Grenze International Journal of Engineering and Technology, June Issue

Steering Towards Safety: Approach of CNN Architectures to Predict Distracted Driving

Aaditi Ghodke[1], Girija Giri[2], Manvi Ankalgi[3], Lavanya Kamble[4], Yogesh Deshpande[5] and Pawan Wawage[6]
[1-6]Vishwakarma Institute of Information Technology/Department of Information Technology, Pune, India
Email: aaditi.22210978@viit.ac.in, girija.22210608@viit.ac.in, manvi.22210423@viit.ac.in, lavanya.22211505@viit.ac.in, yogesh.deshpande@viit.ac.in, pawan.wawage@vit.edu

***Abstract*— Distracted Driving is becoming the critical problem day by day causing accidents, fatalities and injuries worldwide. With the demand of automated intelligent transportation system, Artificial Intelligence (AI) with computer vision technologies will assure the driver safety with the help of automated monitoring system. The research addresses deep learning model implementation along with evaluation for detecting distracted driving incidents. The research investigates how CNN (Convolutional Neural Network) architectures including MobileNetV2, EfficientNet-B0, ResNet-152 and InceptionV3 function to detect driving behaviors using the State Farm Distracted Driver Detection Dataset with thousands of images. The models used are transfer learning methods along with image preprocessing to conduct their training that led to evaluations concerning classification accuracy, loss reduction and confusion matrix analysis. The MobileNetV2 and EfficientNet-B0 models outperformed with the 99.74% and 99.71% validation accuracies respectively. This makes it suitable for embedded system real-time deployments. The research shows how lightweight CNN models achieve successful performance in real-time driver monitoring applications which reduce distraction-caused accidents on the roads. Overall, the technology helps in detecting the distraction in the driving and warns the drivers about not paying the attention.**

***Keywords*— Distracted driver, CNN, MobileNetV2, EfficientNet-B0, Resnet-152, InceptionV3.**

I. INTRODUCTION

Distracted driver is one of the important causes behind increasing road accidents. Any activity that distracts driver from his main task of driving refers to distracted driving. It results into reduced reaction time, less situational awareness, and lost control over the vehicle which increases the risk of accidents. According to Road Accidents in India – 2022 report given by the Ministry of Road Transport and Highways India, 7558 accidents were reported in 2022 due to mobile phone usage during driving [1]. Mobile phone usage creates mental strain on drivers which makes them lose focus from their roadway responsibilities while any hands-free method of phone usage causes "inattentional blindness." Road users who are not in vehicle control positions such as pedestrians together with riders of two-wheelers experience higher levels of danger.
Also, the Infotainment systems present in car can cause 67 major hazards for driver safety by their visual, manual and cognitive demands as drivers use touchscreens and complex menus and multimedia systems. The study was done to examine driver adjustments when operating infotainment systems during dangerous traffic situations.

Grenze ID: 01.GIJET.11.2.519_26

Participants operating infotainment systems in a BMW simulator slowed down and reduced their tasks when encountering hazardous situations according to researcher findings. [2]. Another research investigated the way drivers perform their actions and gaze at their surroundings together with how they distribute their attention undergoes modifications. The study inferred that near crashes were not primarily caused by infotainment system usage but drivers using these systems exhibited changed glance behavior which may adversely affect their driving awareness during certain driving conditions. [3].

Recent trends in road accidents that stem from driver distraction drive the immediate necessity for automated driver monitoring systems. This need can be fulfilled with the help of deep learning algorithms through visual monitoring of driver. Hence, the purpose of this comparative study involves evaluating different architectures of CNN namely MobileNetV2, EfficientNet-B0, ResNet-152 and InceptionV3 for detecting distracting activities of driver behavior through images. The research makes use of transfer learning techniques through existing ImageNet weights that undergo adjustment with distracted driver dataset. The evaluation of algorithms is done on confusion matrix and accuracy of training. The real-time operability of the system is measured through the ability of system to detect the activity of driver correctly. This research study will reveal which architectural design best fits operational road driving applications and which model works most efficiently to identify different driver diversions.

II. Literature Survey

It is important to analyze the previous work done in the field of distracted driver detection using machine learning techniques. This section gives insights into the previous work done.

One of the studies proposed distracted driving detection system utilizes Internet of Things sensors by combining images from a camera's interior-facing view and IMU sensor motion data. In this work, the system utilized Inception-V3 to process images, as well as the Bi-LSTM for sensor signal data, using a Bayesian Network, achieved 87.02% accuracy and provided the digital image denoising for privacy, while still offering a real-time Android operation and modular that allows for the array of connected sensors [4]. Another research is about Random Forest and SVM algorithms together with kNN and Naive Bayes algorithms to detect distracted driving incidents through hierarchical clustering feature selection procedures and permutation importance evaluation methods. The combination of three data modalities generated maximum 91.5% academic accuracy for Random Forest models [5].

A study focused on producing real-time CNN-based systems for embedded systems which detect driving distractions. High model accuracies came at a cost of significant computational expenses when using models like VGG-16 and ResNet. The model ResNet reached a detection accuracy of 92% whereas GoogleNet delivered 89% accuracy and performed at 11Hz. The distracted driving detection system executed real-time operation on Jetson TX1 and NanoPi M3 platforms. The system included a voice alert functionality which gave immediate feedback to drivers during operation [6]. The study proposed the idea that reduces the distraction because of complex infotainment touchscreens system with voice-based Head-up displays and customizable buttons. But the system is not for real-time use [7].

In one of the papers, researchers performed comparative study on CNN, VGG-16, RestNet50 and ensemble of CNN to classify the distracted driver into different categories. Among these models, CNN got 99.5% accuracy, ensemble models performed well with F1 value of 0.724. The dataset used has less diversity and there is chance of overfitting of classes [8]. One of the studies estimated the driver's point of gaze with two regular cameras at face and scene. They used A DPEN (Drivers' Points-of-Gaze Estimation Network) model for training with DPoG dataset. The solution demonstrates high accuracy (AUC 0.97) while being low-cost with non-invasive capabilities yet lacks real-time analysis and diverse driving data capabilities [9].

The researchers created deep learning model based on EfficientDet for real-time detection of distracted drivers through dashboard camera images with EfficientDet-D3 having accuracy of 99.16%. But the study omitted multi-modality input and has limited generalization because of missing real-world validation [10]. In one of the papers, a Bayesian Network received training through simulator data of 92 drivers to predict distracted driving and its optimization used a Genetic Algorithm with 10-fold cross-validation for validation. The system operated with synthetic data while delivering average results for a restricted set of users without integrating multiple input methods [11]. In one of the research projects, a HCF i.e. Hybrid CNN Framework is developed in which the authors have made the hybrid model of Resnet50, Inception V3 and Xcpetion models. They have extracted the features from each model and concatenated them into one vector to predict results. HCF achieved 96.74% accuracy which is better than any single model [12]. In one of the papers, the authors optimized the DNN i.e. Deep Neural Network with GWO (Grey Wolf) algorithm to identify driving patterns. The optimized DNN achieved a great accuracy of 95.13% and a F1 score of 0.9513. The model got better results as compared to the traditional machine learning algorithms such as Random Forest and Support Vector Machine [13].

In the next research studied, the authors proposed an E2DR i.e. Ensemble-Based Distraction Detection with Recommendations, which is an ensemble model designed to detect the distracted drivers. In this research project, the best performing combination was Resnet50 with VGG16 with a test accuracy of 92% [14]. The HSDDD (Hybrid Scheme for the Detection of Distracted Driving) is proposed in the other research, which combined the four models AlexNet, Inception V3, ResNet50, and VGG-16 using Histogram of Oriented Gradients. It achieved the accuracy upto 99.5% [15]. In the other study, the researchers found that to detect the driver distraction postures, the combined model of pre-trained CNN along with Bi-directional LSTM achieved a great result of 93.1% of F1 score and 91.7% of accuracy.

The models effectively taken the spatial features (via CNN) and temporal/spectral features (via BiLSTM) [16]. Past studies used vision or sensor data for driver distraction detection, but most covered few actions or used simulators. Multimodal approaches improved accuracy. The study combined real-world image and sensor data with CNN and LSTM to detect 10 actions effectively [17].One of the types of research focused on detecting driver cognitive distraction using driving performance measures obtained from the CAN-Bus, avoiding the need for external sensors. Support Vector Machine (SVM) models were developed and showed around 74% accuracy, making the approach practical and real-time. However, the system had lower sensitivity and was tested only in a simulated environment [18]. In one of the studies, SafeDrive introduced a wrist-worn device system combined with smartphones to detect distracted driving using real-time gesture recognition and a semi-supervised KNN-SVM model. It achieves over 90% accuracy but is limited to right-hand traffic scenarios. [19].

III. PROPOSED SYSTEM

The study made use of the State Farm Distracted Driver Detection Dataset, hosted on Kaggle [20]. The dataset consists of 22,424 images taken from a camera mounted on the dashboard of a car, all featuring 28 different drivers. Each image is labelled with one of 10 distinct labels pertaining to driving activity.

A. Data Pre-processing

The dataset has separate train and test folders. The train folder contains labelled images, and the test folder contains unlabelled images, which are only for evaluation. To aid in validating the model and to avoid overfitting, the training data was split into training and validation parts, utilizing custom code. After that, all images were resized to a fixed size so they could be inputted into the CNN architectures. All pixel values were normalized with respect to all training images so the model could converge faster during training.

B. Model Used

MobileNetV2 exists as a lightweight architectural design meant for mobile and embedded applications. The combination of inverted residual blocks together with depth wise separable convolutions produces MobileNetV2 with more than ten times fewer parameters than regular CNNs and nearly identical accuracy results. MobileNetV2 functions perfectly in online systems that need to balance processing speed with accuracy performance.

EfficientNet researchers established that a compound scaling strategy to model scaling dimensions of depth width resolution provided the optimal solution to scaled accuracy problems. The EfficientNet-B0 provides a basic example of EfficientNet architecture which the other variations including EfficientNet-B1 and B2 derive from. EfficientNet-B0 became the selected baseline architecture since it provided both high performance accuracy and optimized resource usage. The EfficientNet architecture features a compact model design which attains high accuracy levels to support applications with limited resources.

ResNet represents a deep residual network with 152 layers that solves training difficulties of deep networks through layer-to-layer residual connections to prevent the vanishing gradient issue. The implementation of skip connections enables ResNet to perform identity mappings across layers that allows the training of highly deep models. ResNet-152 demonstrates sufficient capabilities to detect elaborate visual details while requiring extra computational processing.

Through its Inception modules InceptionV3 implements filters in parallel to achieve a multi-scale functionality. The Inception modules combine factorized convolutions and auxiliary classifiers and batch normalization techniques to reduce the neural network complexity effectively. Many different variants of InceptionV3 models currently exist because of its prevalent use in image classification benchmarks. Different models of Inception V3 offer varying levels of depth, width and computational cost.

C. Model Training

After completing the pre-processing steps, four different CNN architectures namely MobileNetV2, EfficientNet-B0, ResNet-152, and InceptionV3 were trained, by utilizing transfer learning and using the pre-trained weights from ImageNet. During this training process, the training vs. validation loss and accuracy curves were monitored to evaluate the model performance and convergence behaviour. Once the models were trained, the performance of the models were assessed on the unlabelled test set. The predictions were produced from the models and dutifully evaluated the overall accuracy and confusion matrix to understand class-wise prediction effectiveness.

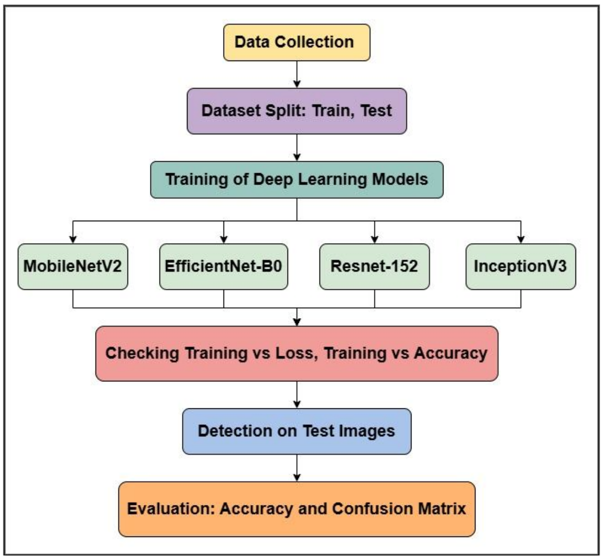


Fig.1. Flow Diagram of the System

The initial step requires retrieval of the State Farm Distracted Driver Detection Dataset data. The data collection includes images that show drivers while engaging in different distracted actions with specified labels. The gathered data gets separated into two essential splits namely training and test directories. The training process requires the training folder to split into training and validation sets through coding to achieve proper model validation. Evaluating model learning in real-time occurs in this step to prevent overfitting and achieve generalization.

The next step involves deep learning model training after preparation of data. The set of prepared data is processed using MobileNetV2 alongside EfficientNet-B0 and ResNet-152 and InceptionV3 as pretrained models. The testing phase performs continuous performance evaluation based on the comparison between training loss and validation loss values as well as training accuracy against validation accuracy. The trained models get deployed for test images detection to determine their accuracy with unobserved data. The final stage involves performing an evaluation through accuracy score assessment alongside confusion matrix usage to determine model performance regarding distracted driving behaviour identification.

IV. Results

In this section, the total of four CNN models are evaluated which includes MobilenetV2, Efficient-B0, Resnet 152 and InceptionV3 on the State Farm Distracted Driver Detection dataset. Each model is tested based on training accuracy, validation accuracy and loss convergence and confusion matrix.

A. Experimental Setup and Dataset

All the models are trained on Google Colab which provides a T4 Tesla GPU (Graphics Processing Unit) as a hardware accelerator with Compute Unified Device Architecture (CUDA) Version 12.2. It has 16 GB GDDR6 (Gigabytes Graphics Double Data Rate type 6) Synchronous Dynamic Random-Access Memory.

The State Farm Distracted Driver Detection dataset is taken from Kaggle, which is a well-known platform providing the machine learning based datasets. The dataset has around 22,424 total images. These images are further split into the training and testing data. And classes in the dataset includes c0 (Safe driving), c1(Texting – right), c2(Talking on the phone - right), c3(Texting - left), c4(Talking on the phone - left), c5(Operating the radio), c6(Drinking), c7(Reaching behind), c8(Hair and makeup), c9(Talking to passenger).

MobilenetV2 Model

The model outperformed with a great performance in only 4 epochs, in which it achieved the training accuracy of 99.95% and validation accuracy of 99.74%. First of all the overall loss by the models was 133.164, but as the fourth cycle completed it got reduced to only 1.383.

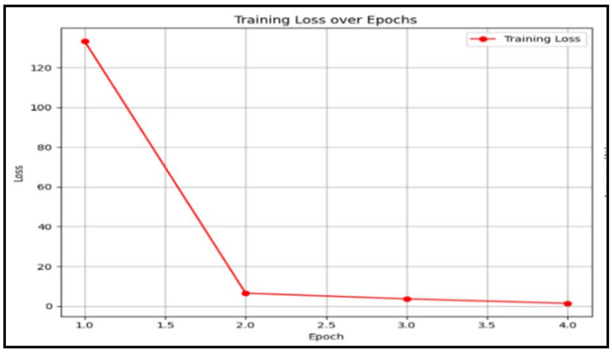


Fig.2. Graph of Training Loss versus Epochs

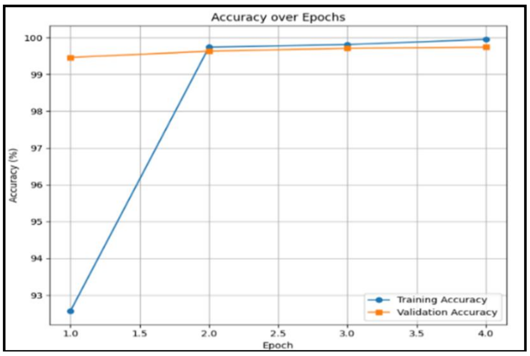


Fig.3. Graph of Accuracy versus Epochs

Fig.4 shows the confusion matrix of the model. In this, it is easily understandable that for most of the classes the model it is predicting well. But the most accurate prediction was for the class 'Safe driving' with 400 accurate predictions. However, the model has several misclassifications for the classes such as 'texting-left', talking on the phone-left'.

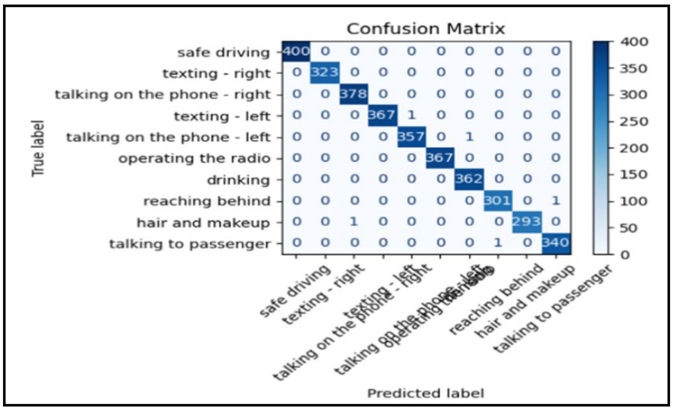


Fig.4. Confusion Matrix of the Mobilenet_v2 Model

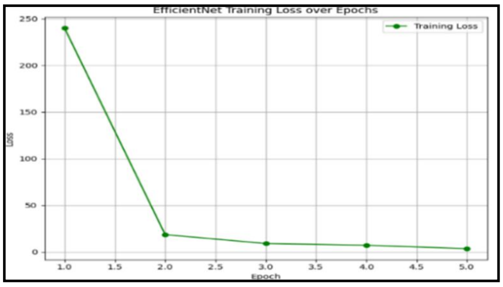
The Fig.5 and Fig.6 shows the correct predictions of the distracted drivers with classes including 'talking on the phone – right', 'texting - right'.

Fig.5. Result Image of Detection

Fig.6. Result Image of Detection

B. Efficientnet-B0 Model

The model performed well with only 5 epochs, in which it achieved the training accuracy of 99.79% and validation accuracy of 99.71%. First of all, the overall loss by the models was 239.829, but as the fifth cycle completed it got reduced to only 3.566.



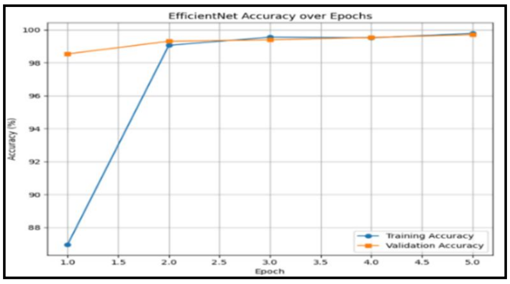


Fig.7. Graph of Training Loss versus Epochs

Fig.8. Graph of Accuracy versus Epochs

Fig.9 shows the confusion matrix of the model. In this, it is easily understandable that for most of the classes the model it is predicting well. But the most accurate prediction was for the class ‘texting-left’ with 389 accurate predictions. However, the model has several misclassifications for the classes such as ‘Safe-driving’, ‘talking on the phone-right'.

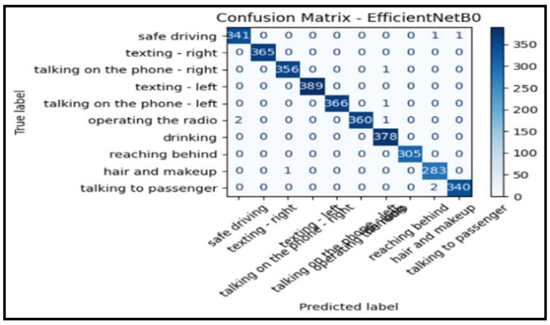


Fig.9. Confusion Matrix of Efficientnet_b0 model

The Fig.10 and Fig.11 shows the correct predictions of the distracted drivers with classes including ‘drinking', ‘talking to passenger’, etc.

Fig.10. Result Image of Detection

Fig.11. Result Image of Detection

C. Resnet 152 Model

The model outperformed with a great performance in only 4 epochs, in which it achieved the training accuracy of 99.63% and validation accuracy of 99.54%. First of all the overall loss by the models was 75.759, but as the fourth cycle completed it got reduced to only 5.188.

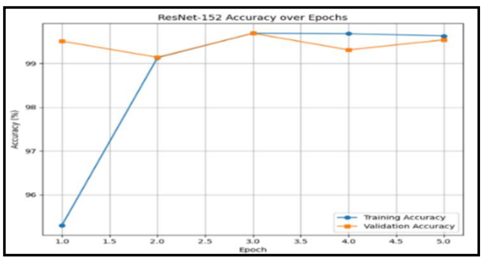


Fig.12. Graph of Training Loss versus Epochs

Fig.13. Graph of Accuracy versus Epochs

Fig.14 shows the confusion matrix of the model. In this, it is easily understandable that for most of the classes the model it is predicting well. But the most accurate prediction was for the class ‘drinking’ with 398 accurate predictions. However, the model has several misclassifications for the classes such as ‘Safe-driving’, ‘texting-right', etc.

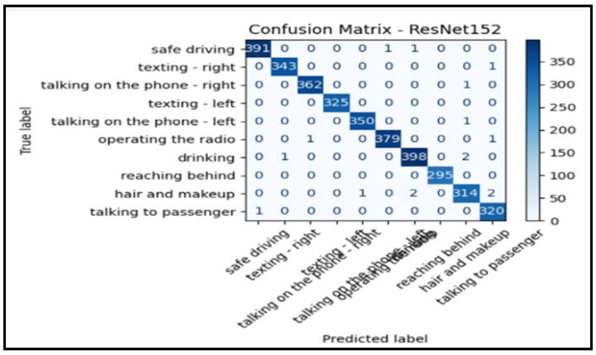


Fig.14. Confusion Matrix of Resnet 152 model

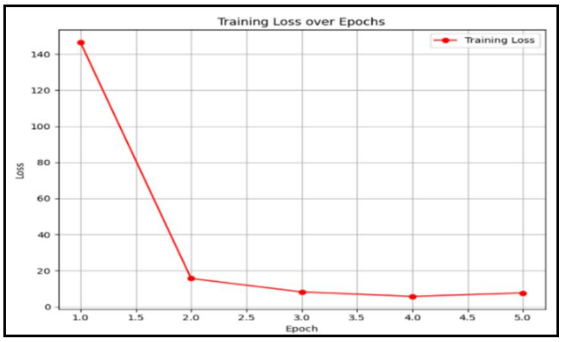
The Fig.15 and Fig.16 shows the correct predictions of the distracted drivers with classes including ‘operating the radio', ‘hair and makeup’.

Fig.15. Result Image of Detection

Fig.16. Result Image of Detection

D. InceptionV3 Model

The model outperformed with a great performance in only 4 epochs, in which it achieved the training accuracy of 99.79% and validation accuracy of 99.54%. First of all the overall loss by the models was 146.440, but as the fourth cycle completed it got reduced to only 5.746.



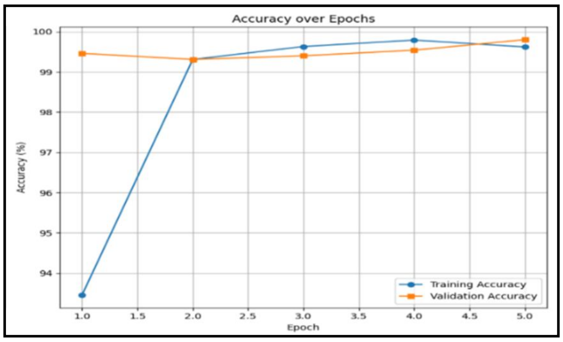


Fig.17. Graph of Training Loss versus Epochs

Fig.18. Graph of Accuracy versus Epochs

Fig.19 shows the confusion matrix of the model. In this, it is easily understandable that for most of the classes the model is predicting well. But the most accurate prediction was for the class ‘drinking’ with 253 accurate predictions. However, this model got most of the misclassifications in which the biggest number of misclassifications are for the class ‘talking on the phone-left'’, followed by the class ‘talking to passenger’ with total 284 misclassifications.

The Fig.20 shows the correct predictions of the distracted driver with class ‘drinking’.

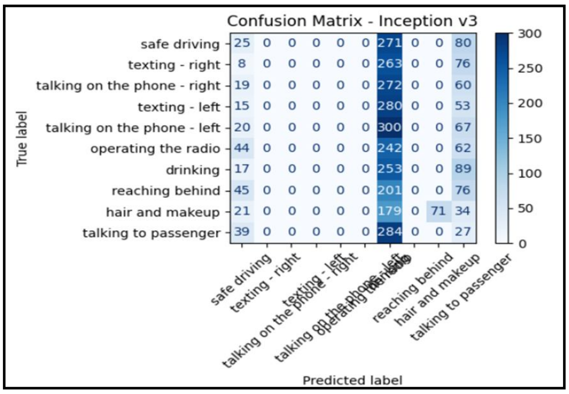


Fig.19. Confusion Matrix of Inception_v3

Fig.20. Result Image of Detection

TABLE I: COMPARISON OF MODELS

Model	Training Accuracy	Validation Accuracy	Loss at First Epoch	Loss at Last Epoch
Mobilenet_v2	99.95%	99.74%	133.6	1.38
Efficientnet_b0	99.79%	99.71%	239.8	3.56
Resnet 152	99.63%	99.54%	75.7	5.18
Inception_v3	99.79%	99.54%	146.4	5.74

V. CONCLUSION

A comparative evaluation is provided within this research about four CNN architectures including MobileNetV2 and EfficientNet-B0 alongside ResNet-152 and InceptionV3 to detect different driver distractions using images. The models reached high levels of training accuracy and validation accuracy while MobileNetV2 offered an optimal combination of performance effectiveness. The study confirms deep learning models operate successfully for detecting real-time distracted driving behavior when utilizers optimize their models for optimal results. Subsequent research needs to achieve these following advancements: Using both image data and sensor measurements from accelerometers and gyroscopes and heart rate sensors strengthens prediction accuracy and enhances environment awareness. Real-time simulation testing of models must occur on-road among various weather and lighting conditions to enhance their operational usefulness. The precision of predictions should boost through personalized driver models which adapt to individual user drive habits during operation.

REFERENCES

[1] Road Accidents in India | Ministry of Road Transport & Highways, Government of India,” Morth.nic.in, 2016. https://morth.nic.in/road-accident-in-india

[2] Platten, F., Milicic, N., Schwalm, M., & Krems, J. (2013). Using an infotainment system while driving–A continuous analysis of behavior adaptations. Transportation research part F: traffic psychology and behaviour, 21, 103-112.

[3] Perez, M. A. (2012). Safety implications of infotainment system use in naturalistic driving. Work, 41(S1), 4200-4204.

[4] Streiffer, C., Raghavendra, R., Benson, T., & Srivatsa, M. (2017, December). Darnet: A deep learning solution for distracted driving detection. In Proceedings of the 18th acm/ifip/usenix middleware conference: Industrial track (pp. 22-28).

[5] Misra, A., Samuel, S., Cao, S., & Shariatmadari, K. (2023). Detection of driver cognitive distraction using machine learning methods. IEEE Access, 11, 18000-18012.

[6] Tran, D., Manh Do, H., Sheng, W., Bai, H., & Chowdhary, G. (2018). Real-time detection of distracted driving based on deep learning. IET Intelligent Transport Systems, 12(10), 1210-1219.

[7] Nejintev, N., Pascal, A., & Istrati, D. (2023). Reducing driver distraction created by the infotainment system. In Conferința tehnico-științifică a studenților, masteranzilor și doctoranzilor (Vol. 2, pp. 53-56).

[8] Darapaneni, N., Arora, J., Hazra, M., Vig, N., Gandhi, S. S., Gupta, S., & Paduri, A. R. (2022). Detection of distracted driver using convolution neural network. arXiv preprint arXiv:2204.03371.

[9] Nguyen, D. V. T., Tran, A., Vu, H. N., Pham, C., & Hoai, M. (2024). Driver Attention Tracking and Analysis. arXiv preprint arXiv:2404.07122.

[10] Sajid, F., Javed, A. R., Basharat, A., Kryvinska, N., Afzal, A., & Rizwan, M. (2021). An efficient deep learning framework for distracted driver detection. IEEE Access, 9, 169270-169280.

[11] Ahangari, S., Jeihani, M., & Dehzangi, A. (2019, August). A machine learning distracted driving prediction model. In Proceedings of the 3rd International Conference on Vision, Image and Signal Processing (pp. 1-6).
[12] Huang, C., Wang, X., Cao, J., Wang, S., & Zhang, Y. (2020). HCF: A hybrid CNN framework for behavior detection of distracted drivers. IEEE access, 8, 109335-109349.
[13] Fu, B., Shang, Q., Sun, T., & Jia, S. (2023). A distracted driving detection model based on driving performance. IEEE Access, 11, 26624-26636.
[14] Aljasim, M., & Kashef, R. (2022). E2DR: A deep learning ensemble-based driver distraction detection with recommendations model. Sensors, 22(5), 1858.
[15] Alkinani, M. H., Khan, W. Z., Arshad, Q., & Raza, M. (2022). HSDDD: A hybrid scheme for the detection of distracted driving through fusion of deep learning and handcrafted features. Sensors, 22(5), 1864.
[16] Mafeni Mase, J., Chapman, P., Figueredo, G. P., & Torres Torres, M. (2020). Benchmarking deep learning models for driver distraction detection. In Machine Learning, Optimization, and Data Science: 6th International Conference, LOD 2020, Siena, Italy, July 19–23, 2020, Revised Selected Papers, Part II 6 (pp. 103-117). Springer International Publishing.
[17] Omerustaoglu, F., Sakar, C. O., & Kar, G. (2020). Distracted driver detection by combining in-vehicle and image data using deep learning. Applied Soft Computing, 96, 106657.
[18] Jin, L., Niu, Q., Hou, H., Xian, H., Wang, Y., & Shi, D. (2012). Driver cognitive distraction detection using driving performance measures. Discrete Dynamics in Nature and Society, 2012(1), 432634.
[19] Jiang, L., Lin, X., Liu, X., Bi, C., & Xing, G. (2018). Safedrive: Detecting distracted driving behaviors using wrist-worn devices. Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, 1(4), 1-22.
[20] Seok Jeong-ro, "State Farm Distracted Driver Detection," Kaggle.com, 2020. https://www.kaggle.com/datasets/rightway11/state-farm-distracted-driver-detection (accessed Apr. 24, 2025).