Intelligent monitoring system for analyzing vehicle drivers state based on adaptive deep learning models

Nickolay Rudnichenko[1], Vladimir Vychuzhanin[1], Tetiana Otradskya[1] and Denys Shvedov[1]

[1] Odessa Polytechnic National University, Shevchenko Avenue 1, Odessa, 65001, Ukraine

Abstract

This paper focuses on the development of an intelligent driver monitoring system based on adaptive deep learning models to enhance road safety. The research explores advanced deep learning techniques, particularly convolutional neural networks and their modifications, such as ResNet50 and MobileNetV2. Special attention is given to the stages of data preprocessing, augmentation, training and testing dataset formation, as well as model training and fine-tuning. A conceptual framework and architecture for an intelligent driver monitoring system have been developed, incorporating two modules based on different deep learning models. An experimental study was conducted to compare the performance of various convolutional neural network (CNN) architectures, including classical CNN, ResNet50, MobileNetV2, EfficientNetB0, and VGG16, in detecting driver fatigue and drowsiness. Signs of overfitting were identified in the ResNet50 and MobileNetV2 models when applied to the selected datasets, highlighting the need for further hyperparameter optimization. The developed testing scripts enable real-time analysis of behavioral indicators of drowsiness and driver distraction. The proposed system is designed for non-invasive and high-precision real-time monitoring of driver conditions, including fatigue, drowsiness, and distraction detection. The findings confirm the effectiveness of adaptive deep learning models for driver state monitoring. The developed system demonstrates the capability to detect signs of fatigue, drowsiness, and distraction, which may help reduce the likelihood of road accidents. Experimental results indicate that the choice of an optimal neural network architecture depends on the specific task requirements and the available computational resources.

Keywords

deep learning, data analysis, intelligent monitoring systems, vehicle drivers state

1. Introduction

The advancement of modern technologies and the increasing computational power make intelligent big data analysis systems essential tools for automating complex processes and making well-founded decisions [1]. The application of intelligent technologies and methods enables the identification of intricate patterns, resource optimization, and enhanced prediction accuracy across various scientific and industrial domains [2]. The growing volume of data necessitates efficient algorithms for processing, interpreting, and utilizing information in real time, emphasizing the significance of developing advanced analytical models. Intelligent data analysis systems contribute to the autonomy and adaptability of technological solutions, ensuring their reliability, efficiency, and security [3].

So, in the modern context, road traffic safety is becoming an increasingly pressing issue, necessitating the implementation of innovative methodologies and technological solutions aimed at minimizing the likelihood of traffic accidents and enhancing driver protection. A crucial aspect of this issue is the physiological and psychological state of the driver, including their level of concentration, degree of fatigue, emotional stability, and ability to respond promptly to changes in road conditions. Consequently, the study and development of highly effective driver state monitoring algorithms have become priority areas in the field of transportation safety.

Traditional driver monitoring approaches based on physiological parameters such as heart rate and galvanic skin response have significant limitations. Their implementation in real-world operational conditions is associated with technical challenges, the need for specialized equipment, and potential discomfort for the driver. In this context, non-invasive monitoring based on video

✉ nickolay.rud@gmail.com (N. Rudnichenko); icst_nuop@ukr.net (V.Vychuzhanin); tv_61@ukr.net (T. Otradskya); studylearnerstudy@gmail.com (D. Shvedov);

⬭ 0000-0002-7343-8076 (N. Rudnichenko); 0000-0002-6302-1832 (V.Vychuzhanin); 0000-0002-5808-5647 (T. Otradskya); 0009-0002-4823-8782 (D. Shvedov)

stream analysis presents a compelling alternative. This approach enables the assessment of driver states by examining visual indicators, including facial expressions, head position, blink frequency and patterns, as well as other markers of fatigue and decreased attention [4].

With advancements in artificial intelligence (AI) and data-driven analysis, the accuracy and reliability of automatic driver state detection have significantly improved. A key role in this progress is played by machine learning (ML) and deep learning (DL) techniques, particularly deep neural networks (DNN), which have demonstrated outstanding performance in computer vision and behavioral pattern recognition. DL enables models to autonomously extract meaningful features from large datasets, eliminating the need for manual feature engineering. State-of-the-art architectures, such as convolutional neural networks (CNNs), recurrent neural networks (RNNs), and transformers, ensure efficient real-time video stream processing, allowing for accurate and timely detection of potentially hazardous driver states [1,3,5].

Thus, the development of driver state assessment approaches based on video analysis using DL techniques represents a promising direction in transportation safety. Intelligent monitoring systems built upon these technologies can promptly respond to changes in driver conditions, mitigating the risk of accidents. Their integration into modern vehicles has the potential to significantly enhance overall road traffic safety.

2. Description of Problem in Literature Review

According to the analysis of a number of literary sources and the opinion of authoritative authors, in practice there are various methods for determining the driver's condition, the most priority and promising of which are based on: wearable sensors, processing the driver's visual conditions and the acoustic environment.

2.1. Methods based on processing biometric information and classic hardware sensors

One of the first and main areas of focus for many researchers and organizations, including automobile companies, is the development of sensors for collecting biometric information. Biometric information about a driver allows us to understand his condition and ability to drive a vehicle. Biometric information includes information such as electrocardiogram, electrodermal activity, blood pressure levels and visceral fat levels, as well as exercise levels, sleep patterns and diet. An important factor is also the correct interpretation of all the above parameters [6].

For a example, the authors of paper [7] conduct a study demonstrating the significant effectiveness of electroencephalography data in monitoring driver states, particularly in detecting drowsiness and loss of attention. To achieve this, they developed a system comprising an EEG recording device, a computational unit capable of signal processing and classification, and a real-time feedback mechanism that alerts the driver and wakes them up by emitting an audio signal. Drawing upon the analysis of the authors' perspectives existing classical methods of measuring heart rate limit or interfere with driver performance. In addition to the completeness and accuracy of measurements, it is very important that the driver monitoring system does not limit or interfere with the driver's performance [8]. Therefore, traditional methods are not suitable for measuring heart rate in a vehicle, and a non-wearable monitoring system is desirable, although the reliability of the data obtained is inferior to that of wearable systems. Such driver monitoring systems should be able to correctly determine the driver's state of readiness without limiting his or her movement [9].

It is worth noting that driver state monitoring using MEMS (Micro-Electro-Mechanical Systems) sensors represents an innovative approach to enhancing road safety. MEMS sensors are characterized by their small size, high sensitivity, and precision, making them ideal for integration into driver monitoring systems. According to [10-12], MEMS continuously collect data on the driver's physiological parameters and movements. The gathered information is processed using machine learning algorithms to detect anomalies or patterns indicative of potential danger. For instance, the system can identify patterns associated with drowsiness or driver distraction. If a potential risk is detected, the system can issue auditory or visual alerts, as well as haptic warnings via seat or steering wheel vibrations. Some advanced systems may also implement active safety measures, such as engaging autopilot functions or initiating an automatic vehicle stop if the driver

fails to respond to warnings [13]. A key aspect of all the reviewed scientific studies is the complexity of their technical reproducibility due to the necessity of multiple integrations and the non-trivial process of configuring operational modes of technical devices, combined with the consideration of individual characteristics and predispositions of specific drivers. However, collectively, the results obtained by the authors indicate the promising potential of MEMS sensors for driver state monitoring.

2.2. Methods based on processing the driver's visual state

Modern DNN generally outperform traditional methods in accuracy and automation. However, they require large datasets, have limited interpretability, and demand high computational resources [14, 15]. These challenges drive the development of hybrid models that combine the strengths of traditional approaches with deep learning techniques. The increasing prevalence of in-vehicle information systems significantly impacts road safety, as their use contributes to visual, manual, and cognitive driver distraction, potentially impairing driving performance. Additionally, drivers frequently engage in secondary activities such as eating, drinking, adjusting the radio, and using mobile devices. These distractions reduce their focus on the road and increase cognitive load, thereby heightening the risk of traffic accidents. One effective method for detecting driver distraction involves analyzing facial orientation and gaze direction. Most modern driver monitoring systems follow a multi-step approach [16]:

1. Face recognition and head tracking – initially, a face detection algorithm is applied, and its results serve as input for a more precise head-tracking system.
2. Facial landmark localization – this step involves identifying key facial features such as the eyes, enabling anthropometric analysis of both the face and head.

One of the most widely used face recognition algorithms is the Viola-Jones method, which has inspired several enhanced versions, such as PICO [17]. This approach refines the standard Viola-Jones object detection framework by employing a cascade of binary classifiers to scan images at multiple scales, achieving high processing speed while maintaining accuracy.

Furthermore, head position in three-dimensional space can be assessed by analyzing its tilt relative to the camera. This evaluation allows for the estimation of head rotation angles, tilt levels, and deviations, providing insights into the driver's gaze direction. Advanced facial analysis methods also incorporate more sophisticated algorithms capable of generating a 3D model of the head and face using a single camera. One of the most well-known systems in this category is based on 49 tracked 2D facial landmarks utilizing the supervised descent method (SDM). In this context, it is also important to note that many modern approaches incorporate tree-based models, Deformable Part Models (DPM), SDM, explicit shape regression, and local binary feature extraction techniques [18]. However, these methods often suffer from performance limitations when exposed to varying lighting conditions. Uneven light sources, asymmetric shadowing on the face and eye region, and abrupt changes in illumination—caused by factors such as shadows from buildings, bridges, and trees—pose significant challenges for accurate facial feature detection. Consequently, further research is required to adapt these algorithms for real-world driving conditions, enhancing the reliability and precision of driver monitoring systems.

2.3. Methods based on the acoustic environment

Previously, one of the primary challenges in studying and developing voice analysis algorithms was the limited availability of training datasets. However, with the advent of voice assistants, researchers and developers have gained access to an almost unlimited variety of speech data from diverse speakers, significantly enhancing the potential for speech analysis.

Acoustic characteristics of speech can be classified according to auditory-perceptual prosodic concepts, including prosody (pitch, intensity, rhythm, pauses, and speech rate), articulation (clarity of speech), and voice quality (e.g., breathy, tense, harsh, hoarse, or modal voice). Modern approaches to speech emotion recognition rely on precise temporal modeling of acoustic feature contours, known as feature level dynamics (FLD). This method results in the extraction of hundreds or even thousands of features used for classification. The process follows a four-step framework [19]:

- The speech signal is segmented into small time frames and smoothed using windowing functions such as the Hamming window.
- Signal processing is performed, including speaker recognition and feature extraction for each individual frame.
- The values of each frame-level feature are aggregated into FLD contours.
- The one-dimensional temporal sequence is projected onto a scalar feature that captures the temporal dynamics of the acoustic contour.

A key advantage of this sequential approach is its enhanced ability to model the contribution of both smaller units (words) and larger segments (phrases) to the prosodic structure of an utterance [20].

2.4. Focus and goal of work

Current methods for assessing driver states based on sensor data and acoustic environment analysis have several limitations that reduce their effectiveness in real-world applications. Physiological sensor-based technologies (e.g., heart rate monitoring or galvanic skin response) face challenges related to invasiveness, complex calibration requirements, and high sensitivity to individual physiological variations. Furthermore, these systems require continuous physical contact with the driver, which can cause discomfort and limit usability. Acoustic analysis-based approaches also exhibit constraints, such as susceptibility to high background noise levels within the vehicle cabin, variations in individual speech patterns, and the need for complex signal processing to achieve high detection accuracy.

Additionally, these methods are less effective when the driver remains silent or exhibits minimal speech activity.

Given these limitations, hybrid approaches that combine computer vision with biometric data analysis present a promising direction for improving driver state monitoring. Specifically, integrating face recognition, head and body posture assessment, and MEMS sensor data enables the development of more robust monitoring systems. Video-based analysis offers a non-invasive means of evaluating driver behavior, while MEMS sensors provide physiological and behavioral insights, enhancing the accuracy of fatigue, drowsiness, and distraction detection.

Thus, the aim of this paper is to develop intelligent monitoring system for analyzing vehicle drivers state based on adaptive deep learning models.

3. System's concept development

3.1. Main functions formalization

To address the outlined problem, the following concept of intelligent monitoring system for analyzing vehicle drivers state on can be proposed:

- Development system's first module (M1) with a DNNs, adapted from existing DL models, for detecting key points on the face and head with the purpose of binary or multiclass classification, aimed at assessing the driver's level of fatigue.
- Development system's second module (M2), also adapted from existing DL models, for detecting distractions affecting the driver during driving.
- Aggregation of the outputs from M1 and M2 to enhance result accuracy and reduce the number of false positives.

To comprehensively assess the condition of a vehicle driver for drowsiness detection through automated recognition and classification of video stream images, followed by an analysis of the driver's focus level or distraction from the traffic process, it is proposed to develop and use two separate modules that implement different DL models, which have models to handle the processing and analysis of data for assessing driver drowsiness: by analyzing head posture considering distractions and by analyzing eye condition. A generalized scheme of the project stages is presented in Figure 1.
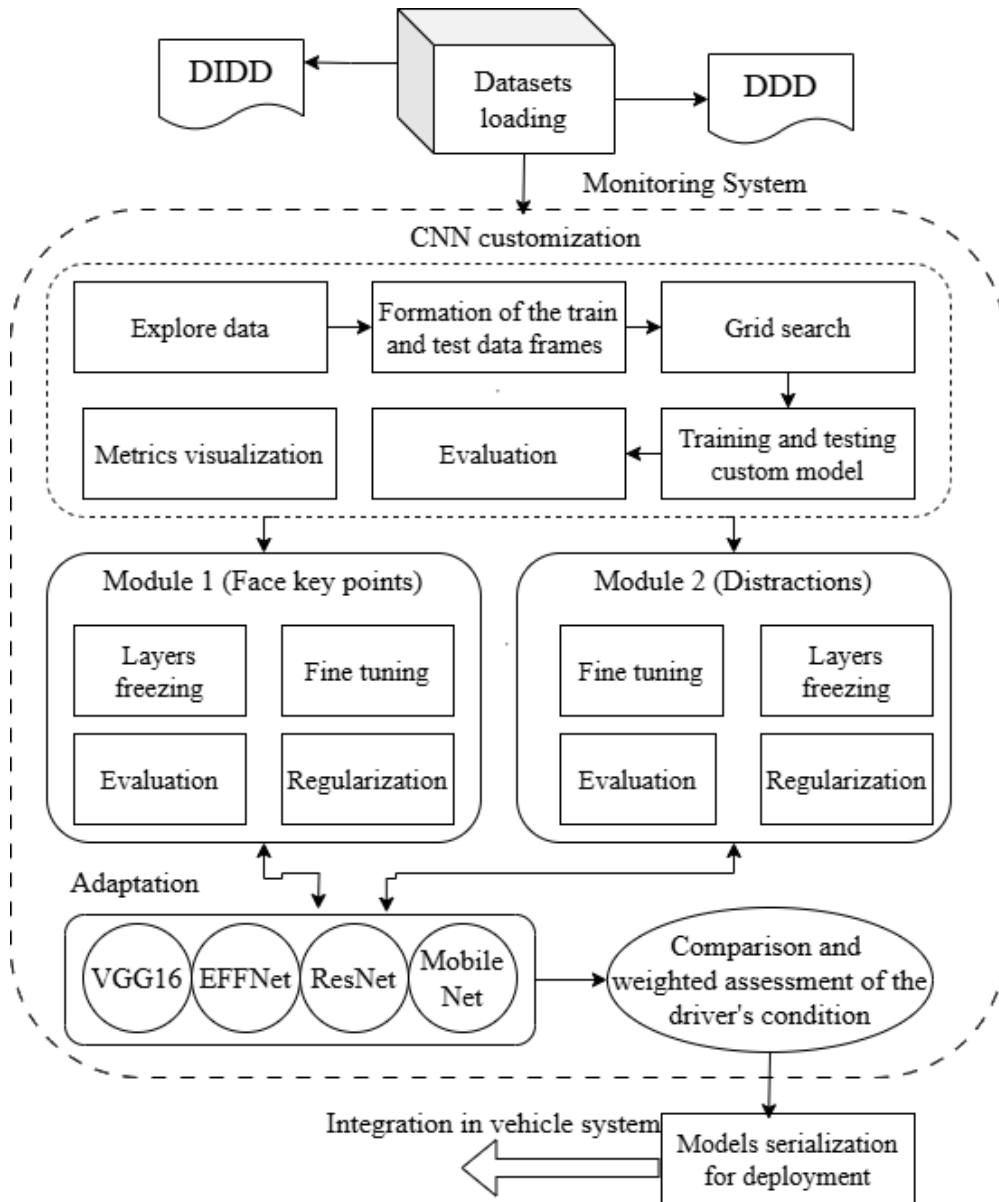
Figure 1: Intelligent system consept stages scheme

The key aspects of the implementation are as follows:

- Data selection and loading. At this stage, a dataset containing images and data regarding the driver's condition (head posture, eye condition) is chosen and loaded into the working environment. For deep learning models like ResNet and MobileNet, high-resolution video frames are loaded, as both models have been pre-trained on large image datasets.

- Data preparation and preprocessing. This step involves standardizing the input format, including resizing images, normalizing pixel values, and augmenting the data. For ResNet and MobileNet, images are resized to a fixed format (e.g., 224×224), and augmentation techniques such as rotation, mirroring, and brightness adjustment are applied.

- Forming training and testing subsets. Based on the size of the data, the dataset is split into training and testing subsets in an 80/20 or 70/30 ratio. Cross-validation is used to enhance the robustness of the models.

- Creating and loading DL models. Pre-trained DL architectures, such as ResNet-50, can be used, with the last fully connected layer being replaced for driver condition classification tasks. The MobileNet model can also be used for lightweight and fast classification, followed by fine-tuning and adding fully connected layers to process specific data.

- Training and fine-tuning models. The training process includes adjusting hyperparameters such as learning rate (0.001-0.01), number of epochs (10-30), and optimization functions (such as Adam or SGD).

- Metrics evaluation and results analysis. At this stage, the models' quality is assessed using appropriate metrics to analyze the driver's condition based on the selected factors.
- Decision making. Based on the data and predictions, decisions are made to adjust system actions accordingly.

In the implementation of the described concept, the adaptive feature fusion mechanism is of key importance, which includes the following stages:
- weighted fusion of features based on the dynamic confidence coefficient of the model;
- Bayesian aggregation of probabilistic predictions to improve the accuracy of determining the driver's state (analysis of the level of drowsiness);
- adaptation of the attention mechanism to focus on the most informative regions of the video stream images;
- optimization of the final assessment of the driver's state using the retrained VGG16 model.

That is, the deployed ResNet, MobileNet and CNN models extract features $F_{cnn} \in \Re^{d_c}, F_{res} \in \Re^{d_r}, F_{mob} \in \Re^{d_m}$ respectively. Then the final representation of the combined features F_{fusion} is defined as:

$$F_{fusion} = w_{res} \cdot F_{res} + w_{mob} \cdot F_{mob} + w_{cnn} \cdot F_{cnn},$$

(1)

where $w_{res}, w_{mob}, w_{cnn}$ – adaptive weights determined through the attention mechanism:

$$w_i = \frac{e^{S_j}}{\sum_j e^{S_j}}, S_j = MLP(F_i)$$

(2)

where $MLP(F_i)$ - a multilayer perceptron that learns to predict the importance of each feature channel.

Bayesian aggregation of model predictions is based on probabilistic combination of predictions of each model:

$$P(y|X) = \sum_i w_i P_i(y|X),$$

(3)

where $P_i(y|X)$ – the probability of predicting the level of sleepiness produced by each model.

The VGG16 model is used to further validate the output representation by using the following model output correction function:

$$F_{opt} = \sigma(WF_{fusion} + b),$$

(4)

where W - learnable transformation matrix, b – bias, σ – activation function (ReLU or softmax).

The final assessment of the driver's condition is calculated as:

$$P_{final}(y|X) = \alpha P(y|X) + (1-\alpha) P_{vgg}(y|X),$$

(5)

where α – weighting factor determined based on the confidence level of the VGG16 model.

Given the labor-intensive nature of creating a custom dataset, which includes aggregation, formatting, and labeling, the decision has been made to use existing publicly available datasets compiled by third-party experts for the training and testing of data analysis models.

3.2. Datasets description

In the development of intelligent system's module for processing and analyzing data to assess driver drowsiness based on head position and distraction factors, the driver-inattention-detection-dataset [21] has been selected. This dataset, presented in grayscale, is highly diverse and includes over 14,000 labeled images distributed across six different classes, providing a broad and varied data range for training, validation, and testing tasks specifically tailored for grayscale image processing.

The dataset is organized into three main directories: training (11,942 grayscale images that have been carefully selected and labeled across six classes), validation (1,922 images used for model tuning and performance evaluation during the development process), test (985 images reserved for final verification and comparative analysis of the models). This dataset covers six classes of driver

behavior: dangerous driving, distracted driving, alcohol consumption, safe driving, drowsy driving, yawning.

For further exploration of the potential intelligent system's M2 for a different, more specialized, and pre-processed dataset focusing on driver eye images, besides the previously discussed ResNet50 and MobileNetV2 models, EfficientNetB0 and VGG16 models were selected. The dataset chosen for this purpose is the Driver Drowsiness Dataset (DDD) [22], which contains extracted and cropped images of drivers' faces from video recordings of real-world cases of drowsiness while driving.

This dataset is intended for the development and training of machine learning and deep learning models capable of detecting signs of drowsiness in drivers by analyzing their eye regions.

Since the data were collected from real video recordings, they reflect a variety of lighting conditions, angles, and other factors, making them valuable for creating robust and reliable drowsiness detection systems.

The DDD includes more than 41,790 images of drivers' faces, and the dataset structure is as follows: RGB images with a size of 227×227 pixels, labeled into two classes — "drowsy" and "alert," involving 28 drivers, each assigned a unique identifier.

3.3. Neural network models development

According to M1 logic implementation all the images uploaded into the system are converted to RGB format and resized to 224x224 pixels at the preprocessing stage. The class labels are encoded using one-hot encoding.

For the experiments, it was decided to use a classical convolutional neural network (CNN) architecture, as well as compare it with pre-trained models such as ResNet50 and MobileNetV2.

The ResNet50 architecture includes residual blocks, which help address the vanishing gradient problem common in deep neural networks. Specifically, the model incorporates GlobalAveragePooling2D layers to reduce feature dimensionality, a fully connected Dense layer with 512 neurons and the ReLU activation function, and a final Dense layer with 6 neurons and softmax activation.

The MobileNetV2 architecture employs depthwise separable convolutions, which significantly reduce computational complexity. For driver state analysis, a similar approach to ResNet50 was used, where the base layers of MobileNetV2 were frozen (using pre-trained weights from ImageNet), and GlobalAveragePooling2D layers, a fully connected Dense layer with 512 neurons and the ReLU activation function, as well as the softmax-activated output layer were added.

The training process for both models is similar to that of ResNet50, but MobileNetV2 offers a lower computational load, making it more efficient in environments with limited computational resources.

The CNN model architecture (Figure 2) consists of several Conv2D convolutional layers with ReLU activation, MaxPooling2D subsampling layers, a fully connected Dense layer with 512 neurons, and Dropout to prevent overfitting, along with an output layer with softmax activation for classifying into 6 classes.
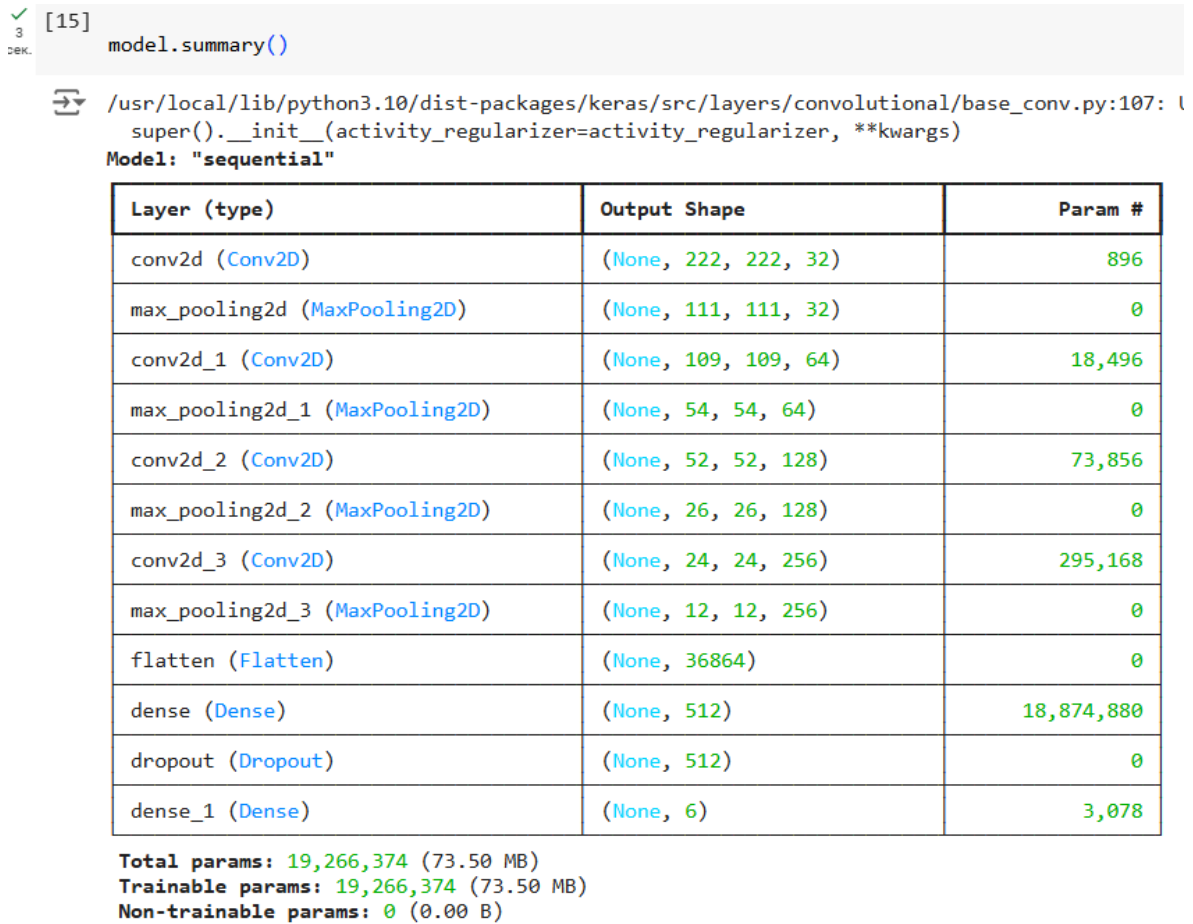
```
✓ [15]
3
сек.    model.summary()
```

```
/usr/local/lib/python3.10/dist-packages/keras/src/layers/convolutional/base_conv.py:107: (
    super().__init__(activity_regularizer=activity_regularizer, **kwargs)
Model: "sequential"
```

Layer (type)	Output Shape	Param #
conv2d (Conv2D)	(None, 222, 222, 32)	896
max_pooling2d (MaxPooling2D)	(None, 111, 111, 32)	0
conv2d_1 (Conv2D)	(None, 109, 109, 64)	18,496
max_pooling2d_1 (MaxPooling2D)	(None, 54, 54, 64)	0
conv2d_2 (Conv2D)	(None, 52, 52, 128)	73,856
max_pooling2d_2 (MaxPooling2D)	(None, 26, 26, 128)	0
conv2d_3 (Conv2D)	(None, 24, 24, 256)	295,168
max_pooling2d_3 (MaxPooling2D)	(None, 12, 12, 256)	0
flatten (Flatten)	(None, 36864)	0
dense (Dense)	(None, 512)	18,874,880
dropout (Dropout)	(None, 512)	0
dense_1 (Dense)	(None, 6)	3,078

```
Total params: 19,266,374 (73.50 MB)
Trainable params: 19,266,374 (73.50 MB)
Non-trainable params: 0 (0.00 B)
```

Figure 2: Main CNN model structure

In M2 implementation the research followed the subsequent steps:
- Data preprocessing, including normalization of images and resizing them to the required dimensions for each model (e.g., 224x224 pixels for most models).
- Data augmentation to increase the diversity of the training set and improve the models' robustness (e.g., rotations, shifts, brightness adjustments).
- Model initialization with pre-trained weights, which accelerates the learning process and improves accuracy.
- The ReLU activation function was used as the optimizer, and Sigmoid as the loss function, with binary cross-entropy applied as the loss function due to the binary classification task.
- Testing was performed by splitting the data into training and testing subsets.
- Model performance evaluation, using metrics similar to those in the previous study, and cross-validation to assess the robustness of the models on different data subsets.

4. Experiments and results analysis

In M1 accuracy, F1-score, precision, recall were used as metrics for assessing the accuracy of the models, each of which was evaluated on a test set. Classic CNN is characterized by a simpler architecture, high performance and base accuracy. The ResNet50 model is characterized by higher accuracy due to pre-trained weights, and MobileNetV2 demonstrates moderate (not very high) accuracy, but is more efficient in terms of consumption of computing resources. At the same time, the ResNet50 model copes best with the "SleepyDriving" and "Yawn" classes.

Comparison of F1 Score and Precision metrics evaluation results for adaprive DL models in M1 is shown in Figure 3. It should be noted that there is a consistent decrease in the loss and an increase in accuracy for each model, indicating the absence of overfitting. The ResNet50 model demonstrates the most stable convergence. Visualization of the results of constructing error matrices for adaptive DL models in M1 is shown in Figure 4.
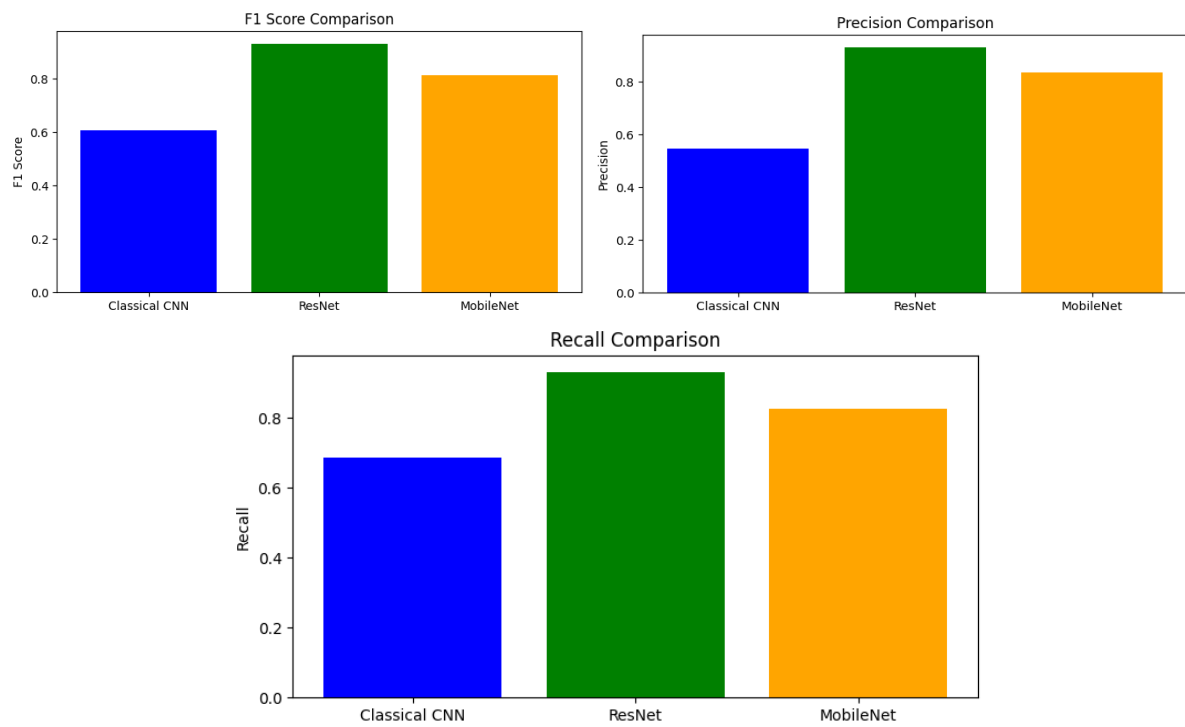
Figure 3: Comparison of F1 Score and Precision metrics evaluation results for adaptive DL models in M1
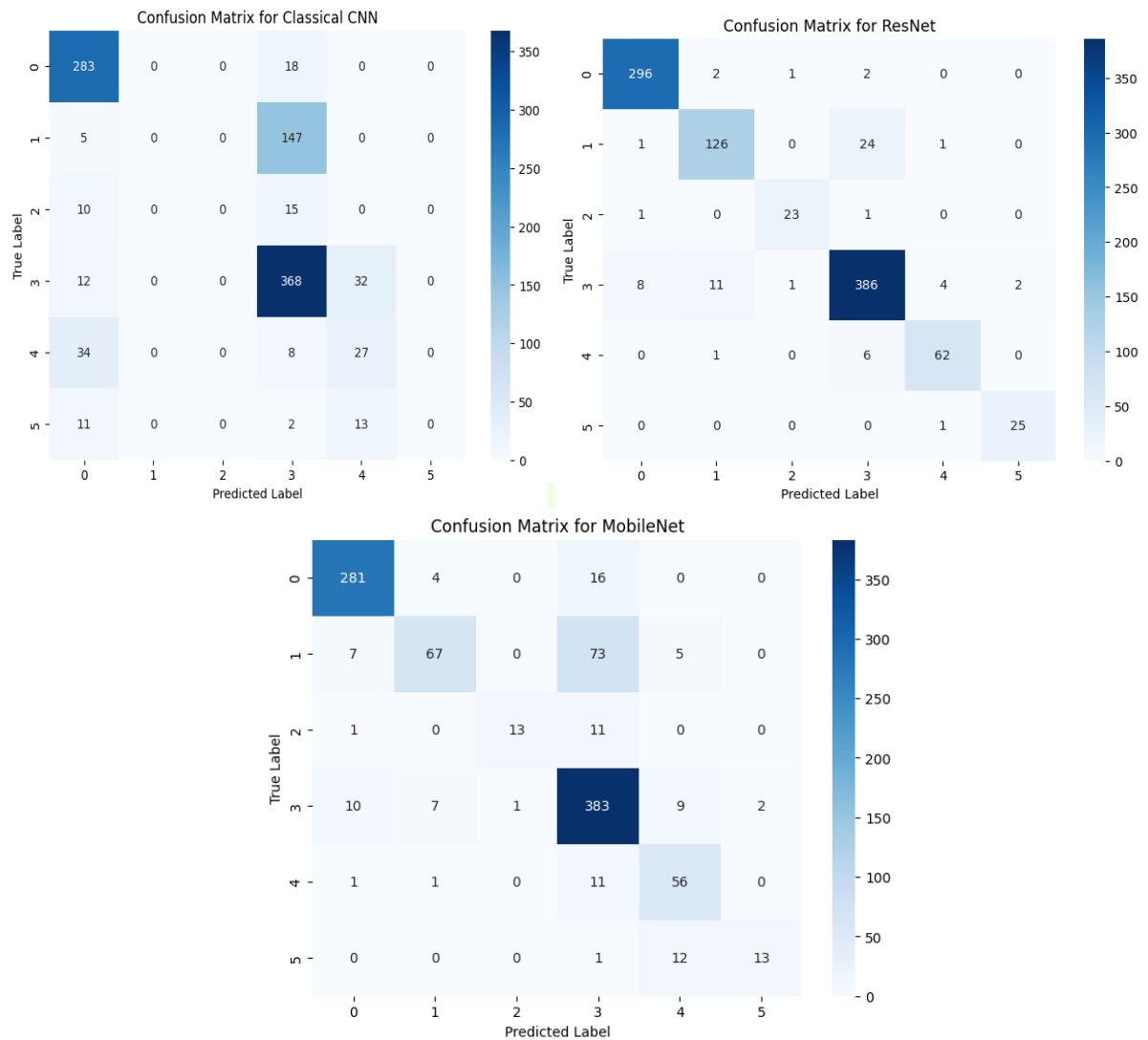

Figure 4: Constructing error matrices results Visualization for adaptive DL models in M1

Rational approaches to improving the accuracy of the loaded models include: fine-tuning by unfreezing the upper layers of the ResNet50 base model and retraining them on additional data; using a smaller learning rate for the unfrozen layers; increasing data variability by applying augmentation techniques such as rotations, brightness adjustments, and horizontal flipping, as well as data mixing (images and labels) to improve model robustness against noise.

Furthermore, there is potential to add additional features, such as the sequence of frames for analyzing the fatigue dynamics, and to increase the number of parameters in the dense layers by adding more layers or neurons to improve the generalization capability of the models.

It is worth noting the consistent decrease in loss and increase in accuracy for each model, indicating the absence of overfitting. The ResNet50 model demonstrates the most stable convergence.

To improve the accuracy of the loaded models, several effective strategies can be considered: fine-tuning by unfreezing the upper layers of the base ResNet50 model and retraining on additional data; using a lower learning rate for the unfrozen layers; increasing data variability through augmentation (such as rotations, brightness adjustments, and horizontal flipping), as well as employing data mixing techniques (images and labels) to improve model robustness against noise.

Additionally, new features can be introduced, such as the sequence of frames for analyzing the dynamics of fatigue, and the number of parameters in the Dense layers can be increased by adding additional layers or increasing the number of neurons, which would enhance the generalization capabilities of the models.

Dependence of values on the number of model training epochs for custom CNN, tuned MobileNet and ResNet is shown in Figure 5.

In M2 we can say, that the difference in model error rates between the training and test sets is minimal, ranging from 3% to 7%, indicating data balance and the high efficiency of fine-tuning models on the constructed datasets using cross-validation.

An analysis of the presented dependencies reveals that the accuracy of the ResNet50 model gradually increases, reaching approximately 0.85 by the end of training, which suggests well-balanced classes and a successful learning process. However, the validation accuracy exhibits some instability: it peaks at around 0.86 during the early epochs but then declines to below 0.80 by the 200th epoch.

This trend may indicate overfitting, as training accuracy continues to increase while validation accuracy decreases.

The accuracy of the MobileNetV2 model initially increases gradually, reaching 0.82 in the later training stages. However, its accuracy improvement is less pronounced compared to other models, and its validation accuracy peaks at 0.84 in the early epochs before declining more significantly than that of ResNet50. This suggests overfitting or potential issues with generalization.

For the EfficientNetB0 model, accuracy also increases with more training epochs, reaching 0.82 in the final stages, albeit at a slower rate compared to other models. Notably, its validation accuracy steadily improves over time, surpassing the training accuracy in later stages and reaching 0.86. This behavior indicates strong generalization capabilities without significant overfitting.

The VGG16 model initially exhibits lower accuracy during training but eventually reaches 0.81. At early stages, its validation accuracy is higher than training accuracy and remains stable at approximately 0.83 by the end of training. This suggests good overall performance, though possible underfitting may need to be addressed.

Summary graph of estimates of training and test accuracies of adaptive DL models is shown in Figure 6.

To test the operation of the created modules and serialized models, test scripts were developed that run the models on prepared videos.

This allowed parallel recognition of driver states in console mode. This approach allows for the prompt analysis of behavioral signs of drowsiness, distraction, and other factors affecting driving safety.
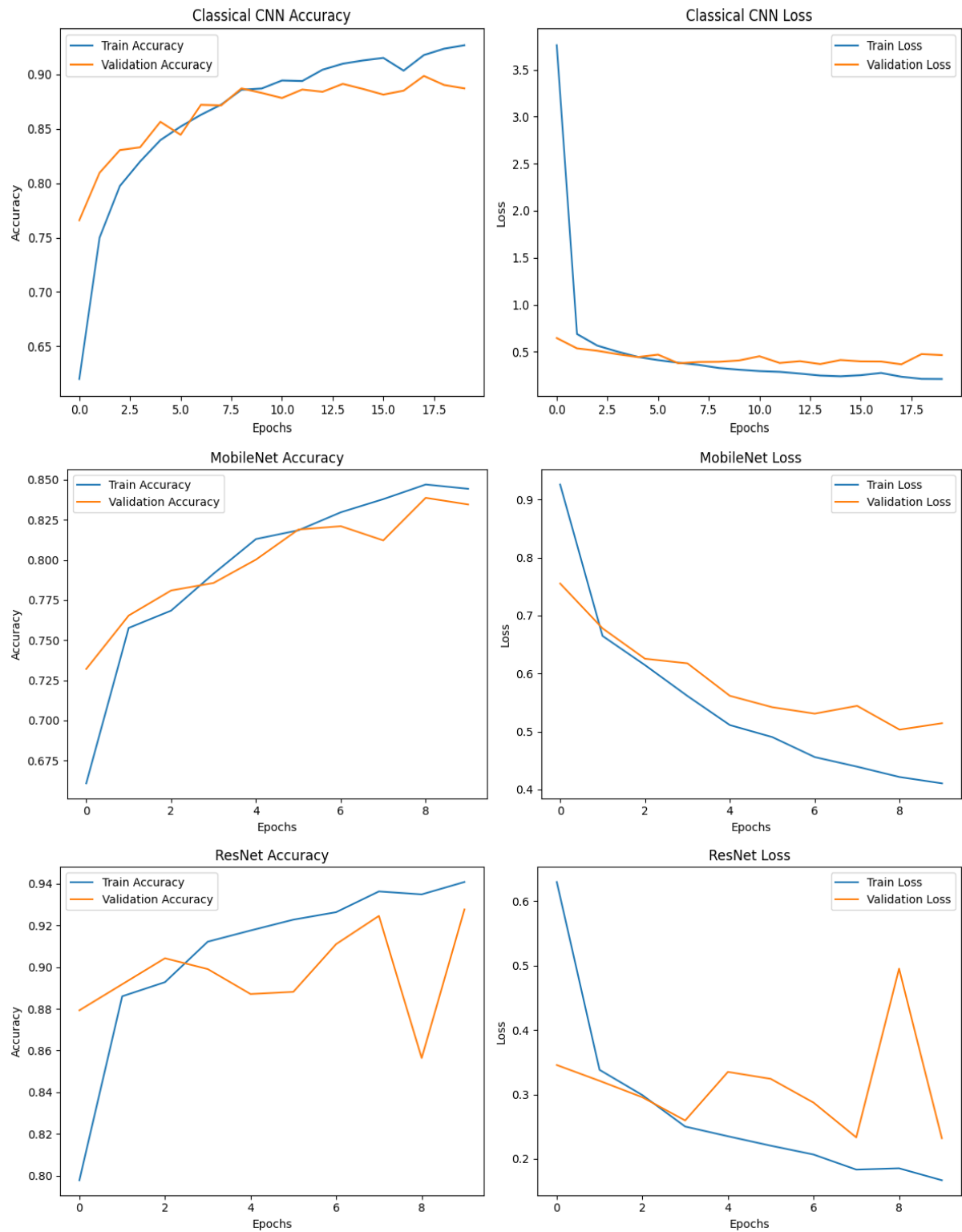
Figure 5: Dependence of values on the number of model training epochs for custom CNN, tuned MobileNet and ResNet

The testing results are presented in Figure 7, where we can see how each module of the system process the video stream and classify the driver's state in real time (evaluates the level of driver's state - drowsiness). Particular attention is paid to the analysis of the stability of the models to changes in lighting conditions, angles, and differences in the anatomical features of vehicle drivers. In summary, the ResNet50 and MobileNetV2 models exhibit signs of overfitting, as the gap between training and validation accuracy increases with more training epochs. In contrast, EfficientNetB0 demonstrates stable performance improvements on both training and validation sets, suggesting its

advantage for this dataset. The VGG16 model maintains consistent but non-optimal results, indicating a potential need for additional hyperparameter tuning or increased training epochs.
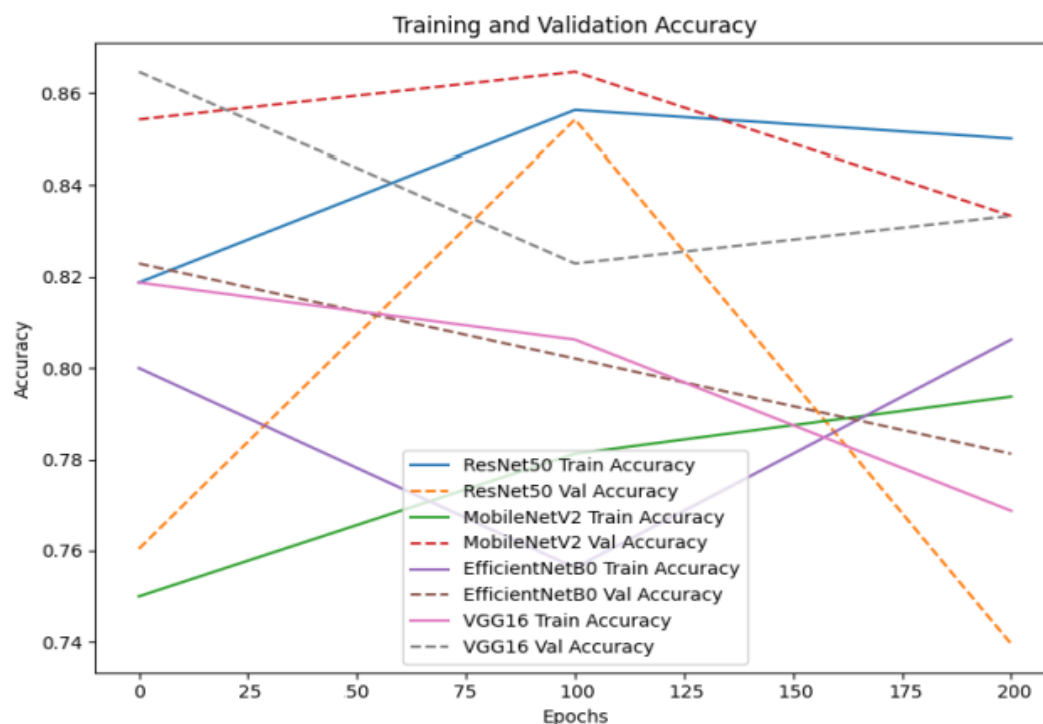

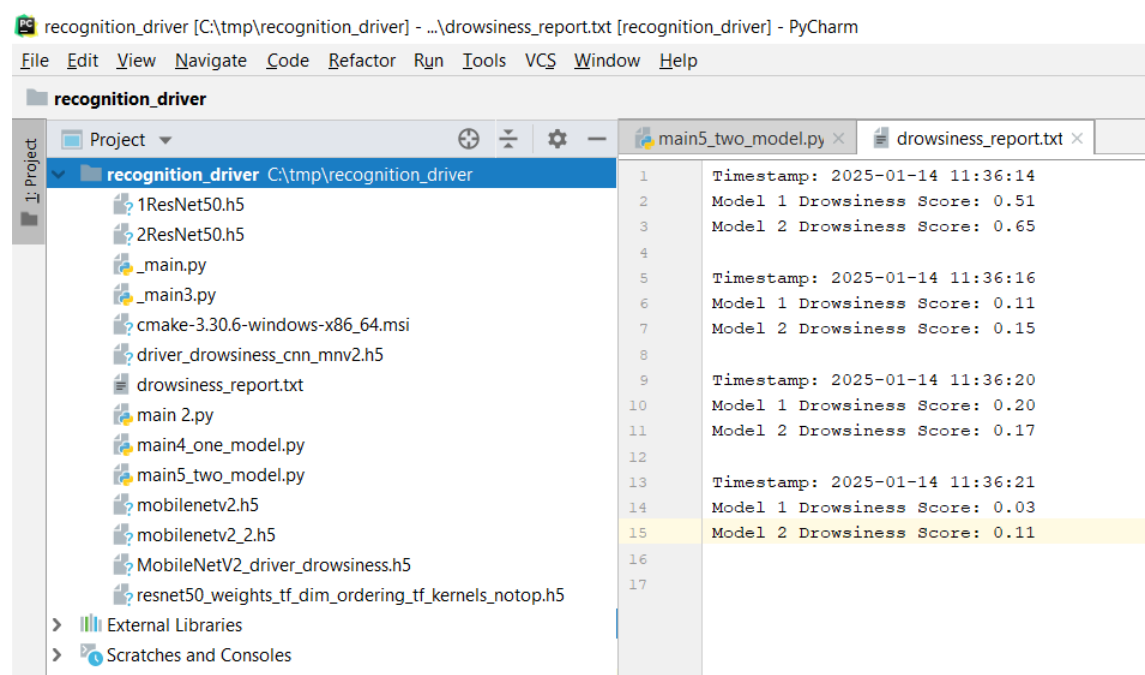**Figure 6:** Summary graph of estimates of training and test accuracies of adaptive DL models

Figure 7: Test enviroment for M1 and M2 usage in PyCharm

5. Conclusions

The research results demonstrate us the effectiveness of fine-tuning and the adaptation of existing DL models, specifically MobileNetV2 and ResNet50, in developed intelligent monitoring system for analyzing vehicle drivers state. By leveraging pre-trained architectures, the models achieve high classification accuracy while reducing computational costs and training time.

The scientific novelty of the developed system lies in the hybrid approach, combining several adaptive deep learning models to improve the accuracy and reliability of real-time driver

monitoring. For the first time, an architecture with two modules based on different convolutional neural networks was implemented, which made it possible to adapt the system to different scenarios and resource constraints. ResNet50, with its residual learning framework, effectively captures complex feature representations but exhibits higher computational demands. In contrast, MobileNetV2, optimized for lightweight and efficient deployment, ensures faster inference while maintaining competitive accuracy, particularly in tasks focusing on eye-region analysis. The results indicate that both models generalize well when fine-tuned on domain-specific datasets, particularly in detecting signs of drowsiness and distraction. As observed, the MobileNetV2 model demonstrates a more accurate assessment of the driver's condition, particularly when analyzing segments containing the ocular region. Moreover, its performance is 2–3 times faster than that of the ResNet50 model. This can be attributed to the fact that ResNet50 considers a broader feature space and possesses a more complex architecture, leading to an increased size of serialized objects and weight values.

However, in cases where the driver's eyes are partially closed or the head is significantly tilted sideways or downward, both models exhibit high confidence levels in detecting driver drowsiness. This finding indicates a high generalization capability of the models and confirms the effectiveness of their fine-tuning on representative datasets. These results suggest that MobileNetV2 may be preferable for resource-constrained real-time systems, whereas ResNet50, due to its deeper architecture, can provide a more detailed analysis of complex scenarios.

Future research efforts should focus on enhancing the accuracy of DL models by implementing the following strategies:

- Integration of multimodal data. Utilizing multiple data sources, such as video recordings, voice signals, biometric indicators, and vehicle movement data, to improve the reliability of driver state assessment.
- Training on large and representative datasets. Expanding the dataset to include a diverse range of drivers across different ages, genders, cultural backgrounds, and driving conditions, ensuring robust generalization.
- Handling rare events. Emphasizing the recognition of rare and critical driver states, such as microsleep episodes or sudden health deterioration, to enhance safety-critical detection capabilities.

A promising direction for the development of the system is the integration of multimodal data and automatic adaptation of the architecture to specific operating conditions.

Declaration on Generative AI

During the preparation of this work, the authors used Grammarly in order to: Grammar and spelling check. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the publication's content.

References

[1] N. Rudnichenko, V. Vychuzhanin, D. Shvedov, T. Otradskya, I. Petrov, Information system for generating recommendations for risk-oriented trading strategies based on deep learning, in: Proceedings of the 7th Workshop for Young Scientists in Computer Science & Software Engineering (CS&SE@SW 2024), 2024, ceur-ws.org/Vol-3917, pp. 110-119.
[2] N. Rudnichenko, V. Vychuzhanin, T. Otradskya, D. Shvedov, Intelligent System for Processing and Forecasting Financial Assets and Risks. in: CMIS-2024 Computer Modeling and Intelligent Systems 2024, ceur-ws.org/Vol-3702, pp. 251- 262.
[3] V. Vychuzhanin, N. Rudnichenko, A.Vychuzhanin, A. Rychlik, Diagnosis Intellectualization of Complex Technical Systems, in: ICST-2023 Information Control Systems & Technologies 2023, ceur-ws.org/Vol- 3513, pp.352- 362
[4] R. Chinthalachervu, I. Teja, M. Ajay Kumar, N. Sai Harshith, T. Santosh Kumar, Driver Drowsiness Detection Using Machine Learning, in: International Conference on Electronic Circuits and Signalling Technologies, 2325, 2022, 012057. doi: doi:10.1088/1742-6596/2325/1/012057
[5] S.A. El-Nabi, W. El-Shafai, E.-S.M. El-Rabaie, K.F. Ramadan, F.E. Abd El-Samie, and S. Mohsen, Machine learning and deep learning techniques for driver fatigue and drowsiness detection: a

review. Multimedia Tools and Applications 83 (2024) 9441–9477. doi: 10.1007/s11042-023-15054-0.

[6] Sabu, S., Driver Drowsiness Detection and Warning System. International Journal for Research in Applied Science and Engineering Technology 9 5 (2022) 657-660. doi:10.22214/ijraset.2021.34246

[7] Giridhar S., Rhithik Raj K., Sreeram K.P., Nithish M., Aswathy R., Harini S. An Intelligent System for Preventing Accidents Due to Driver Distractions. Procedia Computer Science 235 (2023) 2196-2204. doi: 10.1016/j.procs.2024.04.208

[8] S. Nandhini, V. Venkatasubramanian, C. Aparna, Drowsiness Detection Using Facial Features, Image Processing, and Machine Learning. Computer Networks, Big Data and IoT (2022) 565-574. doi: 10.1007/978-981-19-0898-9_452.

[9] S Amutha, P. Vamsi, K. M. Reddy, C. Mahesh, P. V. K. Reddy, K. A. Reddy, Driver Drowsiness Detection System using OpenCV and Keras, in: Conference: 2023 International Conference on Data Science, Agents & Artificial Intelligence (ICDSAAI), 2023. doi: 10.1109/ICDSAAI59313.2023.10452590

[10] A. Deepa, R.R.D. Sai, K. Milind, Y. Vijayalata, K. Rahul, Drowsiness detection using IoT and facial expression, in: Proceedings of the international conference on cognitive and intelligent computing. Springer, Singapore, 2023, pp 679–692. doi:10.1007/978-981-19-2358-6_61

[11] S.A. Shiney, R. Seetharaman, N. Mageshwari, S. Saikiran, B.K. Sulaksha, S.Sai Pravin, P.L.M. Dheepak, Methodology for Driver Drowsiness Detection Based on OpenCV and Dlib Libraries, in: Conference: 2024 5th International Conference on Smart Electronics and Communication (ICOSEC)(2024). doi: 10.1109/ICOSEC61587.2024.10722297

[12] S.A. Ajagbe, O.A. Oki, M.A. Oladipupo, A. Nwanakwaugwu, Investigating the efficiency of deep learning models in bioinspired object detection, in: 2022 International conference on electrical, computer and energy technologies (ICECET), pp 1–6. doi:10.1109/ICECET55527.2022.9872568

[13] Y. Albadawi, M. Takruri, M. Awad, A review of recent developments in driver drowsiness detection systems. Sensors 22 5) (2022) 2069 doi:10.3390/s22052069

[14] T. Arakawa, Trends and future prospects of the drowsiness detection and estimation technology. Sensors 21 23 (2021) 7921. doi: 10.3390/s21237921.

[15] A. M. Băiaşu, C. Dumitrescu, Contributions to driver fatigue detection based on eye-tracking. International Journal of Circuits, Systems and Signal Processing 15 (2021) 1–7. doi: 10.46300/9106.2021.15.1.

[16] S. Bajaj, L. Panchal, S. Patil, K. Sanas, H. Bhatt, S. Dhakane, A real-time driver drowsiness detection using OpenCV, DLib. ICT Analysis and Applications (2023) 639–649. doi: 10.1109/ICAC3N56670.2022.10074245.

[17] M. Dua, R. Singla, S. Raj, A. Jangra, Deep CNN models-based ensemble approach to driver drowsiness detection. Neural Computing & Applications 33 8 (2021) 3155–3168. doi: 10.1007/s00521-020-05209-7.

[18] I. A. Fouad, A robust and efficient EEG-based drowsiness detection system using different machine learning algorithms. Ain Shams Engineering Journal 14 (2022) 101895. doi: 10.1016/j.asej.2022.101895.

[19] S. S. Jasim, A. K. A. Hassan, Modern drowsiness detection in deep learning: A review. Journal of Al-Qadisiyah for Computer Science and Mathematics 14 3 (2022) 119. doi: 10.29304/jqcm.2022.14.3.1023.

[20] M. T. A. Dipu, S. S. Hossain, Y. Arafat, F. B. Rafiq, Real-time driver drowsiness detection using deep learning. International Journal of Advanced Computer Science and Applications 12 7 (2021) 844–850. doi:10.14569/IJACSA.2021.0120794.

[21] Driver inattention detection dataset. URL: https://www.kaggle.com/datasets/zeyad1mashhour/driver-inattention-detection-dataset

[22] Driver drowsiness dataset ddd. URL: https://www.kaggle.com/datasets/ismailnasri20/driver-drowsiness-dataset-ddd