Driver Drowsiness Detection Using Deep Learning *

Saurabh Kumar Singh
Department of CSE
NIT Karnataka
Surathkal, Mangalore, India
saurabhmsvm@gmail.com

Annappa B.
Department of CSE
NIT Karnataka
Surathkal, Mangalore, India
annappa@ieee.org

Saraswati Koppad
Department of CSE
NIT Karnataka
Surathkal, Mangalore, India
saraswatikoppad@gmail.com

***Abstract*—Drowsiness is the main factor that leads to road accidents. Statistics show that drowsy drivers are responsible for accidents. A National Sleep Foundation survey indicates that 20% of drivers are fatigued and sleep deprived when operating a vehicle. This study offers a novel method for detecting driver drowsiness in this setting by utilizing deep learning techniques. The study leverages the MRL Eye dataset to capture eye states for analysis. To distinguish between drowsy and awake states, single architectures like InceptionNetV3, MobileNetV2, and EfficientNetB2 are used, along with a stacked ensemble model that combines these architectures, achieving highest accuracy of 0.9621%. This study highlights the potential of deep learning to improve road safety while advancing our awareness of and response to the problems associated with drowsy driving.**

Keywords: Driver Drowsiness, Deep Learning, InceptionV3, MobileNetV2, Stacking Ensemble

I. Introduction

Drowsy driving is a critical safety concern, posing a significant threat on the road. The National Highway Traffic Safety Administration (NHTSA) [1] estimates that drowsiness while driving causes 100,000 police-reported crashes each year, with 50,000 injuries and around 800 fatalities as a result. These alarming statistics highlight the urgent need for effective driver drowsiness detection systems.

Such systems analyze various data sources to assess a driver's alertness. These can be categorized into:

Physiological signals: Techniques like Electroencephalography (EEG) and Electrocardiography (ECG) measure brain activity and heart rate, but they require intrusive sensors.

Vehicle-based data: Steering wheel movement, lane departure, and braking behavior provide insights, but require additional hardware installations.

Behavioral data [2]: This approach analyzes driver actions captured by a camera, focusing on facial features, head movements, and eye activity like blinking and yawning. This method offers several advantages: it's non-intrusive, cost-effective, and requires minimal additional equipment.

Deep learning [3] has revolutionized image analysis, making it a powerful tool for driver drowsiness detection. Deep learning models can automatically learn complex patterns from image data, effectively identifying signs of fatigue in a driver's facial expressions and eye movements.

This research leverages the Media Research Lab (MRL) Eye dataset [4], specifically designed for driver drowsiness detection tasks. The study explores the impact of different activation functions within deep learning models, which introduce non-linearity and allow the model to capture complex relationships in the data. By investigating three pre-trained deep learning architectures – InceptionV3, EfficientNetB2, and MobileNetV2 – with various activation functions in their final layers, the research aims to identify the optimal combination for accurate driver drowsiness detection.

To achieve this, the study evaluates both single stand-alone architecture models and ensemble deep learning model that combine their outputs to identify the optimal combination of model architecture and activation function for accurate driver drowsiness detection. The research advances deep learning based driver drowsiness detection systems and ultimately improves road safety by evaluating the data.

II. Related Work

Elena Magán et al. [5] explore the landscape of driver drowsiness detection within the realm of advanced driving assistance systems (ADAS), unveiling a multifaceted terrain of research endeavors. Past studies have traversed diverse methodologies encompassing physiological signals, vehicle-based metrics, and vision-based techniques to discern driver fatigue. In their paper, Elena Magán et al. undertake a non-intrusive approach aimed at alerting drowsy drivers without undue alarm. Employing two key methodologies—a neural network model and a fusion of deep learning-based feature extraction with a fuzzy logic system—the study showcases comparable accuracy rates of approximately 65% on training data and 60% on test data. Notably, the fuzzy logic-based system exhibited a specificity of 93%, effectively mitigating false alarms. Analyzing sequences of facial images captured over 60-second intervals, the study employed recurrent and convolutional neural networks alongside deep learning methodologies for feature extraction. While the results signify a promising stride, the study concedes to modest overall accuracy rates, indicating scope for refinement. Thus, while laying a sturdy groundwork for future exploration, the proposed methodologies beckon further enhancement to bolster detection precision.

Shanmanth Guduru et al. [6] embark on a crucial exploration into the domain of driver drowsiness detection, underscoring its paramount significance in ensuring road safety and advocating for the development of robust alert systems to mitigate potential accidents. Their study builds upon prior research endeavors, which have traversed various technological avenues, including parameters such as speed, RPM, and facial expressions, with an escalating interest in harnessing deep learning methodologies for heightened accuracy. Introducing an innovative model for driver drowsiness detection, the paper integrates Transfer Learning [7] and OpenCV techniques, yielding commendable success rates surpassing existing technologies. Through the application of CNN and Inception models on eye datasets, Shanmanth Guduru et al. achieve an impressive accuracy exceeding 87.4%, showcasing the efficacy of their approach. Particularly noteworthy is the model's exceptional capability in discerning between open and closed eyes, indicative of its potential for precise drowsiness detection. Nonetheless, challenges persist in mitigating human error and refining the model's applicability for real-world deployment, necessitating ongoing research and refinement efforts.

M Suriya et al. [8] embark on a significant exploration of driver drowsiness detection, illuminating its pivotal role in road safety and advocating for the imperative need for effective detection systems. Their comprehensive literature survey reveals a landscape of diverse methodologies employed in previous studies, ranging from EEG sensors to CNN-based models, all aimed at monitoring physiological and mental conditions to mitigate accidents. Proposing a novel approach, the research amalgamates EEG sensors with a CNN-based hybrid model, striving to accurately track driver fatigue levels and provide real-time alerts. Through meticulous monitoring of EEG signals and leveraging CNN algorithms, the model adeptly identifies drowsiness and fatigue states, proffering timely alerts to forestall accidents. Results demonstrate the seamless integration of EEG signals and CNN algorithms, thereby achieving high accuracy and efficiency in detecting driver drowsiness, ultimately enhancing road safety. By transmitting signals to the driver's mobile device and activating alerts within the car system, the model effectively notifies drivers of their drowsy or fatigued states, potentially averting accidents. Nonetheless, challenges persist in real-world implementation, necessitating further validation and testing across diverse driving conditions and demographics, highlighting the imperative for ongoing refinement and optimization efforts.

Praveen Tumuluru et al. [9] address the critical issue of driver drowsiness, emphasizing the importance of this problem for road safety and the requirement to put detection systems in cars to minimize accidents caused by drowsy drivers. The research introduces a stacked ensemble model that compares lightweight models, including MobileNet-V2, SqueezeNet [10], and ShuffleNet [11], aiming to achieve efficient drowsiness detection. By utilizing a stacked ensemble approach, the study combines multiple lightweight models to enhance accuracy while reducing computational complexity. The usefulness of the suggested model is demonstrated by experiments on the NTHU-DDD dataset, where the models that were chosen have shorter inference times and reduced computing complexity. The results reveal that the stacked ensemble model performs promisingly in detecting driver drowsiness, balancing high accuracy with minimized computational demands.Nonetheless, the study notes some shortcomings, including the lack of extensive discussion on real-world implementation challenges and the need for further exploration of the model's effectiveness across diverse driving conditions and scenarios.

Saad Arif et al. [12] address the critical issue of drowsy driving and underscore the necessity for effective detection systems to enhance road safety. Their research introduces a passive brain-computer interface (pBCI) scheme that employs EEG signals for spatially localized drowsiness detection during driving tasks. The study highlights the significance of neuro-physiological signatures in differentiating between drowsy and alert states for early detection. By focusing on the spectral signatures of EEG biosignals, the research aims to accurately detect drowsiness through the analysis of spectral band powers and ratios of δ, θ, α, and β rhythms, utilizing machine learning classifiers for classification. The optimal ensemble model achieved the highest results at the F8 electrode position in the right frontal cortex, with an accuracy of 85.6% in the categorization of sleepiness, along with good precision, recall, F1-score, specificity, and area under the ROC curve.Methods involved acquiring EEG signals from multiple brain regions during simulated driving tasks, extracting and analyzing spectral signatures using feature selection methods and machine learning classifiers. This innovative approach underscores the potential of EEG-based pBCI systems in providing early and accurate detection of driver drowsiness, thereby enhancing road safety.

III. Methodology

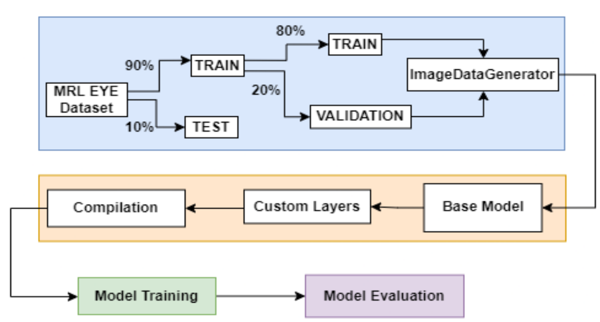

Figure 1: Proposed Workflow of Driver Drowsiness Detection System

Figure 1 describes the setup and design of the proposed detecting system. The dataset is enhanced using a variety of data augmentation techniques, such as rescaling, rotation range, shear range, zoom range, and height shift range, before being fed into the various CNN models. InceptionV3, MobileNetV2, EfficientNetB2 and ensemble model are used for the classification challenge.

A. Dataset Description

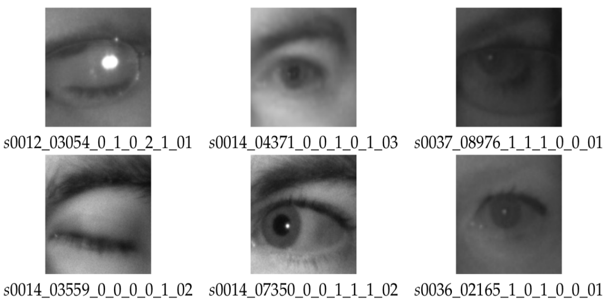

Figure 2: MRL Eye Dataset Samples and Annotations [4]

A vast collection of human eye images, the MRL Eye Dataset [4] is intended for applications such as eye identification, gaze estimation, and eye-blinking frequency analysis within the framework of driver behavior analysis. This collection provides a large-scale library of infrared photos taken with different devices and under varying lighting circumstances, which satisfies the demand for real-world testing data. The publicly available dataset is suitable for testing various features and trainable classifiers because it offers both high- and low-resolution infrared images. Some sample images and annotations from the MRL Eye dataset are shown in Figure 1.

B. Dataset Characteristics

This study makes use of the MRL eye dataset, which contains information from 37 individuals, comprising 33 men and 4 women. With 84,898 photos in all, it provides a significant amount of data for study. The gender of the subject (0 for males, 1 for women) and whether or not they wear spectacles (0 for no, 1 for yes) are annotated on each image. The dataset gives details about the eye state in each image, which is classified as open (1) or closed (0). Three categories are used to annotate reflections on the eyes: none (0), small (1), and big (2). Based on the lighting conditions at the time of capture, images are categorized as either good (1) or bad (0). The file also contains images from three different sensors: the 640 × 480 resolution Intel RealSense RS 300 sensor (sensor ID 01), the 1280 x 1024 resolution IDS Imaging sensor (sensor ID 02), and the 752 x 480 resolution Aptina sensor (sensor ID 03).

C. Dataset Pre-pocessing

To ensure comprehensive coverage,the MRL Eye dataset is split into separate training and testing sets, with 10% of the data set reserve for testing and 90% of the data set reserved for training. This split is based on the eye state in each image, ensuring a balanced and effective evaluation of the model's performance. Within the training set, an additional partitioning strategy allocates 80% for actual model training and 20% for validation, fostering the model's iterative refinement by enabling continuous performance monitoring during training. Augmentation of the training and validation datasets plays a pivotal role in enhancing model robustness and adaptability to real-world scenarios. Leveraging techniques such as rotation, shear, zoom, and shifts in width and height, the augmentation process introduces synthetic variations into the data, mimicking diverse environmental conditions and occlusions that drivers may encounter. This augmented dataset, meticulously curated for training and validation, is seamlessly integrated into the model training pipeline using Keras' ImageDataGenerator, ensuring efficient data loading and processing. Images are uniformly resized to 80x80 pixels, facilitating computational efficiency without compromising on information content. Moreover, to evaluate the model's performance on unseen data, testing data is used to ensure a strict demarcation from the training and validation data. This meticulous data preprocessing pipeline ensures that the model is rigorously trained on diverse, augmented datasets, equipping it with the robustness and adaptability required for accurate driver drowsiness detection in real-world scenarios.

D. IncpetionV3

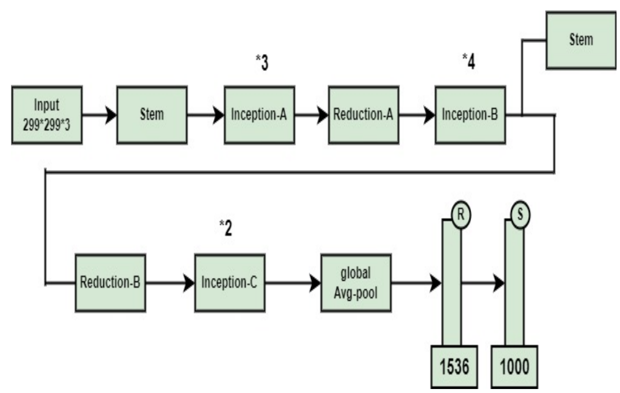

Figure 3: InceptionV3 Architecture [13]

InceptionV3 [13] stands out as a CNN architecture renowned for its remarkable performance in image classification tasks. Developed by researchers at Google, It comprises a total of 48 layers, organized into various blocks, with each block containing inception modules. These inception modules utilize convolutional filters of different sizes within the same layer, allowing the network to capture features across multiple spatial scales effectively. InceptionV3 employs rectified linear unit (ReLU) activation functions throughout the network, facilitating nonlinear transformations essential for learning complex patterns in the data. Figure 2 provides a visual representation of the InceptionV3's architecture. Pretrained on the extensive ImageNet dataset, which encompasses millions of labeled images across numerous categories, InceptionV3 serves as a robust foundation for transfer learning. Its versatility extends beyond image classification, finding applications in diverse tasks such as object detection, semantic segmentation, and medical image analysis. With its depth, parameter efficiency, and stellar performance, InceptionV3 continues to be a cornerstone in the realm of deep learning, empowering researchers and practitioners alike to tackle intricate visual recognition challenges with confidence.

E. MobileNetV2

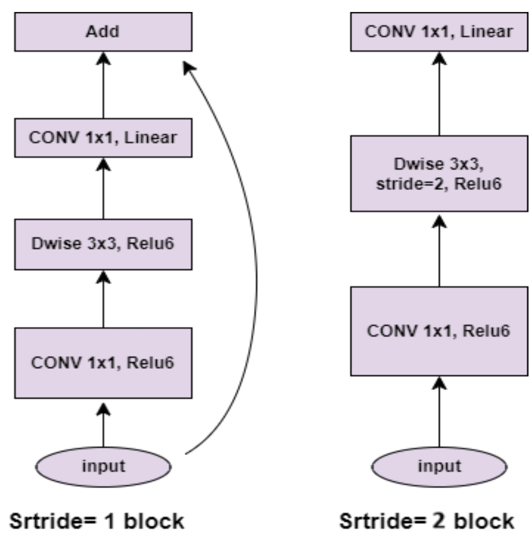

Figure 4: MobileNetV2 Architecture [14]

MobileNetV2 [14] is a state-of-the-art convolutional neural network architecture with a focus on efficiency for mobile and embedded vision applications. It builds upon the success of MobileNetV1 by introducing novel elements like inverted residual blocks. These blocks, typically consisting of around 18-20 layers, leverage depthwise separable convolutions, pointwise linear convolutions, and linear bottlenecks to capture complex features while minimizing computational overhead. Figure 3 contains the architecture of MobileNetV2. MobileNetV2 achieves a remarkable balance between accuracy and efficiency by incorporating techniques such as shortcut connections and these efficient inverted residual blocks. Additionally, it relies primarily on linear and ReLU (Rectified Linear Unit) activation functions throughout the network. These activations enable essential non-linear transformations for the network to learn discriminative features from input images. Its lightweight design, typically with fewer parameters compared to other models, and efficient processing make it particularly well-suited for deployment on resource-constrained devices like smartphones. This allows for the widespread use of deep learning models in real-world applications. Deep learning experts can now create small, powerful vision systems that function well across multiple platforms thanks to MobileNetV2, which is still a mainstay in the area.

F. EfficientNetB2

EfficientNetB2 [15] is part of the EfficientNet family, distinguished for its exceptional performance while maintaining efficiency in terms of parameters and computational demands. Developed by Google researchers, EfficientNetB2 represents an optimization pinnacle, finely balancing model complexity, computational resources, and accuracy. It achieves this through compound scaling, which systematically adjusts the network's depth, width, and resolution. This method enables EfficientNetB2 to efficiently capture intricate image features across varying scales and resolutions. Central to its architecture are advanced techniques such as depthwise separable convolutions, linear bottleneck layers, and squeeze-and-excitation blocks, which optimize the network's efficiency and expressive capacity. These components facilitate substantial model compression without compromising on performance. Furthermore, EfficientNetB2 incorporates a combination of activation functions, including ReLU and linear, strategically placed throughout the network to enable nonlinear transformations essential for learning discriminative features. Pretrained on datasets like ImageNet, EfficientNetB2 serves as a robust starting point for transfer learning, empowering rapid development and deployment of high-accuracy computer vision systems across diverse applications. Its blend of efficiency, accuracy, and versatility solidifies its status as a preferred choice for researchers and practitioners navigating the demands of modern deep learning.

G. Stacking Ensemble Model

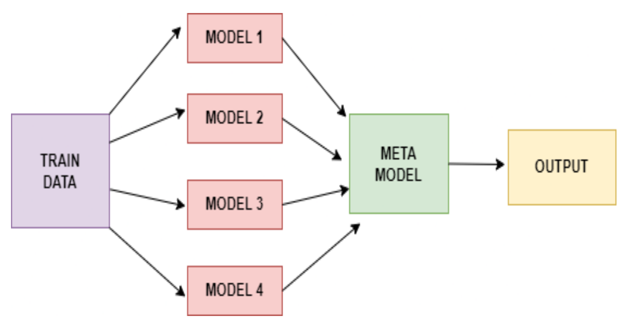

Figure 5: Stacked Ensemble Model [16]

In this project, we employ Stacking Ensemble model [16], an advanced method of ensemble modeling, to harness the collective predictive power of multiple base models, including Convolutional Neural Networks (CNNs). Stacking operates through a multi-layer architecture, where base classifiers, such as CNNs, independently generate predictions on the dataset. These predictions from each base model serve as the input for a meta-model, situated in the subsequent layer, which aggregates them to produce the final prediction. This iterative process involves multiple layers, with each layer utilizing the predictions from the preceding layer as input. Importantly, each layer of the Stacking Ensemble model can consist of different types and configurations of base models, allowing for a high degree of flexibility and adaptability. This adaptability is a key strength of Stacking, as it enables researchers to experiment with various combinations of base models and meta-model to optimize predictive performance. Furthermore, Stacking Ensemble models are not limited to a fixed number of layers; additional layers of meta-models can be added to further refine and enhance predictive accuracy. This flexibility, coupled with the ability to incorporate diverse base models, makes Stacking Ensemble models a highly versatile and effective approach for tackling classification tasks across a wide range of domains.

IV. Experimental Results

In this experimental evaluation, we explored two distinct methodologies aimed at driver drowsiness detection, each leveraging a suite of deep learning architectures. The first approach involved the deployment of standalone pre-trained models, including InceptionV3, EfficientNetB2, and MobileNetV2, each operating autonomously as a feature extractor. These models were seamlessly integrated into training pipeline, characterized by a uniform configuration encompassing optimization techniques such as the Adam optimizer with a categorical cross-entropy loss function. To fortify our models against overfitting and enhance their generalization capabilities, we incorporated a dropout regularization mechanism with a dropout rate of 0.2. This regularization technique, implemented across all standalone architectures, aimed to curb the reliance of the models on specific features and foster robustness in classification tasks.

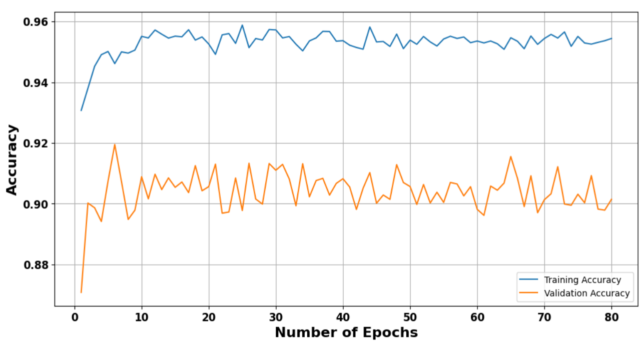

Figure 6: InceptionV3 model accuracy for training and validation

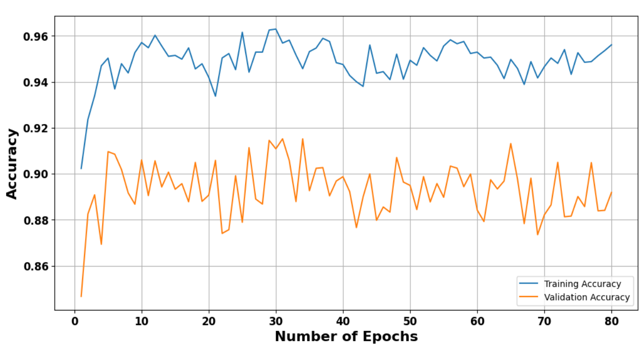

Figure 7: MobileNetV2 model accuracy for training and validation

Transitioning to the second methodology, we delved into the realm of stacked ensemble architectures, an innovative approach that leverages the collective predictive power of multiple models. Here, we amalgamated the outputs of InceptionV3, EfficientNetB2, and MobileNetV2, channeling them into a series of dense layers for classification. This ensemble architecture, while inheriting the optimization framework of its standalone counterparts, also employed a dropout rate of 0.2, ensuring consistency in regularization efforts across all models. In addition, Figures 6, 7, 8 and 9 visually depict the accuracy trends observed during both training and validation phases as the models learn distinctive features crucial for detecting drowsiness. These figures offer valuable insights into how the models perform over the course of training, aiding in the understanding of their learning dynamics and effectiveness.

Moreover, Table 1 presents a detailed overview of the test accuracy achieved by employing various activation functions, providing a comparative analysis of their performance and informing subsequent model selection and refinement strategies. In both methodologies, we ensured consistency in training protocols by employing a fixed number of epochs, with each model trained for 10 epochs and a batch size of 8.

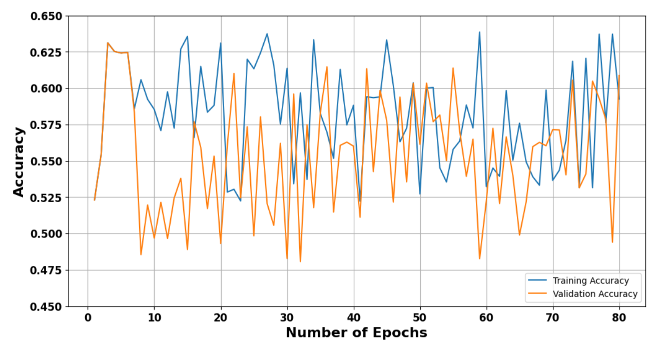

Figure 8: EfficientNetB2 model accuracy for training and validation

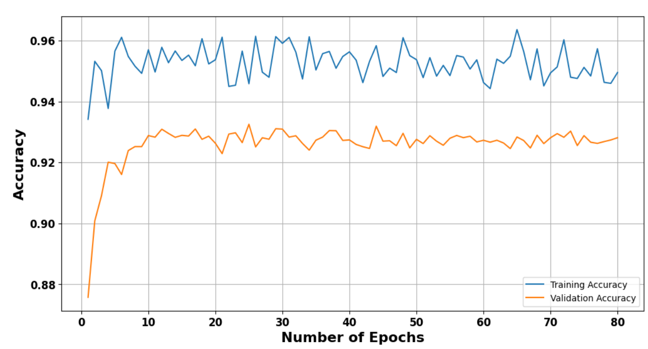

Figure 9: Stacked Ensemble model accuracy for training and validation

TABLE I: TEST ACCURACY REPORT

Model	ReLU	Leaky ReLU	SELU	ReLU6	SiLU	GeLU
InceptionV3	0.9577	0.9495	0.9567	0.9382	0.9411	0.9410
MobileNetV2	0.9458	0.9342	0.9527	0.9418	0.9386	0.9428
EfficientNetB2	0.6231	0.6543	0.6312	0.6253	0.6242	0.6246
Stacked Ensemble	0.9557	0.9510	0.9621	0.9505	0.9494	0.9581

This standardized approach allowed for fair comparison and evaluation across different architectures. Additionally, early stopping and learning rate reduction callbacks were applied universally to promote model convergence and prevent potential overfitting issues, ensuring the stability and reliability of the trained models. Through these standardized practices, we aimed to establish a robust foundation for evaluating the efficacy of different deep learning architectures in driver drowsiness detection tasks. By adhering to common training protocols and incorporating consistent regularization techniques, we facilitated a comprehensive analysis of model performance, thus advancing the frontier of driver safety technology. In this project, accuracy was calculated using categorical accuracy, which measures the proportion of correctly classified samples out of the total number of samples in the dataset.

V. Conclusion

The proposed driver drowsiness detection system aims to discern whether the driver is experiencing drowsiness or remains alert during their time behind the wheel. Through the use of advanced machine learning methods, specifically Convolutional Neural Networks (CNNs), I have demonstrated the capabilities of standalone and stack ensemble models in improving accuracy. CNNs are effective in correctly identifying drowsiness indicators; InceptionV3 came in top at 0.9577%, followed by EfficientNetB2 at 0.6543% and MobileNetV2 at 0.9527%. The investigation into ensemble learning demonstrated the advantages of merging several models once more, yielding an astounding accuracy of 0.9621% with the stacked ensemble model. The shortcomings of standalone models can now be overcome with the help of ensemble approaches, which provide more accuracy and reliability in identifying even the most minute indicators of driver drowsiness. In order to further enhance temporal modeling and context-aware detection, future work will concentrate on augmenting the system's capabilities. This will involve investigating more sophisticated model architectures including recurrent neural networks [17] and attention processes. Furthermore, an attempt will be made to incorporate adaptive learning systems, make use of edge computing for making decisions in real time, guarantee user-centric design for alerts that are not obtrusive, and carry out long-term research to assess the safety and long-term effectiveness.

References

[1] National Safety Council. Drivers are falling asleep behind the wheel, 2020. Accessed: 2024-05-20.

[2] Zhou Dengwen, Shen Xiaoliu, and Weiming Dong. Image zooming using directional cubic convolution interpolation. *IET Image Processing*, 6:627–634, 08 2012.

[3] Francis Quintal Lauzon. An introduction to deep learning. In *2012 11th International Conference on Information Science, Signal Processing and their Applications (ISSPA)*, pages 1438–1439, 2012.

[4] R. Fusek. Pupil localization using geodesic distance. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 11241 LNCS:433–444, 2018.

[5] Elena Magán, M. Paz Sesmero, Juan M. Alonso-Weber, and Araceli Sanchis. Driver drowsiness detection by applying deep learning techniques to sequences of images. *Applied Sciences*, 2022.

[6] Sasi Preetham Ch, Shanmanth Guduru, Vinay Ronagala, Harisudha Kuresan, and Samiappan Dhanalakshmi. Automatic system for driver drowsiness detection system using deep learning. In *2023 International Conference for Advancement in Technology (ICONAT)*, pages 1–4, 2023.

[7] Fuzhen Zhuang, Zhiyuan Qi, Keyu Duan, Dongbo Xi, Yongchun Zhu, Hengshu Zhu, Hui Xiong, and Qing He. A comprehensive survey on transfer learning. *Proceedings of the IEEE*, 109(1):43–76, 2021.

[8] M Suriya, K Preethie, M Amitha, M Sumithra, I Sumaiya Fathima, and M Vishnu. An efficient driver drowsiness detection using deep learning. In *2023 9th International Conference on Advanced Computing and Communication Systems (ICACCS)*, volume 1, pages 2036–2041, 2023.

[9] Praveen Tumuluru, S.Sai Kumar, N. Sunanda, Jaswanth Sai Koduri, Teegela Ayyappa, and Kilaru Balasankar. Sddd: Stacked ensemble model for driver drowsiness detection. In *2023 5th International Conference on Smart Systems and Inventive Technology (ICSSIT)*, pages 1374–1380, 2023.

[10] Forrest N. Iandola, Song Han, Matthew W. Moskewicz, Khalid Ashraf, William J. Dally, and Kurt Keutzer. Squeezenet: Alexnet-level accuracy with 50x fewer parameters and ¡0.5mb model size, 2016.

[11] Xiangyu Zhang, Xinyu Zhou, Mengxiao Lin, and Jian Sun. Shufflenet: An extremely efficient convolutional neural network for mobile devices, 2017.

[12] Saad Arif, Saba Munawar, and Hashim Ali. Driving drowsiness detection using spectral signatures of eeg-based neurophysiology. 14:1153268, 03 2023.

[13] Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jon Shlens, and Zbigniew Wojna. Rethinking the inception architecture for computer vision. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2818–2826, 2016.

[14] Mark Sandler, Andrew Howard, Menglong Zhu, Andrey Zhmoginov, and Liang-Chieh Chen. Mobilenetv2: Inverted residuals and linear bottlenecks. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4510–4520, 2018.

[15] Mingxing Tan and Quoc Le. Efficientnet: Rethinking model scaling for convolutional neural networks. 05 2019.

[16] Stamatios-Aggelos Alexandropoulos, Christos Aridas, Sotiris Kotsiantis, and Michael Vrahatis. *Stacking Strong Ensembles of Classifiers*, pages 545–556. 05 2019.

[17] Robin M. Schmidt. Recurrent neural networks (rnns): A gentle introduction and overview, 2019.