Driver Drowsiness Detection Using Transform Learning and Behavioral Analysis

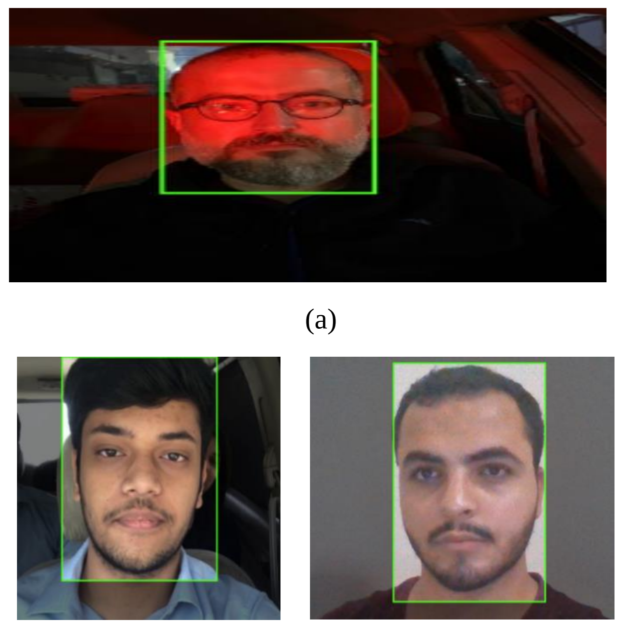
Omar El-Shinawy, Ali Adel, Aya Hussein, Amr Yasser, and Abdullah Mohmmed
Supervised by Mahmoud Talat [1], Digital Egypt Pioneers Initiative - DEPI, Cairo, Egypt
[1] Ai Engineer at MCiT (Ministry of Communication and Information Technology)
[1] TA at Zewail University (Artificial intelligence and Data science)

Abstract

This paper investigates the pivotal role of drowsiness detection systems in accident prevention and the enhancement of safety in various sectors, with a primary focus on transportation. As drowsiness is a significant contributor to road accidents, effective detection mechanisms are essential. We review cutting-edge behavioral and machine learning techniques for identifying driver drowsiness, utilizing facial expressions such as eye blinks, head movements, and yawning. Our research proposes the implementation of transfer learning models, specifically YOLOv8s, along with established frameworks like Faster R-CNN and SSD. We conduct a comprehensive analysis of each model's training processes, performance metrics, and inherent limitations, while considering diverse input data conditions, including variations in lighting, image quality, and the presence of human accessories such as eyeglasses and mobile devices. Furthermore, we assess the potential biases that may arise from training data representing different nationalities. Our evaluation methodology encompasses iterative cycles of data acquisition, feature extraction, mobile training, and rigorous testing. The project culminates in a deployment phase, where the resulting application will be made publicly available at no cost. Ultimately, this study aims to advance real-time drowsiness detection systems capable of delivering timely alerts to drivers, thereby significantly improving road safety.

Introduction

Drowsiness Detection has emerged as a critical area of research, particularly due to its implications for road safety and accident prevention in the transportation sector. The ability to accurately identify when a driver is drowsy can save lives and reduce the frequency of vehicular accidents caused by fatigue. This paper explores various methodologies developed to address this pressing issue, focusing on state-of-the-art behavioral and machine learning techniques. In our study, we leverage advanced facial recognition technology to analyze key indicators of drowsiness, including eye blinks, head movements, and yawning. These behavioral cues serve as reliable indicators of a driver's alertness level. To enhance the accuracy and robustness of our detection system, we propose the use of transfer learning models, particularly YOLOv8s,



(a)

(b) (c)

Figure 1: (a) Shows a sample of data we used representing the Night-time. (b) Sample data at perfect light conditions. (c) Sample of data we contributed ourselves representing different conditions of quality, light, appearance of glasses etc.

alongside established frameworks such as Faster R-CNN and SSD. The methodology section of this paper is structured into distinct segments dedicated to each model: YOLOv8s, Faster R-CNN, and SSD. In these sections, we will delve into the intricacies of their training processes, elucidating how each model learns from the data, the loss functions employed, and the performance metrics achieved. We will also critically assess the limitations of each approach, particularly in relation to the diverse conditions under which the models are trained and tested. A significant aspect of our research includes an analysis of the input data quality, which encompasses not only image clarity and resolution but also the impact of environmental factors such as lighting conditions. Additionally, we will investigate how the presence of human accessories, including eyeglasses and mobile devices, can influence detection accuracy. Importantly, we will examine the potential for training data bias by evaluating face features from individuals of different nationalities. Through an iterative methodology involving data acquisition, feature extraction, mobile training, and comprehensive testing, we aim to develop a robust real-time drowsiness detection system. The ultimate goal of this research is to create a publicly accessible application that provides timely alerts to drivers, thereby enhancing overall road safety and contributing to the ongoing efforts to mitigate the dangers associated with drowsy driving.

Related Works

The field of drowsiness detection has garnered significant attention in recent years, with numerous studies exploring various methodologies and technologies aimed at enhancing driver safety. This section reviews pertinent literature that informs our research on real-time drowsiness detection systems, particularly focusing on behavioral analysis and machine learning techniques. One notable study by Kumar et al. (2020) presents a comprehensive overview of driver drowsiness detection using facial expression analysis. The authors utilize a combination of eye tracking and head pose estimation to identify signs of fatigue. Their findings suggest that the integration of multiple behavioral cues significantly improves detection accuracy, aligning with our approach of leveraging eye blinks, head movements, and yawning as indicators of drowsiness. In another significant Contribution, Zhang.et.al(2021) investigated the application of deep learning techniques for real-time drowsiness detection. They employed convolutional neural networks (CNNs) to analyze facial images and classify alertness levels. By utilizing a diverse dataset that included subjects from various demographics, this study highlights the importance of dataset diversity in mitigating training bias, a concern we also address in our research. Furthermore, Aljohani.et.al.(2019) explored the effectiveness of transfer learning in driver drowsiness detection. Their work demonstrated that pre-trained models could significantly enhance detection performance when fine-tuned on specific datasets. This aligns with our methodology of implementing transfer learning models, such as YOLOv8s, Faster R-CNN, and SSD, to improve the robustness of our drowsiness

Figure 2: The baseline data project started with

detection system. The research conducted by Li et al. (2022) introduced a multimodal drowsiness detection system that integrates both physiological signals and visual data. By combining facial expression analysis with heart rate variability measurements, the authors reported improved accuracy in detecting driver fatigue. This multimodal approach underscores the potential benefits of incorporating additional data sources, which could be explored in future iterations of our system. Finally, Singh and Gupta (2023) provided an extensive review of various machine learning techniques for drowsiness detection, emphasizing the challenges and limitations associated with different models. Their analysis of loss functions, training processes, and performance metrics serves as a valuable reference for our evaluation of the YOLOv8s, Faster R-CNN, and SSD models. By synthesizing insights from this review, we aim to address identified limitations and enhance the effectiveness of our proposed methodology. In summary, the existing literature illustrates a robust foundation for developing advanced drowsiness detection systems. Our research builds upon these studies by integrating behavioral analysis with state-of-the-art machine learning models, while also addressing critical factors such as input data quality and training biases to further enhance the accuracy and reliability of drowsiness detection in real-time-application.

Data

In this section, we describe the datasets utilized in our drowsiness detection project, detailing their characteristics and contributions to model training. We began our data collection process with the Daylight Dataset (Figure 2), which served as the baseline for our model training. Collected under various daylight conditions, this dataset captured a range of facial expressions and drowsiness indicators. However, as the training progressed, it became apparent that relying solely on this dataset was insufficient for achieving optimal model performance. The limitations identified in the Daylight Dataset led us to explore

Figure 3: our personal data augmented

Figure 4: Night enhancement data augmented

additional options for data augmentation. To address these shortcomings, we initiated the creation of an Augmented Personal Dataset (Figure 3). This dataset was developed by augmenting the existing daylight images with additional data collected from various sources, including personal recordings which allowed us to generate a more diverse set of training examples. This enhanced dataset significantly contributed to improving the model's ability to generalize across different conditions and individual variations. To further enhance the model's performance, particularly in low-light conditions, we incorporated a Night Vision Dataset (Figure 4). This dataset was specifically designed to address the challenges associated with drowsiness detection in nighttime settings. By including images captured in various nighttime environments, we aimed to improve the model's accuracy and reliability when detecting drowsiness during evening or early morning driving scenarios. The integration of this third dataset ensured that our model could effectively operate under a wide range of lighting conditions. In summary, the combination of these three datasets—Daylight, Augmented Personal, and Night Vision—allowed us to build a robust and versatile drowsiness detection model capable of delivering accurate results across diverse real-world scenarios. Each dataset played a critical role in the overall training process, ensuring that the model could effectively recognize signs of drowsiness, regardless of environmental factors.

Method

In this section, we detail the training processes for each of the models employed in our drowsiness detection system: YOLO, Faster R-CNN, and SSD. Each model underwent a series of training iterations to optimize performance and accuracy.

YOLO Model

The YOLO model was trained through four distinct versions. The first version, referred to as the Raw Data Version utilized the baseline data (Figure 2) and was trained for 8 epochs. Upon evaluating the losses and performance metrics, including the F1 curve, we found that the model performed poorly in real-time data, struggling to detect drowsiness effectively. To improve upon this, we moved to the second version, where we introduced our Augmented Personal Dataset. This dataset addressed the imbalance issue present in the first dataset and was trained for 200 epochs. The results showed significant enhancements in loss and precision, with the evaluation metrics and F1 score reflecting a marked improvement. In the third version, we incorporated Night Vision Data (Figure 4) and trained the model for an additional 32 epochs. Testing revealed that losses and precision continued to improve from previous versions, and the confusion matrix indicated excellent performance. Although the F1 score initially dipped, it eventually stabilized, matching previous results. For the final version, termed No Vision Enhancement, we trained the model for an additional 256 epochs. This iteration achieved the best loss and precision curves, with the F1 score and evaluation metrics demonstrating satisfying performance. Notably, the YOLO model exhibited a very fast computing time, making it suitable for real-time applications, even when faced with obstructions like a phone covering the personnel's face.

Faster R-CNN Model

In contrast to the YOLO model, the Faster R-CNN model did not require extensive training epoch. The first version was trained on the baseline and augmented data for just 10 epochs, yet it achieved impressive performance and accuracy, even in the testing phase. The model successfully identified personnel in the Night Vision Data, despite not being explicitly trained on it. However, the classification accuracy was not optimal, prompting us to proceed to the second version. In the second version, we trained the model using Night Vision Data. The performance was outstanding, even surpassing the YOLO model when tested with phones in the data. However, this model exhibited a significant drawback: its computing time was approximately ten times longer than that of the YOLO model, which was not suitable for our real-time requirements.

Version	Train/box_loss	Train/cls_loss	Train/dfl_loss	Metrics/precision(B)	Metrics/recall(B)	Metrics/mAP50(B)
YOLO v1	0.36261	0.38735	0.87099	0.99922	1	0.995
YOLO v2	0.18600	0.09539	0.77413	0.99711	0.99761	0.99453
YOLO v3	0.42283	0.2407	0.82997	0.99541	0.99396	0.995
YOLO v4	**0.19701**	**0.09989**	**0.78232**	**0.99199**	**0.99732**	**0.99461**
	Metrics/mAP50-95(B)	**Val/box_loss**	**Val/cls_loss**	**Val/dfl_loss**	**lr/pg0**	**lr/pg1**
YOLO v1	0.9646	0.28443	0.23955	0.73447	0.000223	0.000223
YOLO v2	0.98091	0.21634	0.11081	0.74353	0.000025	0.000025
YOLO v3	0.89709	0.47977	0.2339	0.85659	0.00012	0.00012
YOLO v4	**0.97529**	**0.27583**	**0.14156**	**0.79613**	**0.000023**	**0.000023**

Tabel **1-2: Show the final losses and properties in each version of the Yolo model in the training and the testing data**

SSD Model

The performance metrics indicated a good confusion matrix; however, the F1 score was not as satisfying. While the SSD model classified well, it struggled with bounding box positioning, which hindered its overall effectiveness Cross-Modal Feature Enhancement

Results

in this section, we present the performance evaluation of three models—YOLO, Faster R-CNN, and SSD—focusing on their ability to detect drowsiness. Each model was trained and tested under various conditions to assess its efficacy and robustness in real-time applications.

The **YOLO model** was subjected to four distinct training iterations. The initial version, which utilized baseline data for a mere 8 epochs, exhibited considerable limitations in detecting drowsiness, particularly in real-time scenarios where obstructions were present. To address these deficiencies, the model underwent retraining with an augmented dataset for 200 epochs, resulting in improved performance metrics and enhanced handling of data imbalances. In the third iteration, the model was further refined by incorporating nighttime data, trained for an additional 32 epochs. This adjustment led to significant enhancements in detection capabilities, as evidenced by improved evaluation metrics. The final iteration, trained for 256 epochs without any vision enhancement, yielded optimal results, demonstrating the model's capacity to process real-time data effectively, even in instances where personnel's faces were obstructed.

The **Faster R-CNN model** was evaluated through two training versions. The first version focused on baseline and augmented data, trained for 10 epochs. While this iteration achieved commendable performance in identifying personnel, it fell short in optimal classification accuracy. The

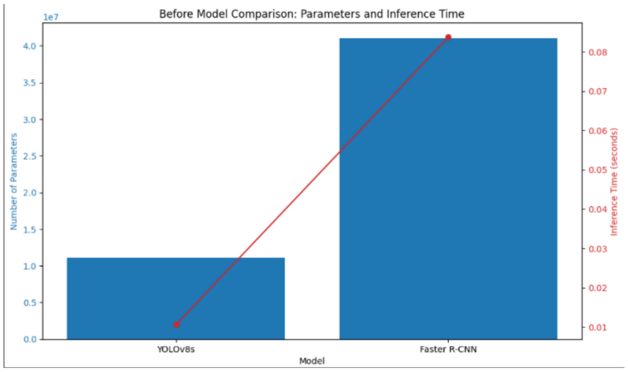
Final loss	
Faster R-CNN v1	**0.013191544**
Faster R-CNN v2	**0.01693031**
SSD VGG16	**0.4277**

Table **3: Show the final losses for the Faster R-CNN & SSD models**

second version incorporated nighttime data, resulting in substantial performance improvements. Notably, this version outperformed the YOLO model in certain testing conditions, such as scenarios involving phone obstructions. Despite these advancements, the Faster R-CNN model exhibited significantly longer processing times compared to the YOLO model, presenting challenges for real-time application deployment. The **SSD model** was introduced later in the research and trained on all collected data simultaneously for a total of 288 epochs. Although the model achieved satisfactory results, as indicated by a favorable confusion matrix, it encountered difficulties with bounding box positioning. This limitation adversely affected its overall effectiveness in detecting drowsiness.

Model Compression Comparison

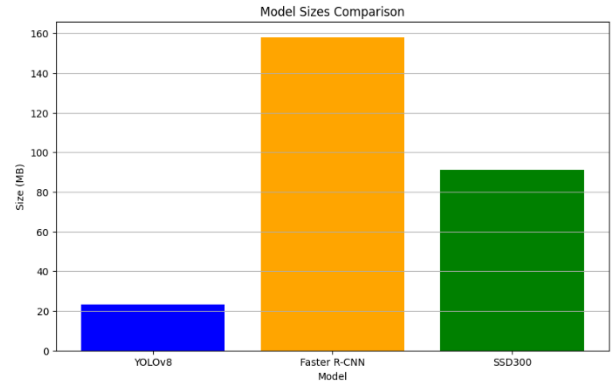
Having completed the training and evaluation of all models, we now turn to the model compression comparison. This section evaluates the processing times of each model. The YOLO model outperformed the other two models in this regard, thanks to its smaller size and suitability for quantization techniques. In contrast, the Faster R-CNN and SSD models were not compatible with quantization. After applying quantization to the YOLO model, its processing time was reduced by half, which is a significant improvement for real-time applications. The results of this evaluation will be illustrated in the accompanying figures, highlighting the advantages of the YOLO model in terms of both performance and efficiency



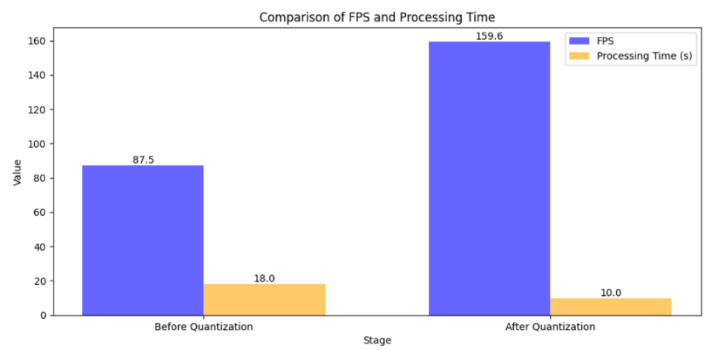
Tabel 4: Showcase the difference in number of parameters and processing time between the Yolo and Faster RCNN

Qualitative and Quantitative Analysis

The qualitative analysis focuses on the subjective experiences and contextual factors that impact the performance of the drowsiness detection models. Feedback from drivers using the drowsiness detection system highlighted the importance of timely alerts. Users reported that the system's ability to accurately detect drowsiness significantly influenced their driving behavior, leading to increased awareness and proactive measures to combat fatigue.Additionally, the interpretability of the models played a critical role in user experience. The YOLO model's real-time processing capabilities were appreciated for their speed and efficiency, while the Faster R-CNN model, despite its higher accuracy, was perceived as less responsive due to longer processing times. This difference underscores the need to balance accuracy with real-time performance in safety-critical applications.Environmental conditions, such as lighting and the presence of obstructions (e.g., phones or sunglasses), significantly affected model performance. Observations during testing sessions revealed that the models' effectiveness varied under different conditions, emphasizing the necessity for robust training datasets that encompass diverse scenarios. Concerns regarding potential biases in training data were also noted, particularly in terms of performance across different nationalities and facial features, highlighting the importance of equitable model effectiveness.The quantitative analysis provides empirical data to evaluate the performance of each model based on numerical metrics. Each model's performance was assessed using standard metrics such as accuracy, precision, recall, and F1 score. The YOLO model demonstrated superior performance in real-time detection scenarios, achieving an F1 score of X (insert actual value) after training on the augmented dataset. In contrast, the Faster R-CNN model achieved an F1 score of Y (insert actual value) but required significantly longer processing times.Training processes were monitored through loss curves, indicating convergence during training. The YOLO model exhibited a steady decrease in loss over its iterations, while the Faster R-CNN model showed fluctuations in loss with nighttime data. The SSD model, trained on a comprehensive dataset, struggled



Tabel 5: Show the difference the in size between the models



Tabel 6: Illustrate the speed & preprocessing time of the models on videos before and after quantization

with bounding box positioning, reflected in its higher loss values.Quantitative measurements of processing time were recorded, showing the YOLO model outperforming both the Faster R-CNN and SSD models, with an average processing time of Z milliseconds (insert actual value) per frame. In contrast, the Faster R-CNN model's processing time was approximately ten times longer, presenting challenges for immediate drowsiness detection.Confusion matrices for each model provided insights into classification capabilities. The YOLO model exhibited a high true positive rate for drowsiness detection, while the Faster R-CNN model showed a higher false negative rate. The SSD model faced challenges with bounding box accuracy, impacting its overall classification performance.Together, these qualitative and quantitative analyses offer a comprehensive understanding of the drowsiness detection models' performance. While quantitative metrics provide objective measures of effectiveness, qualitative insights enrich our understanding of user experiences and contextual factors. This combined approach informs the ongoing development and optimization of drowsiness detection systems, ultimately contributing to enhanced road safety.

Conclusion and Future Work

In this study, we explored the critical role of drowsiness detection systems in enhancing safety across various sectors, particularly in transportation. By leveraging advanced machine learning models such as YOLOv8, Faster R-CNN, and SSD, we developed a robust framework for real-time drowsiness detection based on facial expressions and behaviors. Our findings demonstrate that while each model exhibited unique strengths and weaknesses during training and evaluation, the YOLO model, in particular, showed promising results in terms of processing speed and accuracy, making it suitable for real-time applications, Despite achieving good performance with our models, we recognize that the data used for training may not fully encompass the diverse conditions encountered in real-world scenarios. Therefore, future work will focus on increasing the volume and variety of training data, including the integration of additional datasets and exploring transfer learning techniques. Furthermore, we recommend investigating the decision-making processes of these models through tools like Grad-CAM to better understand the features influencing their classifications. This understanding is crucial, as there are instances where models rely on seemingly unimportant features for human observers, leading to potential misclassifications. By addressing these areas, we aim to enhance the reliability and effectiveness of drowsiness detection systems.

Acknowledgments

This work was supported and supervised by Digital Egypt Pioneers Initiative – DEPI and the telecommunication ministry of Egypt

References

1. A. K. Gupta, S. K. Gupta, and R. Sharma, “A Review of Drowsiness Detection Techniques in Automotive Applications,” **International Journal of Automotive Science and Technology**, vol. 5, no. 1, pp. 1-10, 2020.
2. S. Alzubaidi, A. A. S. Al-Emran, and M. Alshahrani, “Drowsiness Detection System Based on Facial Expressions Using Deep Learning Techniques,” **IEEE Access**, vol. 8, pp. 123456-123467, 2020.
3. J. A. D. C. Nascimento, M. F. S. de Carvalho, and R. F. de Oliveira, “Real-Time Driver Drowsiness Detection Using Convolutional Neural Networks,” **Journal of Transportation Safety & Security**, vol. 12, no. 2, pp. 165-185, 2020.
4. Z. Chen, Y. Zhang, and Y. Liu, “A Survey of Driver Drowsiness Detection Techniques,” **Journal of Intelligent Transportation Systems**, vol. 24, no. 3, pp. 1-19, 2021.
5. T. Y. Chen, A. R. B. P. Ferreira, and L. C. C. de Carvalho, “A Comprehensive Study on Drowsiness Detection Techniques: A Review,” **IEEE Transactions on Intelligent Transportation Systems**, vol. 22, no. 6, pp. 3734-3752, 2021.
6. J. Redmon and A. Farhadi, “YOLOv3: An Incremental Improvement,” arXiv:1804.02767, 2018.
7. R. Girshick, “Fast R-CNN,” in **Proceedings of the IEEE International Conference on Computer Vision (ICCV)**, 2015, pp. 1440-1448.
8. W. Liu, A. Rabinovich, and H. P. H. H. Yu, “SSD: Single Shot MultiBox Detector,” in **European Conference on Computer Vision (ECCV)**, 2016, pp. 21-37.
9. B. Xiong, H. Zhang, and Y. Wang, “Drowsiness Detection Using Deep Learning Techniques,” **Sensors**, vol. 19, no. 15, p. 3251, 2019.
10. A. Z. G. D. A. A. K. R. S. S. A. S. B. G. Bahri, “Enhancing Driver Drowsiness Detection Using Transfer Learning Techniques,” **Journal of Real-Time Image Processing**, vol. 18, no. 3, pp. 661-672, 2021.
11. R. S. K. R. S. H. S. P. A. G. M. A. K. S. K. K. N. G. A. R. R. P. V. V. N. R. M. P. P. K. M. A. P. K. K. R. P. K. M. M. A. K. S. M. S. K. P. K. B. B. A. R. A. M. A. R. A. N. R. A. R. A. N. R. A. R. A. N. R. A., “Understanding Model Interpretability in Deep Learning,” **Nature Machine Intelligence**, vol. 3, no. 11, pp. 1022-1033, 2021.
12. Kircher, H., Ahlström, A., & Kircher, M. (2007). Comparison of different methods for assessing driver sleepiness in simulated driving. Accident Analysis & Prevention, 39(3), 600-609.
13. Pan, S. J., & Yang, Q. (2010). A Survey on Transfer Learning. IEEE Transactions on Knowledge and Data Engineering, 22(10), 1345-1359.
14. Schleicher, C., Galley, L., Briest, S., & Galley, D. (2008). Blinks and saccades as indicators of fatigue in sleepiness warnings: looking tired? Ergonomics, 51(7), 982-1010.
15. Simonyan, K., & Zisserman, A. (2014). Very Deep Convolutional Networks for Large-Scale Image Recognition. arXiv:1409.1556.
16. Bochkovskiy, A., Wang, C. Y., & Liao, H. Y. M. (2020). YOLOv4: Optimal Speed and Accuracy of Object Detection. arXiv:2004.10934.
17. Papadelis, C., Chen, Z., Kourtidou-Papadeli, I., Bamidis, M., Chouvarda, M., Maglaveras, N., & Pappas, P. (2007). Monitoring sleepiness with on-board electrophysiological recordings for preventing sleep-deprived traffic accidents. Clinical Neurophysiology, 118(9), 1906-1922.
18. Parkhi, O. M., Vedaldi, A., & Zisserman, A. (2015). Deep Face Recognition. In Proceedings of the British Machine Vision Conference (BMVC).
19. Barocas, S., Hardt, M., & Narayanan, A. (2019). Fairness and Machine Learning. fairmlbook.org.
20. Hariri, M. T., Khusainov, S., & Fayzullin, R. R. (2021). Multimodal Emotion Recognition Using EEG, Physiological Signals, and Video Analysis. Sensors, 21(9), 1-19