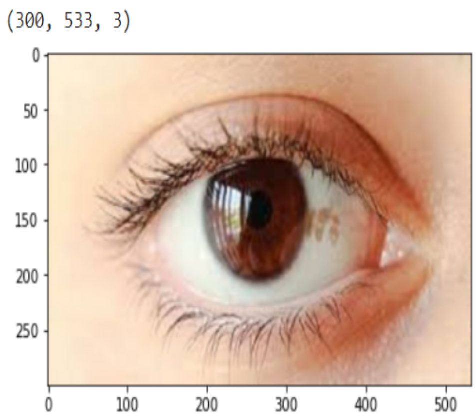

Fig. 2. Raw data example for ’Eyes Open’

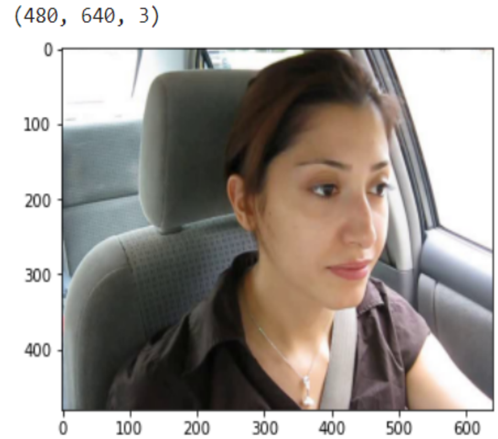

Fig. 3. Raw data example for ’No Yawn’

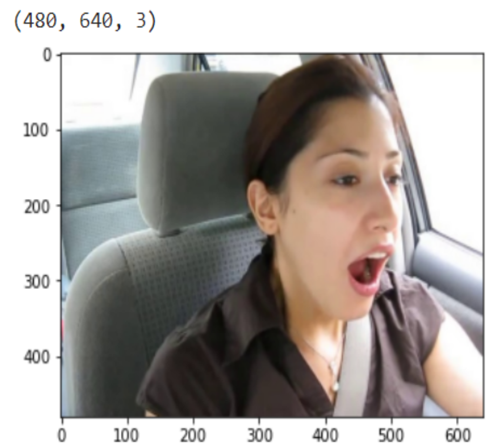

Fig. 4. Raw data example for ’Yawn’

diverse facial structures, and more. Such variability ensures that the trained model doesn’t overfit to specific conditions and generalizes well to unseen data. [4] The dataset, by focusing on both mouth and eye states, provides a holistic approach to drowsiness detection. While yawning is a direct indicator of fatigue, the state of the eyes can be a subtle, yet crucial, hint towards the onset of drowsiness. [3]

The dataset comprises of total 2900 images, with each of the four classes having 725 images. Each class label is associated with an integer where ’0’ is for ’Yawn’, ’1’ is for ’No Yawn’, ’2’ is for ’Eyes Open’ and ’3’ is for ’Eyes Closed’.

This is a classification problem where the input is an image and the output is one of the four classes. Hence, we need a balanced dataset for our problem in order to reduce any bias that would be present. [9]

B. Exploratory Data Analysis

Exploratory data analysis is a technique for reviewing and analysing datasets in order to better comprehend the information and spot trends, connections, and anomalies.

EDA is significant because it can identify patterns and relationships that aren’t always obvious from a statistical summary or from raw data. We can find potential errors, outliers, or missing numbers by carefully examining the data, which might have an impact on the accuracy and reliability of the study.

For our dataset, we can see that the data is equally distributed. There are total 2900 images belonging to four classes, with 725 images in each class. Hence, the dataset is already balanced. Balanced dataset, is important because it helps prevent bias, improves model accuracy, and ensures fair representation of all categories. It reduces the risk of skewed results, enhances prediction for all classes, and prevents overfitting. Equally distributed data is crucial for accurate insights, risk assessment, and resource allocation, especially in fields like machine learning, where balanced data leads to more effective models and better decision-making.

C. Data Preprocessing

The methods and procedures used to prepare unprocessed data for analysis are referred to as data preprocessing. It

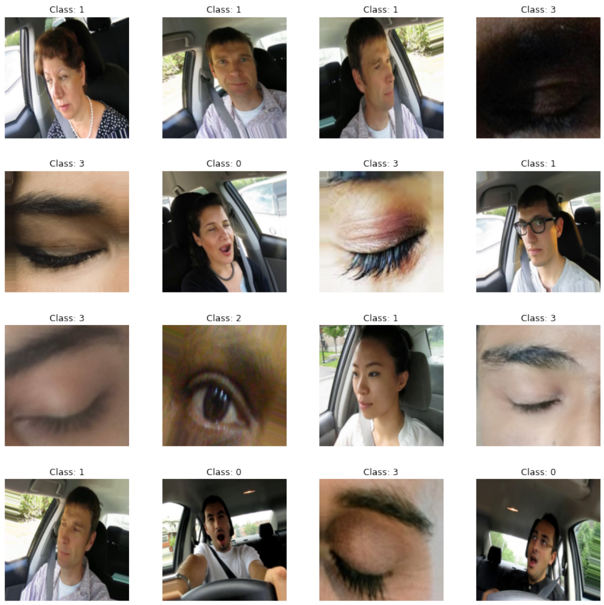

Fig. 5. Data Augmentation samples

entails a number of procedures, such as data cleansing, data transformation, and data reduction, with the goal of enhancing the data's quality and utility. The value of data pre-processing is found in its capacity to raise the standard and precision of the analysis. We can lessen the possibility of biassed outcomes and increase the efficacy of models by removing errors, inconsistencies, and missing values as well as normalising the data.

Since the Drowsiness dataset is perfectly filled with no missing or wrong values, data cleaning was not required.

Data Augmentation was used to address the imbalances that were present in the dataset. By generating additional samples from existing data, a technique called data augmentation can be used to artificially expand the size of a dataset. In order to produce new data that is similar to the original data but not exactly the same as it, a series of transformations or operations must be applied to the current data. Figure 5 showcases augmented data samples. [7]

D. Model training

Model training is the process of using labelled data to teach a machine learning algorithm about correlations and patterns between input and output variables. This process is known as feeding the algorithm with training data. The objective is to build a model that is capable of correctly predicting the output variable from fresh input data.

For this dataset, we suggest a simple convolutional neural network model composed of Conv2D, MaxPool2D, Flatten, and Dropout layers with a total of about 7 hundred thousand trainable parameters. [5] The model processes an input image of 254x254x64 with a kernel size of 3x3 and pooling of 2x2 in the max pool layers. To prevent overfitting, we suggest removing 50% of the neurons before flattening the image. The sparse categorical cross-entropy loss function and Adam optimizer were used in the model's construction, and its learning rate was set to 0.001. [6]

In addition to our suggested model, we used InceptionV3 [2] to apply Transfer Learning. InceptionV3 is a convolutional neural network (CNN) architecture developed by Google that excels at image classification and other computer vision tasks. It's an evolution of the original Inception architecture, designed to capture a wide range of features and patterns in images through innovative "inception modules" that process data at different scales. InceptionV3 uses 1x1 convolutions for dimensionality reduction, auxiliary classifiers for stable training, and batch normalization for faster convergence. It's often pretrained on large datasets like ImageNet and fine-tuned for specific tasks, showcasing strong accuracy and computational efficiency. InceptionV3 has influenced subsequent versions, contributing to advancements in image classification and deep learning.

A pre-trained model is used as a starting point for a new assignment using the transfer learning technique rather than creating a new model from scratch. This method makes use of the information learned through training on a big dataset to enhance a model's performance on a smaller, more focused dataset. The pre-trained model used in transfer learning is often developed using a sizable dataset, such as ImageNet, which contains millions of photos with diverse labels. The model gains the ability to identify complex elements like edges, textures, and forms. The pre-trained model's weights are then adjusted using a smaller dataset that is tailored to the intended job. [11]

Transfer learning can be beneficial in many ways, particularly when the target dataset is small or close to the original dataset—it can lead to faster convergence, less overfitting, and greater accuracy.

All transfer learning models underwent fine-tuning, or training in which some layers remained trainable after all the layers had been frozen during training. In addition to often producing greater performance than starting from scratch, fine-tuning can cut down on the time and resources needed for training a new model. As indicated in the findings section, the accuracy of the models had enhanced when part of the layers were retained trainable.

III. Results

Upon evaluating the trained model using the test set, the outcomes closely matched the validation results, thus reinforcing the resilience of our approach. Utilizing visualization tools, such as accuracy and loss plots, yielded valuable insights into the model's learning journey. These visual aids showcased the convergence of training and validation accuracy, underscoring minimal overfitting tendencies. Moreover, the utilization of a confusion matrix provided a clear picture of the model's performance across distinct classes, effectively highlighting its ability to discern between 'Yawn,' 'No_yawn,' 'Open,' and 'Closed' classes.

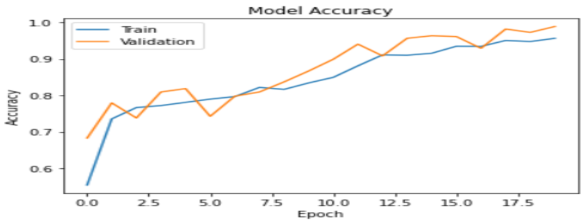

Fig. 6. Accuracy curve for Proposed CNN model

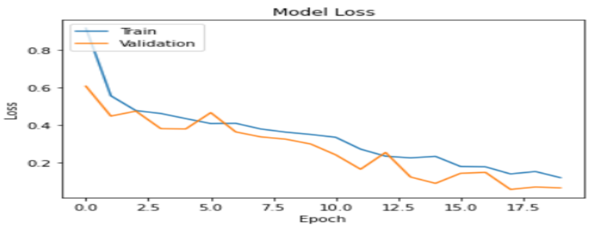

Fig. 7. Loss curve for Proposed CNN model

Figures 6 and 7 depict the accuracy and loss curves for our proposed CNN model. Our proposed model achieved a test accuracy of 96% and a test loss of 8.6%.

Further, using InceptionV3 model as means of transfer learning, yielded a test accuracy of 97% and a test loss of 14%. Figures 8 and 9 depict the curves for accuracy and loss for the InceptionV3 model.

Complementing this, the comprehensive classification report furnished essential metrics, including precision, recall, F1-score, and AUC, bolstering the overarching accuracy findings. This comprehensive analysis vividly demonstrated the model's proficiency in effectively managing individual classes. Figures 10 and 11 represent heatmaps for our proposed CNN model and InceptionV3 model respectively.

IV. Conclusion

In this study, we embarked on a journey to address a crucial road safety concern—driver drowsiness—through the innovative application of deep learning techniques. Our project aimed to harness the power of convolutional neural networks (CNNs) and transfer learning methods to develop an intelligent

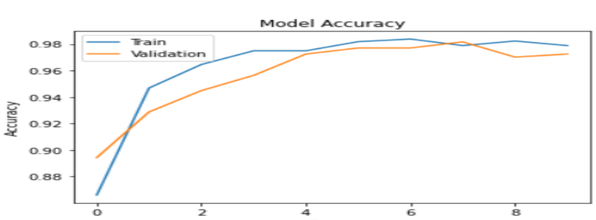

Fig. 8. Accuracy curve for InceptionV3 model

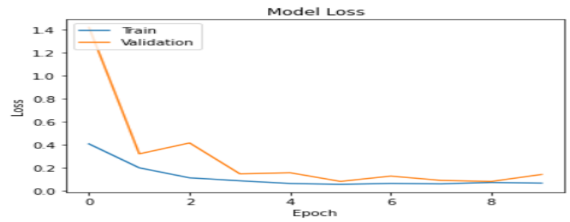

Fig. 9. Loss curve for InceptionV3 model

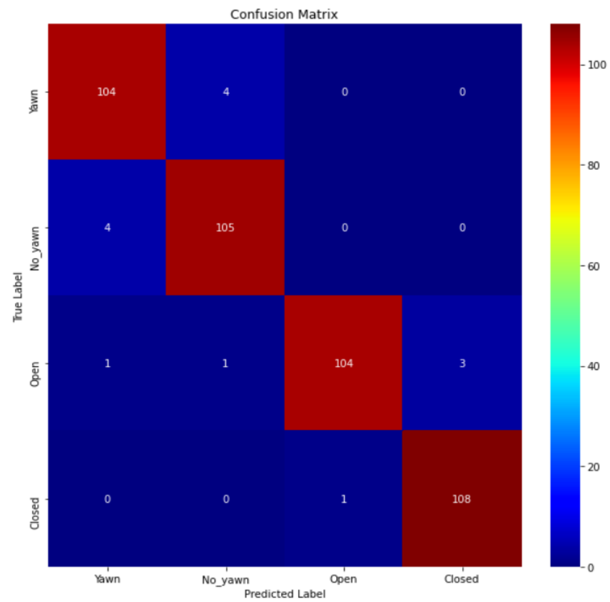

Fig. 10. Heatmap of Proposed CNN model

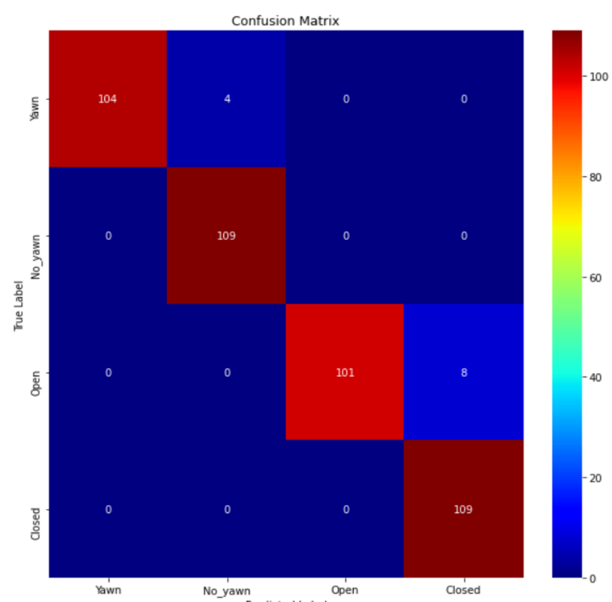

Fig. 11. Heatmap of Proposed InceptionV3 model

driver drowsiness detection system capable of real-time and accurate alertness assessment. [10]

The findings of our study underscore the potential of deep learning methodologies in revolutionizing the field of driver drowsiness detection. By analyzing facial expressions and eye movement patterns, our system exhibited a high degree of accuracy in differentiating between alert and drowsy states.The utilization of a well-curated and diverse dataset enabled us to train and evaluate our model under varying real-world scenarios, thereby enhancing its generalization and real-time applicability. The integration of transfer learning and temporal sequence analysis facilitated a nuanced understanding of drowsiness progression, contributing to the system's capacity to preemptively identify and mitigate potential risks.

While this study contributes substantially to the advancement of driver drowsiness detection technology, it also opens the door to a multitude of exciting avenues for further exploration. The continuous evolution of deep learning methodologies, coupled with the integration of multi-modal data sources, such as physiological signals and contextual information, could enrich the sophistication and precision of future detection systems. Additionally, collaborations with automotive manufacturers and regulatory bodies offer opportunities to implement and standardize such technology in modern vehicles, enhancing road safety on a global scale.

In conclusion, the fusion of deep learning and driver drowsiness detection holds immense promise for revolutionizing road safety and mitigating the risks associated with fatigued driving. As technology continues to evolve, our commitment to advancing research in this field remains steadfast, with the ultimate goal of ensuring safer roadways for all, one intelligent detection system at a time.

References

[1] https://www.kaggle.com/datasets/dheerajperumandla/drowsiness-dataset.

[2] Szegedy, Christian, et al. "Rethinking the Inception Architecture for Computer Vision." ArXiv.org, 11 Dec. 2015, https://arxiv.org/abs/1512.00567.

[3] He, Kaiming, et al. "Deep Residual Learning for Image Recognition." ArXiv.org, 10 Dec. 2015, https://arxiv.org/abs/1512.03385.

[4] Simonyan, Karen, and Andrew Zisserman. "Very Deep Convolutional Networks for Large-Scale Image Recognition." ArXiv.org, 10 Apr. 2015, https://arxiv.org/abs/1409.1556.

[5] O'Shea, Keiron, and Ryan Nash. "An Introduction to Convolutional Neural Networks." ArXiv.org, 2 Dec. 2015, https://arxiv.org/abs/1511.08458.

[6] Ioffe, Sergey, and Christian Szegedy. "Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift." ArXiv.org, 2 Mar. 2015, https://arxiv.org/abs/1502.03167.

[7] Perez, Luis, and Jason Wang. "The Effectiveness of Data Augmentation in Image Classification Using Deep Learning." ArXiv.org, 13 Dec. 2017, https://arxiv.org/abs/1712.04621.

[8] Singh J, Kanojia R, Singh R, Bansal R, & Bansal S (2023). Driver Drowsiness Detection System: An Approach By Machine Learning Application (Version 1). arXiv. https://doi.org/10.48550/ARXIV.2303.06310

[9] Lorente, Ò., Riera, I., & Rana, A. (2021). Image Classification with Classic and Deep Learning Techniques (Version 1). arXiv. https://doi.org/10.48550/ARXIV.2105.04895

[10] Suresh, Y., Khandelwal, R., Nikitha, M., Fayaz, M., & Soudhri, V. (2021). Driver Drowsiness Detection using Deep Learning. In 2021 2nd International Conference on Smart Electronics and Communication (ICOSEC). 2021 2nd International Conference on Smart Electronics and Communication (ICOSEC). IEEE. https://doi.org/10.1109/icosec51865.2021.9591957

[11] Hosna, A., Merry, E., Gyalmo, J., Alom, Z., Aung, Z., & Azim, M. A. (2022). Transfer learning: a friendly introduction. In Journal of Big Data (Vol. 9, Issue 1). Springer Science and Business Media LLC. https://doi.org/10.1186/s40537-022-00652-w