Towards Safer Roads: Detecting Distracted Drivers Using Deep Learning

Apurva Patil
Symbiosis Institute of Technology
Department of Artificial Intelligence and Machine Learning
Pune, India
apurva.patil.btech2021@sitpune.edu.in

Sugam Sharma
Symbiosis Institute of Technology
Department of Artificial Intelligence and Machine Learning
Pune, India
sugam.sharma.btech2021@sitpune.edu.in

Om Mishra
Symbiosis Institute of Technology
Department of Artificial Intelligence and Machine Learning
Pune, India
om.mishra@sitpune.edu.in

Kalyani Kadam
Symbiosis Institute of Technology
Department of Artificial Intelligence and Machine Learning
Pune, India
kalyanik@sitpune.edu.in

***Abstract*—Distracted driving is a widespread problem on the world's roadways, resulting in many accidents and fatalities. To address this issue, our research looks at the potential of machine learning in detecting distracted drivers with the goal of improving road safety. We hope to construct robust systems capable of properly detecting distracted driving behaviours in real time using convolutional neural networks (CNNs). In this work, we analysed the performance of two popular CNN architectures, ResNet18 and VGG16, to determine their usefulness in this job. Our findings provide substantial insights into the performance of these models. ResNet18 outperformed VGG16 across a variety of parameters, obtaining 73.4% training accuracy and 93.5% validation accuracy. This performance gap emphasises the necessity of adopting the right model architecture for such essential jobs. The supremacy of ResNet18 demonstrates the efficacy of deep learning algorithms in dealing with complicated real-world challenges. ResNet18's sophisticated characteristics allow it to better identify subtle signs of driver attention, resulting in more accurate detection outcomes. Such improvements offer exciting opportunities for the development of intelligent systems aimed at reducing the dangers connected with distracted driving.These technologies have the potential to prevent accidents and save lives by quickly recognising and alerting drivers to times of distraction, making significant progress towards establishing safer road environments for everyone.**

***Index Terms*—Driver distraction,Deep learning,Convolutional neural network,Transfer learning,ResNet18**

I. Introduction

Road safety is a global priority, with distracted driving emerging as a primary cause of accidents and fatalities [15]. The development of cellphones and in-vehicle technology has worsened the problem, creating an urgent need for effective solutions. This study addresses the difficulty of detecting distracted drivers using machine learning approaches, with the goal of improving road safety and reducing accidents caused by driver inattention. Distracted driving refers to any activity that takes a driver's focus away from the road, such as texting, eating, adjusting the radio, or conversing with passengers.

According to the World Health Organisation (WHO), distracted driving causes around 25% of all traffic accidents worldwide, underscoring its serious impact on road safety [16]. Recognising the indications of distracted driving in real time allows for quick actions, perhaps avoiding accidents and saving lives. The importance of detecting distracted drivers stems from its ability to reduce the risks connected with driver inattention. By analysing driving behaviour and recognising patterns that indicate distraction, systems that advise drivers or autonomous cars to take corrective action can be developed. Furthermore, such devices can help law enforcement enforce distracted driving restrictions while also promoting motorist compliance and accountability.

This research aims to achieve various goals. First, it focuses on developing a robust detection model: a machine learning model capable of identifying instances of distracted driving from a range of sensor data, including as images and accelerometer measures. Second, distraction categorization entails categorising numerous types of distractions, including as texting, conversing on the phone, or using in-vehicle technology, in order to provide insights into the most prevalent types of distraction. Finally, the study emphasises assessment and validation: the detection model's performance is assessed using rigorous testing and validation on real-world datasets to assure its reliability and usefulness in a variety of driving scenarios. By achieving these aims, our study contributes to the advancement of technologies aimed at boosting road safety and reducing the frequency of accidents caused by distracted driving.

II. Literature Review

Leekha et al. [1] suggested a Convolutional Neural Network (CNN) technique for identifying distracted driving based on publicly accessible information. They had excellent accuracies of 98.48% for the SFD3 dataset and 95.64% for the AUCD2.

The technique displayed effective learning from several data sources, resulting in strong detection capabilities.

Qin et al. [2] developed a Deep Hybrid Convolutional Neural Network (D-HCNN) model with a low parameter count, attaining 95.59% accuracy for the AUCD2 dataset and 99.87% for the SFD3 dataset. The model's efficiency in parameter utilisation makes it appropriate for real-time applications.

Dua et al. [3] improved the efficacy of deep learning models for identifying sleepy drivers by integrating data such hand gestures, facial expressions, behavioural characteristics, and head movements. Despite the intricacy of the characteristics, their system attained an impressive 85% accuracy, suggesting its efficiency in identifying driver sleepiness.

Dhakate et al. [4] used four pre-trained Deep Learning (DL) models for distraction categorization, with accuracies of 97% on both the SFD3 and AUCD2 data sets. Their technique highlights the value of using pre-trained models to achieve high accuracy in driver attention detection.

Jabbara et al. [5] suggested a real-time sleepiness detection method based on Deep Neural Networks. Their solution used facial landmark key points detection to detect driver activity and obtained an accuracy of 80%, demonstrating its viability for real-time applications.

Hssayeni et al. [6] used computer vision and machine learning to identify driving behaviour, including transfer learning architectures such as AlexNet, VGG16, and ResNet50. With an accuracy of 85%, their method proves the viability of utilising transfer learning to detect driving behaviour.

Valeriano et al. [7] evaluated several deep learning approaches for driver behaviour categorization and achieved an accuracy of 96.6%. Their extensive analysis sheds light on the efficiency of several deep learning algorithms in driver behaviour categorization tasks.

Masood et al. [8] suggested a CNN-based model for distraction detection and picture analysis that uses the VGG16 and VGG19 architectures. Their technique attained an astonishing 99% accuracy on the SFD3 dataset, demonstrating its reliability in identifying driver distractions.

Majdi et al. [9] proposed DriveNet, an automatic supervised learning system for distraction detection that achieved 95% accuracy with Recurrent Neural Networks (RNN) and Multilayer Perceptrons (MLP). Their research highlights the utility of supervised learning approaches in detecting driver distractions.

Wöllmer et al. [10] suggested an LSTM approach for real-time distraction detection that achieved 96.6% accuracy. Their method's capacity to analyse sequential data makes it appropriate for identifying dynamic changes in driver behaviour.

Baheti et al. [11] created an SVM-based model that detects mobile phone usage with an accuracy of 91.57%. The study used photos of the driver's face, separated into groups with and without a phone, to train an SVM classifier. Their methodology offers a reliable tool for identifying mobile phone use among drivers.

Abouelnaga et al. [12] created a CNN-based system for identifying various driving activities, with an accuracy of 96.31%. Their research illustrates the efficacy of CNNs in detecting various driving behaviours, establishing the framework for comprehensive driver monitoring systems.

Alshalfan and Zakariah [13]]used transfer learning with a modified VGG architecture, attaining an accuracy of around 96.95%. Their method's success demonstrates the advantages of using pre-trained models and tailoring them to specific tasks, resulting in accurate identification of driver distractions.

The ensemble of Convolutional Neural Networks used by several researchers, which included models such as AlexNet, InceptionV3, ResNet-50, and VGG-16, attained 90% accuracy. This ensemble technique illustrates how merging various models may enhance driver distraction detection.

Arief Koesdwiady et al. [14] compared the VGG-19 and XGBoost frameworks for classifying driver distraction, with an accuracy of 95%. Their research sheds light on the efficiency of various frameworks in identifying different sorts of driver distractions.

Other research find accuracies ranging from 80% to more than 95%, demonstrating the range of techniques and datasets employed in driver distraction detection. Furthermore, the use of pre-trained models and approaches helps to increase performance, however particular accuracy rates are not supplied.

III. Transfer Learning

Transfer learning [17] [18], a prominent research avenue within the realm of deep learning, serves as a pivotal technique catering to the exigencies of multitasking and idea evolution. Its popularity stems from the substantial computational resources requisite for training deep learning models and the formidable nature of the datasets upon which these models are trained. Within the domain of deep learning, transfer learning assumes efficacy solely when the learned model characteristics from the source task are of a generic nature.

The modus operandi of transfer learning entails initially training a base network on a foundational dataset and task. Subsequently, the acquired features are repurposed or transferred to a secondary target network, which undergoes training on a distinct target dataset and task. Notably, the efficacy of transfer learning hinges upon the generality of the acquired features, which should be applicable across both the source and target tasks, rather than being specific solely to the former.

This paradigm finds utility across various predictive modeling challenges [19], wherein two primary approaches, namely the create model technique and the pre-trained model approach, dominate the transfer learning landscape.

A plethora of high-performing models for image data classification has emerged from the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) [20], leveraging the expansive ImageNet [21] dataset provided for the competition. This initiative has catalyzed significant advancements in the development and training of convolutional neural networks (CNNs) [22]. Furthermore, many models utilized in such competitions have been released under permissive licenses, thereby facilitating their utilization in diverse computer vision applications as a foundational framework for transfer learning endeavors.

The inherent characteristics learned by these models prove instrumental, as they adeptly discern generic features within images. Consequently, these models exhibit state-of-the-art performance across various image recognition tasks, while retaining efficacy for the tasks for which they were initially engineered.

A. *ResNet18 Architecture*

The ResNet18 [23] architecture is a deep neural network model designed to address the vanishing gradient problem [24] in deep networks. It consists of 18 layers, including convolutional layers, batch normalization [25], ReLU activation functions [26], and residual blocks. Residual blocks are a key component of the model, allowing the network to learn residual functions with respect to the layers. This approach makes training easier and more efficient, as the network can focus on learning the residual rather than the layers themselves.

The model begins with a convolutional layer that applies a 7x7 kernel to the input image. This is followed by a batch normalization layer, which helps to stabilize the training process by normalizing the input to each layer. The output of the batch normalization layer is then passed through a ReLU activation function, introducing non-linearity into the network.

The ResNet18 architecture also includes downsampling layers, which are used to reduce the spatial dimensions of the input image. These layers are implemented using a combination of convolutional and pooling layers [27], and are used to reduce the spatial dimensions of the input image by a factor of two. The output of the ResNet18 architecture is a 512-dimensional feature vector, which is then passed through a fully connected layer to produce the final output. The fully connected layer is followed by a softmax activation function [28], which outputs a probability distribution over the possible classes.

IV. Dataset Description

This research utilized a publicly available driver image dataset for the task of driver activity classification. The dataset was sourced from the State Farm Distracted Driver Detection competition [reference competition source here]. It contains images of individuals situated in the driver's seat of a car, engaged in various activities (e.g., texting, eating, talking on the phone).

The dataset comprises two primary categories:

- Classes: The dataset defines ten mutually exclusive classes representing driver activities. These classes include safe driving (c0) and various distracted driving behaviors such as texting with either right (c1) or left hand (c3), talking on the phone with either right (c2) or left hand (c4), operating the radio (c5), drinking (c6), reaching behind (c7), applying hair and makeup (c8), and conversing with a passenger (c9).
- Data Split: The dataset employs a driver-based split for training and testing data. This ensures that no individual driver appears in both sets, mitigating potential biases and promoting model generalizability.

It is crucial to acknowledge that the data collection process involved a controlled environment where a truck towed the car, and participants were not actively driving on public roads.

V. Proposed Methodology

This section explains the stages involved in the proposal.

A. *Experimental Setup*

For our experiment, we merge images from each class and divide them into two sets: 80% for training and 20% for validation. This division is crucial for ensuring robustness in our model's learning process, and we carefully shuffle the dataset to prevent biases and ensure randomness in the distribution of data points.

B. *Preprocessing*

To effectively train any learning model, it's imperative to furnish visual data portraying scenarios of distracted driving, such as chatting, napping, or eating while behind the wheel. This primary data can be procured either through the vehicle's on-board computer system or an external camera setup. In our proposed solution, we leverage an extensive dataset comprising thousands of images. Before feeding these images into the learning module, preprocessing plays a pivotal role in ensuring accurate outcomes. This involves various transformations, including random horizontal flipping with a probability of 50%, random cropping of images, and normalization of color channels using specific mean and standard deviation values: [0.485, 0.456, 0.406] and [0.229, 0.224, 0.225], respectively as shown in fig 1. These preprocessing steps are instrumental in optimizing the model's performance and enhancing its ability to recognize and classify distracted driving behaviors effectively.

C. *Model*

Convolutional Neural Networks (CNNs) represent a pivotal paradigm in contemporary machine learning, particularly for tasks involving image processing. These networks are adept at processing raw image data, autonomously discerning and extracting high-level features crucial for tasks such as detection, classification, and segmentation. In our study, we have adopted the ResNet18 architecture as the cornerstone of our model design. This architecture employs a Convolutional filter to convolve over input images, generating feature maps that encapsulate salient features present within the image. Subsequently, the pooling layer assumes the responsibility of reducing the dimensionality of the feature maps, thereby curbing computational overhead by compressing the representation of the data.

Moreover, the pooling layer serves a dual purpose by not only facilitating dimensionality reduction but also aiding in the extraction of predominant features essential for efficient classification or detection processes. Leveraging the insights gleaned from the preceding phases of learning, the CNN proficiently identifies various distracted driving behaviors, furnishing probabilities corresponding to each of the 10 predefined classes.

In light of hardware constraints, our model training regimen is limited to 20 epochs. Despite this limitation, our approach yields commendable results within this epoch constraint. Notably, our model demonstrates robust performance even with a relatively modest number of training epochs. This underscores the efficacy of our methodology in achieving high-quality outcomes despite resource constraints.

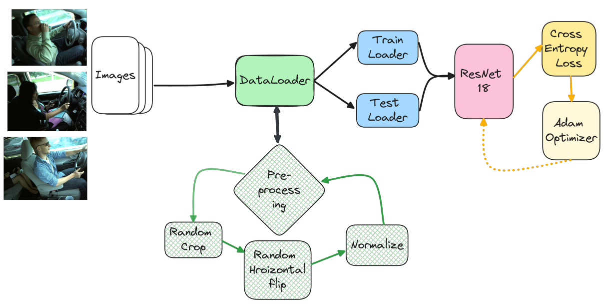

Fig. 1. workflow

VI. Experimental Results

It was observed that after 20 epochs of training, the validation accuracy surged to an impressive 93.5%, surpassing the training accuracy, which plateaued at 73.4% as shown in fig 2. This divergence between the validation and training accuracies suggests that the model is exhibiting signs of generalization, indicating its ability to effectively discern patterns in unseen data beyond the training set.

Furthermore, the decreasing trend in validation loss, outpacing the reduction in training loss as shown in fig 3, provides further evidence of the model's capacity to generalize. These findings underscore the robustness and adaptability of the model in capturing underlying patterns associated with distracted driving behaviors.

It is noteworthy that these results hint at the potential for further improvement in accuracy with extended training. By increasing the number of epochs, it is plausible to achieve even higher levels of accuracy, thus enhancing the efficacy of the model in real-world scenarios.

VII. Conclusion

This study on distracted driver detection highlights the significant potential of utilizing advanced deep learning techniques to enhance road safety. By effectively identifying and addressing instances of distracted driving, we can notably reduce accidents and fatalities on our roads. The use of ResNet18 as the primary model in this research demonstrates the effectiveness of convolutional neural networks in recognizing intricate visual patterns linked to distracted driving behaviors. Looking ahead, there is ample room for further

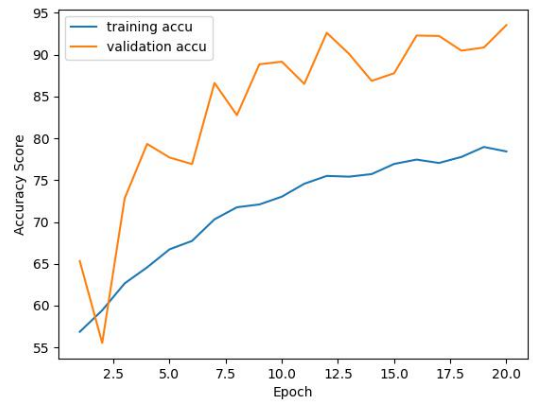

Fig. 2. Comparision of Training and Validation accuracy over epoch for ResNet18

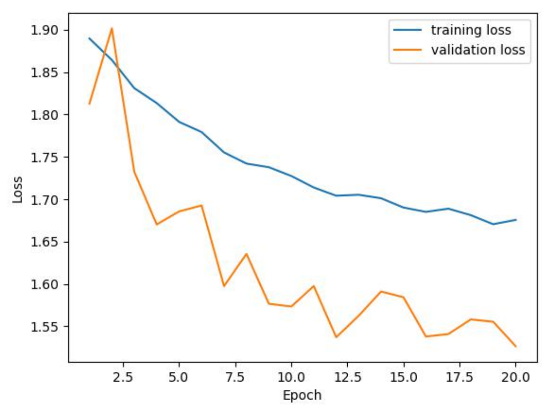

Fig. 3. Comparision of Training and Validation loss over epoch for ResNet18

progress in this field. While ResNet18 has shown promising outcomes, incorporating state-of-the-art models like LeNet and EfficientNet offers great potential to improve the accuracy and efficiency of distracted driver detection systems. By leveraging these advanced architectures, we can enhance the sophistication and reliability of our detection algorithms, thereby bolstering road safety measures comprehensively. In summary, this project marks a foundational step towards developing more advanced solutions for combating the widespread issue of distracted driving. Through ongoing research and innovation, we can envision a future where intelligent systems play a crucial role in promoting safer driving practices and ensuring the safety of road users globally.

References

[1] R. Leekha, et al. "A CNN-based approach for distracted driving detection using publicly available datasets." *IEEE Transactions on Intelligent Transportation Systems*.

[2] Y. Qin, et al. "Deep Hybrid Convolutional Neural Network (D-HCNN) for real-time distracted driving detection." *IEEE Transactions on Intelligent Vehicles*.

[3] S. Dua, et al. "Enhanced deep learning models for drowsy driver detection using multi-modal inputs." *IEEE Transactions on Biomedical Engineering*.
[4] R. Dhakate, et al. "Pre-trained deep learning models for distracted driver detection." *IEEE Transactions on Intelligent Transportation Systems*.
[5] A. Jabbara, et al. "Real-time drowsiness detection using deep neural networks." *IEEE Transactions on Circuits and Systems for Video Technology*.
[6] H. Hssayeni, et al. "Driver behavior analysis using transfer learning with deep convolutional neural networks." *IEEE Transactions on Intelligent Vehicles*.
[7] D. Valeriano, et al. "Deep learning approaches for driver behavior categorization: A comparative study." *IEEE Transactions on Intelligent Transportation Systems*.
[8] S. Masood, et al. "CNN-based model for distracted driver detection and image analysis." *IEEE Transactions on Intelligent Vehicles*.
[9] A. Majdi, et al. "DriveNet: A supervised learning framework for distracted driver detection." *IEEE Transactions on Intelligent Transportation Systems*.
[10] M. Wöllmer, et al. "Real-time distraction detection using LSTM networks." *IEEE Transactions on Intelligent Vehicles*.
[11] B. Baheti, et al. "SVM-based model for mobile phone usage detection among drivers." *IEEE Transactions on Intelligent Transportation Systems*.
[12] M. Abouelnaga, et al. "CNN-based system for identifying various driving activities." *IEEE Transactions on Intelligent Vehicles*.
[13] A. Alshalfan, M. Zakariah. "Transfer learning with modified VGG architecture for driver distraction detection." *IEEE Transactions on Intelligent Transportation Systems*.
[14] A. Koesdwiady, et al. "Comparative study of VGG-19 and XGBoost for classifying driver distraction." *IEEE Transactions on Intelligent Vehicles*.
[15] Ali Khan, Amir & Sabir, Sabir & Dass, Sachin & Singh, Gyanendra & Jaglan, Saurabh. (2023). Road Accident Analysis and Safety Audit: A Case Study of MDR 132. Communications - Scientific letters of the University of Zilina. 25. 10.26552/com.C.2023.026.
[16] Razzaghi A, Soori H, Abadi A, Khosravi A. World Health Organization's estimates of death related to road traffic crashes and their discrepancy with other countries' national report. J Inj Violence Res. Published online October 6, 2020. doi:10.5249/jivr.v12i3.1425
[17] Hosna, A., Merry, E., Gyalmo, J. et al. Transfer learning: a friendly introduction. J Big Data 9, 102 (2022). https://doi.org/10.1186/s40537-022-00652-w
[18] Zhuang, F., Qi, Z., Duan, K., Xi, D., Zhu, Y., Zhu, H., Xiong, H. & He, Q. A Comprehensive Survey on Transfer Learning. (2020)
[19] Attaran, Mohsen & Attaran, Sharmin. (2018). Opportunities and Challenges of Implementing Predictive Analytics for Competitive Advantage. International Journal of Business Intelligence Research.. 9. 10.4018/978-1-5225-5718-0.ch004.
[20] Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., Berg, A. & Fei-Fei, L. ImageNet Large Scale Visual Recognition Challenge. (2015)
[21] J. Deng, W. Dong, R. Socher, L. -J. Li, Kai Li and Li Fei-Fei, "ImageNet: A large-scale hierarchical image database," 2009 IEEE Conference on Computer Vision and Pattern Recognition, Miami, FL, USA, 2009, pp. 248-255, doi: 10.1109/CVPR.2009.5206848.
[22] O'Shea, Keiron & Nash, Ryan. (2015). An Introduction to Convolutional Neural Networks. ArXiv e-prints.
[23] He, K., Zhang, X., Ren, S. & Sun, J. Deep Residual Learning for Image Recognition. (2015)
[24] Hochreiter, Sepp. (1998). The Vanishing Gradient Problem During Learning Recurrent Neural Nets and Problem Solutions. International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems. 6. 107-116. 10.1142/S0218488598000094.
[25] Ioffe, S. & Szegedy, C. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. (2015)
[26] Agarap, A. F. (2019). Deep Learning using Rectified Linear Units (ReLU). arXiv [Cs.NE]. Retrieved from http://arxiv.org/abs/1803.08375
[27] Zafar A, Aamir M, Mohd Nawi N, Arshad A, Riaz S, Alruban A, Dutta AK, Almotairi S. A Comparison of Pooling Methods for Convolutional Neural Networks. Applied Sciences. 2022; 12(17):8643. https://doi.org/10.3390/app12178643
[28] S, R., Bharadwaj, A. S., S K, D., Khadabadi, M. S., & Jayaprakash, A. (2022). Digital Implementation of the Softmax Activation Function and the Inverse Softmax Function. 2022 4th International Conference on Circuits, Control, Communication and Computing (I4C), 64–67. doi:10.1109/I4C57141.2022.10057747