Analysis of Computer Vision-Based Physiological Indicators for Operator Fatigue Detection

Batol Hamoud, Walaa Othman, and Nikolay Shilov
Saint Petersburg Federal Research Center of the Russian Academy of Sciences (SPC RAS)
St. Petersburg, Russia
bkhamud, walaa_othman@itmo.ru, nick@iias.spb.su

***Abstract*—Detecting human mental fatigue is crucial because it significantly impacts work efficiency, especially in system operation control. In our earlier research, we developed a video-based system for detecting mental fatigue using physiological indicators analyzed through deep learning models. However, this approach is computationally expensive and requires processing of videos with various machine learning models to estimate multiple vital signs and facial features. To improve its efficiency, this paper introduces feature importance techniques that identify the most significant indicators of fatigue. This not only enhances our model's performance but also provides insights into the factors contributing to fatigue. Additionally, we transitioned from using traditional machine learning techniques (i.e. Random Forest) to a more advanced architecture such as the Tabular Transformer. This shift offers several advantages, including better generalization and effective handling of tabular data, resulting in an impressive accuracy of 89%. Our findings promise to deliver more accurate and reliable assessments of mental fatigue in real-world applications.**

***Keywords*—operator fatigue detection; computer vision; physiological indicator, features importance, tabular transformers.**

I. INTRODUCTION

Mental fatigue is becoming a serious issue in the advanced work environment. When employees are given tasks that require higher levels of complexity and the cognitive demands on them build up, the results can be disadvantageous in terms of productivity, decision-making, and personal well-being [1]. However, with the advancement of technology, organizations now have the ability to proactively monitor and address mental fatigue, ensuring their employees remain engaged, focused, and resilient.

The advantages of monitoring mental fatigue in the workplace are significant. It enhances both individual well-being and work performance, while also boosting the organization's success and reducing the risk of errors. Employees who have healthy mental states are more likely to make good decisions, work with their colleagues more effectively, all of which positively impact the organization's financial and social performance.

The primary aim of fatigue detection approaches is to identify the early warning signs of mental fatigue and subsequently alert the individual to this potentially dangerous state. The majority of these techniques use advanced machine learning classifiers or deep learning models to construct highly effective models capable of accurately detecting signs of fatigue [2]. Much of the existing research on fatigue monitoring has focused on identifying signs of fatigue during cognitively demanding tasks, such as driving [3]. However, there is some research and effort to detect fatigue in more natural situations, where individuals are not actively engaged in challenging cognitive activities [4]. These non-intrusive techniques often utilize remote or webcam-based eye-tracking to monitor variations in an individual's pupil response, blink rate, and eye movement patterns, which can serve as reliable indicators of fatigue. However, there are several drawbacks to depending exclusively on these features. The individual variability in these physiological measures can complicate finding a consistent and generalizable thresholds or patterns for detecting fatigue. Additionally, these features can be affected by environment conditions, such as lighting, task demands, and emotional states, which might not be directly tied to fatigue, resulting in false positives or negatives. Fatigue is a complex phenomenon, and these eye-related and mouth-related features may only reflect a limited aspect of the fatigue experience, potentially overlooking other significant factors.

This motivated us to propose our previous machine learning-based approach to assessing human mental fatigue using video recordings of working operators [5]. The incorporated features, estimated by deep learning models, encompass head movement, vital signs, and eye and mouth states. This approach eliminates the need for manual detection or external sensors, offering a more practical and scalable solution for organizations. Our results demonstrated that the Random Forest technique consistently achieved high accuracy (98%) and F1 score (94%) in detecting mental fatigue.

Our approach to fatigue detection can be quite resource-intensive, as it typically involves processing videos with different machine learning and deep learning models to estimate multiple vital signs and both eyes and mouth states. To tackle this issue and enhance our model's efficiency, we opted in this paper to utilize feature importance techniques to highlight the most significant features that play a crucial role in fatigue detection. By excluding less critical features, we can lower the computational demands of our model and concentrate on the primary indicators of fatigue. This not only boosts our model's performance but also helps us to better understand the factors that lead to fatigue. Additionally, we explored moving from Random Forest, to a more sophisticated architecture such as the Tabular Transformer. The benefits of adopting the Tabular Transformer include improved generalization, effective management of tabular data, and better performance, all of which can contribute to more precise and dependable fatigue assessments in practical scenarios.

This research makes important contributions to the understanding of mental fatigue as follows:

1) Finding Key Physiological Features: We identified Heart Rate, Oxygen Saturation, Blood Pressure (both Systolic and Diastolic), and Average Pitch as the most important physiological signs for measuring mental fatigue. By focusing on these key features, we improved the accuracy of fatigue detection while also making the process less resource-intensive.
2) Improving Analytical Methods: We used several different methods to analyze which features are most important, including SHAP analysis and Permutation Importance. This approach shows the value of using various techniques to get a complete picture of what affects mental fatigue, helping to clarify the complexities involved in its assessment.
3) Using Advanced Deep Learning Models: We switched to a tabular transformer model, which led to better performance, achieving an accuracy of 89%. This shows that transformer models can effectively handle complex data relationships, suggesting new paths for future research in this area.

The paper is structured in three main sections. Section II, describes numerous studies on the impact of fatigue on physiological indicators, as well as various approaches to feature estimation and the datasets used for fatigue and drowsiness detection. Section III, provides an overview of our previous work and presents our framework, which includes using feature importance techniques and a specific procedure to obtain a more effective fatigue detection model. Finally, Section IV outlines the experiments conducted and the results of employing the proposed approach enhancement.

II. Literature review

In this section, we explore research that examines the connection between fatigue and various physiological indicators. This will be followed by an overview of the models and methodologies employed to derive these indicators, along with a brief review of the datasets collected for the purpose of detecting fatigue and drowsiness.

A. Relationship between mental fatigue and physiological indicators

In the domain of mental fatigue detection, researchers tend to rely on correlation between mental fatigue and various physiological indicators. However, this correlation is not robust and may differ significantly for particular people. Among these physiological parameters are vital signs, such as blood pressure, heart rate, and respiratory rate, as well as head movement and ocular characteristics, including pupil diameter, blink rate, and others. This section provides an overview of existing research exploring the relationship between mental fatigue and these physiological parameters, highlighting their relevance in detecting and understanding cognitive fatigue.

1) Vital Signs: Vital signs are essential for the detection of mental fatigue. For instance, a study investigated the impact of inducing a hypnotic state of fatigue on the respiration rate and blood pressure of labor employees [6]. The researchers identified significant changes in these physiological parameters after participants completed demanding tasks. Notably, blood pressure exhibited a marked upward trend, while respiration rate decreased. In another investigation, features related to heart and respiratory rates were utilized to train classifiers aimed at detecting both physical and mental fatigue [7]. The authors employed Random Forest [8] and causal Convolutional Neural Network (cCNN) [9]. Additionally, in a study involving 14 participants performing a complex task in a flight simulator, these indicators were significantly influenced by mental workload [10]. During the high-stress landing phase, systolic and diastolic blood pressures were elevated compared to other phases, and respiratory activity slowed but deepened after landing, contributing to HRV changes. These physiological changes, combined with subjective reports and task performance, confirmed that landing represents a period of heightened mental workload for pilots. Moreover, a study evaluating the cardiovascular and subjective stress responses to combined physical and mental workloads revealed significant findings regarding blood pressure changes [11]. The introduction of mental stressors during standardized computer work resulted in a notable increase in blood pressure compared to baseline measurements, with this elevation persisting even after the stressors were removed. Specifically, diastolic pressure continued to rise during subsequent control sessions, indicating a prolonged impact of the experienced stress.

2) Head Pose: Head movement, including nodding and sudden shifts in position, has been a focal point in fatigue analysis. An innovative driver fatigue detection system [12] integrates a residual channel attention network (RCAN) with head posture estimation, utilizing Retinaface for facial localization and recording five key facial landmarks. Additionally, a study [13] employed a single-axis MEMS accelerometer alongside statistical and fractal analysis to discern fatigue characteristics. Another approach [14] in this field utilized the XSENS motion capture system to monitor drivers' head posture motions, combined with a modified bidirectional long short-term memory (BiLSTM) deep neural network for sequence-to-sequence classification. Experiments conducted on 15 subjects using a driver-in-loop simulator demonstrated the effectiveness of this method, achieving high performance. Notably, the way individuals nod can signify varying cognitive states. Research indicates that up-nods may reflect a cognitive shift following other people contribution or conversation (such as colleagues), whereas down-nods suggest stability in cognitive state [15]. This differentiation in nodding behavior offers valuable insights into the cognitive processes underlying communication and social interaction.

3) Eyes and Mouth Features: There is a significant correlation between the features of the eyes and mouth and the experience of mental fatigue. Analyzing ocular attributes, including pupil size, blink frequency, and mouth movements such as yawning and mouth openness, has proven effective in detecting signs of fatigue and drowsiness [16]. The authors of [17] employed a pre-trained model using Histogram-Oriented Gradients (HOG) [18] and a linear Support Vector Machine (SVM) to detect facial features and calculate eye aspect ratio (EAR), mouth opening ratio (MOR), and nose length ratio (NLR) to detect the drowsiness. They applied adaptive thresholding for initial classification of behaviors like blinking and yawning, followed by machine learning algorithms to distinguish between drowsy and non-drowsy states. Research has shown that cognitive workload is closely linked to changes in ocular behavior, including increased pupil dilation, higher blink frequency, reduced average fixation duration, and slower saccadic movements [19]. Eyes and mouth related metrics have been consistently associated with mental effort, highlighting their importance in evaluating both fatigue and cognitive demands.

4) Summary of the Relationship Among Mental Fatigue and Physiological Indicators: In discussing the key findings regarding physiological changes experienced by individuals after engaging in cognitively demanding tasks, it is noted that those exhibiting signs of fatigue show distinct ocular characteristics. These include an increased blink rate, larger pupil size, decreased saccadic velocity, and a higher ratio of eye closure. Additionally, vital signs indicate that fatigue is associated with a decrease in respiratory rate and an elevation in blood pressure. Furthermore, fatigued individuals tend to exhibit more frequent jaw opening [20] and head nodding compared to energetic people.

B. Physiological Indicator Estimation Based on Computer Vision

The concept of acquiring physiological indicators has attracted significant interest within the domains of computer vision and deep learning. This interest stems from the potential applications of these estimations, which may facilitate the reduction or elimination of reliance on traditional medical devices. Furthermore, such advancements could enable the monitoring of individuals in scenarios where attaching devices to the body is impractical, such as during driving. In this section we will present the state of the art approaches devoted for this purpose.

1) Vital Signs: This study [21] introduces a novel approach to heart rate estimation using photoplethysmography (PPG) that combines hybrid artifact removal, signal reconstruction, and deep learning techniques. This method effectively addresses motion artifacts in dynamic environments, outperforming traditional methods. Another article [22] presents a novel signal quality ranking and fusion (SQRF) approach for improving non-contact heart rate (HR) estimation using remote photoplethysmography (rPPG). By analyzing multiple regions of interest on the face and employing wavelet synchrosqueezed transform to enhance signal stability, the method significantly reduces mean absolute error (MAE) by up to 58.7% compared to traditional single-region methods. As for blood pressure, study [23] introduces a non-contact, video-based blood pressure estimation method (V-BPE), addressing the limitations of traditional contact-based devices. V-BPE leverages Pulse Transit Time and subject-specific parameters, such as blood vessel length, derived through computer vision and demographic data, to estimate blood pressure. Moreover, research described in [24] proposes RBP-CNN, a network for non-contact blood pressure (BP) estimation using remote photoplethysmography (rPPG) from facial videos. This network uses residual convolution, local and global attention mechanism, By extracting features like blood volume pulse (BVP), heart rate (HR), age, and BMI from facial videos, the authors used Random Forest to fuse these features to identify implicit BP-related features. Additionally, many research efforts focused on respiration rate (RR) estimation, for example, study [25] introduced ACTNet, a dual-branch network for non-contact respiratory rate estimation using facial videos. Combining a CNN for capturing subtle facial color changes with a Transformer for long-term temporal modeling, ACTNet effectively fuses local features and global information for achieving effective remote RR estimation. Furthermore, paper [26] presented an end-to-end deep learning method for estimating respiratory rate from thermal video data, utilizing a detection transformer to identify the facial region of interest. By employing 3D convolutional neural networks and bi-directional long short-term memory layers, the method effectively estimates respiratory signals while addressing phase shifts with a novel loss function. Finally, oxygen saturation estimation has received considerable attention in the field of deep learning, similar to the other vital signs. for instance, study [27] introduced CCSpO2Net, a camera-based contactless oxygen saturation (SpO2) estimation model designed for clinical and laboratory settings. By combining a spatial feature extractor and a global temporal feature extractor, along with pixel-level skin region detection using the SAM model, CCSpO2Net effectively estimates SpO2 from facial video data. In addition, article [28] presented a novel method for estimating oxygen saturation (SpO2) levels using smartphone cameras and video-based photoplethysmography (PPG). The proposed framework analyzes 20-second facial videos through a cloud-based server employing deep learning techniques to extract remote PPG signals and predict SpO2 levels using a Support Vector Regression (SVR) model.

2) Head Pose: A significant amount of research has been dedicated to head pose estimation due to its critical role in monitoring and understanding an individual's state. The authors of [29] introduced a robust method for head pose estimation using a multi-loss convolutional neural network trained on the 300W-LP dataset. Unlike traditional approaches relying on keypoint detection and 2D-to-3D correspondence, this method directly predicts Euler angles (yaw, pitch, roll) from image intensities through joint classification and regression. Another study [30] introduced HyperFace, a deep convolutional neural network (CNN) algorithm for simultaneous face detection, landmark localization, head pose estimation, and gender recognition. HyperFace enhances performance by fusing intermediate CNN layers and leveraging multi-task learning to benefit from the harmony among the tasks. Two variants are proposed: HyperFace-ResNet, which builds on ResNet-101 for improved accuracy, and Fast-HyperFace, which prioritizes speed using a high-recall face detector. Moreover, article [31] proposed a method for accurate 3D head pose estimation using a commodity depth camera. The approach registers a morphable face model for depth data using particle swarm optimization (PSO) combined with the iterative closest point (ICP) algorithm, eliminating the need for explicit initialization or training. This method handles large pose angles and partial occlusions by adapting to visible facial regions and generalizes across different depth sensors.

3) Eye and Mouth states: Eye closure ratio and yawning detection has captured a lot of interest due to their significant role in characterizing states of drowsiness and fatigue. Study [32] presented a novel method for accurate and robust yawn detection. The approach uses a 3D deep learning network with Low Time Sampling rate (3D-LTS) to extract spatial and temporal features for subtle facial action recognition, along with a keyframe selection algorithm to eliminate redundant frames and outliers. Another approach [33] was to extract the mouth region using a face and landmark detector, then uses a pre-trained CNN for spatial features and a combination of 1D-CNN and bi-directional LSTM (Bi-LSTM) for temporal yawn evaluation. Regarding the eyes features, this work [34] presented a real-time algorithm for eye blink detection using video from standard cameras. It relies on robust facial landmark detectors to estimate the eye openness level through the Eye Aspect Ratio (EAR) in each frame. Blinks are identified either via an SVM classifier analyzing EAR patterns over time or a hidden Markov model coupled with a state machine. A further examination of eye aspect ratio measurement is presented in [35] where Viola-Jones method was utilized for facial detection to accurately identify the position of the right eye. They systematically collected six coordinates that define the eye by moving clockwise around the eye region, starting from the left corner. Following this, they applied a formula introduced in [34] to calculate the eye aspect ratio (EAR) and established a threshold of 0.3 for the aspect ratio within their system.

C. Datasets for fatigue and drowsiness detection

This section presents an overview of several commonly used datasets specifically created for detecting fatigue and drowsiness. Our research focuses on video datasets being recorded using standard cameras without physical sensors.

The YawDD dataset [36] consists of two publicly accessible sub-datasets containing RGB videos. In these videos, participants simulate driving scenarios while the vehicle remains stationary. The first sub-dataset features a camera located under the front mirror, resulting in 322 videos (three or four video per suject) with and without eyewear, representing a range of ethnic backgrounds. The second sub-dataset comprises 29 videos captured from a camera mounted on the dashboard in front of the driver, showcasing various mouth states, including normal, talking/singing, and yawning.

The researchers behind [37] developed a multimodal database named DROZY, which was created using data collected from 14 participants. These individuals participated in three consecutive sessions of a modified psychomotor vigilance test (PVT), originally proposed by [38], under conditions of increasing sleep deprivation. The PVT is designed to provide an objective assessment of vigilance and, by extension, drowsiness. For every participant and each PVT session, the database contains synchronized raw data, including polysomnography (PSG) signals, Karolinska Sleepiness Scale (KSS) scores [39], PVT results (such as reaction times), and near-infrared (NIR) intensity and range images of the face.

The SUST-DDD dataset [40] consists of 2074 videos recorded by participants using their mobile phone cameras in real-world driving situations, capturing both fatigued and normal states. Nineteen participants were asked to record videos using their personal phones placed in front of the driver's seat whenever they felt drowsy or alert. Importantly, participants were not instructed to perform specific actions while driving, ensuring the naturalness and safety of the driving experience.

The authors of [41] developed the Licensed Crane Operators dataset, which includes videos gathered through interviews with five experienced crane operators. These videos capture three specific behavioral states: alertness, reduced vigilance, and fatigue. The recordings were taken from multiple camera angles and conducted in a variety of settings, such as computer workstations, simulated or real driving environments, and simulated crane operations. The participants display a wide range of facial features, behaviors, and ethnic diversity.

In conclusion, the integration of physiological indicators, advanced computational approaches, and high quality datasets offers a sophisticated framework for understanding and detecting mental fatigue. This approach not only enhances our ability to monitor cognitive states but also paves the way for practical applications in critical fields such as transportation, healthcare, and workplace safety. As we mentioned, physiological indicators, such as vital signs, head movements, and ocular and mouth features, have shown strong correlation with mental fatigue. These indicators provide a foundation for understanding mental fatigue and overload. Furthermore, advancements in computer vision and deep learning have enabled non-contact estimations and real-time monitoring of these indicators. These methods eliminate the need for traditional medical devices, making them particularly valuable in scenarios where physical sensors are impractical, such as driving or workplace monitoring. In addition, the availability of robust datasets specifically designed for fatigue and drowsiness detection further strengthens this framework. Therefore, in this paper, we aim to enhance our approach of detecting mental fatigue [5] using indicators estimated through our deep learning models [42]–[47] applied to a video dataset that provide a more reliable fatigue metric [48]. This improvement will be achieved by focusing on the feature importance techniques and exploring how eliminating some features can affect the accuracy. Our objective is to maintain the high performance of our model while also reducing its computational cost.

III. Approach

The section provides a quick review of the proposed approach aimed at enhancement, including a brief description of the dataset utilized and the deep learning models applied to estimate the physiological indicator. Furthermore, it outlines the enhancement pipeline, detailing the techniques employed for the analysis of feature importance, the process of feature selection, the development of models based on the refined feature set, and the improvement of the resulting model through feature transformation.

A. Dataset Description

In our previous work [5], the OperatorEYEVP dataset, introduced in [48], was utilized for the development of a fatigue detection system. This dataset includes video recordings of ten distinct individuals engaged in various activities at three different times of the day over a span of eight to ten days. The experimental protocol began each day with a sleep quality survey conducted prior to the morning session. This was followed by the VAS-F questionnaire, a choice reaction time task (CRT), reading a scientific-style text, performing the ”Landolt rings” correction test, playing the ”Tetris” game, and a second CRT. On average, the total duration of these sessions was approximately one hour. To evaluate fatigue levels, we focused on the results of the Landolt test, specifically mental performance, as it effectively captures the cognitive and attentional dimensions associated with fatigue. Based on the experiments described in [5], a threshold for mental performance was established. Values below this threshold were classified as indicative of a fatigue state, while values above it represented a non-fatigued state.

As previously mentioned, our primary objective is to detect fatigue through the analysis of physiological indicators estimated via deep learning and computer vision techniques. To achieve this, we annotated each minute of the aforementioned dataset videos with several key indicators obtained using computer vision models, including blood pressure, heart rate, oxygen saturation, and respiratory rate. Additionally, we estimated head pose using Euler angles (roll, pitch, yaw), calculated ratios of eye closure and mouth openness, and assessed characteristics related to breathing patterns, such as rhythmicity and stability.

B. Deep Learning Models

This section provides a brief overview of the models used for the estimation of physiological indicators.

1) Respiratory Rate and Breathing Characteristics: The respiratory rate model [42] involved detecting chest keypoints with OpenPose, followed by displacement analysis using the SelFlow neural network. The displacement data was processed through signal processing techniques to enhance accuracy, resulting in a mean absolute error of 1.5 breaths per minute. However, the model's effectiveness is limited by its inability to accommodate body movement, making it unsuitable for scenarios where the subject is in motion, such as walking or driving.

2) Heart Rate: The heart rate estimation model [43] utilized in this study follows a structured approach, beginning with the extraction of facial regions and processing through a Vision Transformer model. The outputs are then analyzed using a block structure that incorporates various layers to calculate heart rates through a weighted averaging scheme. Although this model shows improved accuracy over previous methods, it may yield inaccurate results for subjects with extreme heart rates due to insufficient training data.

3) Blood Pressure: The blood pressure estimation model [44] begins with the identification of the left and right cheeks as Regions of Interest (ROIs) in video frames. A Convolutional Neural Network (CNN) was employed to extract spatial features, using EfficientNet architectures for systolic and an ensemble approach for diastolic blood pressure estimation. The outputs were then processed through a Long Short-Term Memory (LSTM) network to capture temporal features, followed by fully connected layers to derive blood pressure values. The model achieved mean absolute errors of 11.8 mmHg for systolic and 10.7 mmHg for diastolic blood pressure, with accuracies of 89.5% and 86.2%, respectively. However, challenges included a lack of diversity in the training dataset, particularly regarding skin tones, which affected the model's predictive accuracy for individuals with darker skin.

4) Oxygen Saturation: The estimation of oxygen saturation in this study was based on the method proposed in [45], which involved several key steps. Initially, the face region was extracted using the 3DDFA_V2 framework, followed by feature extraction through the pre-trained VGG19 model. The resulting features were then processed using the XGBoost algorithm to estimate oxygen saturation values. The model achieved Mean Absolute Errors (MAE) of 1.17% and 0.84% on two test datasets, respectively. However, a significant challenge was the limited representation of SpO2 levels below 85 in the training samples, which could hinder accuracy for individuals with specific health conditions.

5) Head Pose: The head pose estimation model [46] starts with face detection using the YOLO Tiny framework. Following this, a 3D face reconstruction technique aligns facial landmarks to ensure accurate detection, even for partially visible features. The analysis of landmark transitions across frames allows for the calculation of Euler angles, providing head pose information. However, the model is limited to estimating head angles up to 70 degrees, which restricts its applicability in scenarios requiring detection of larger angles and results in slower performance compared to other detectors.

6) Eye and Mouth states: The eye state is determined using a trained model that processes the detected face from the FaceBoxes framework to indicate whether the eyes are open or closed. Mouth state detection is achieved through a modified MobileNet model [47], which boasts an accuracy of 95.20%. However, the model's reliance on a private dataset for validation may limit the generalizability of its findings to the broader population.

C. Previous proposed approach

Our proposed approach in [5], as shown in Figure 1, involves the extraction of physiological indicators from operator videos using a set of specialized models. These indicators encompass heart rate, respiratory rate, blood pressure, oxygen saturation, eye closure ratio, head pose, among others, and are calculated on a per-minute basis to provide continuous data for the fatigue detection process. Once extracted, these indicators serve as input features for our fatigue detection model, which evaluates the operator's fatigue state based on mental performance (AU) derived from the Landolt rings test. This test offers insights into the operator's cognitive state. Our hypothesis posits that mental performance diminishes as fatigue increases. To predict fatigue states, we assessed various methods, including Support Vector Classifier (SVC), logistic regression, Multi-Layer Perceptron (MLP), decision tree, XGBoost, and Random Forest. The results indicated that the Random Forest model achieved the highest F1 score of 0.947 for fatigue prediction based on vital signs.

This approach, while effective, is computationally intensive due to the reliance on deep learning models with complex architectures for estimating physiological indicators. As a result, it poses challenges for deployment on devices with limited computational resources. To address this limitation, the focus of this article is on reducing the number of features used, thereby minimizing the reliance on multiple models. Achieving this requires a thorough analysis of feature importance through various tests and techniques to identify the most impactful features for fatigue detection. The goal is to optimize the system's performance while evaluating how this reduction affects key metrics, such as accuracy and F1 score, ensuring the model remains effective despite the simplification.

D. Used Techniques for Features Importance Analysis

Feature importance methods help to understand which features contribute most to machine learning model's predictions. This section summarizes five common techniques that are used in this work: SHAP (Shapley Additive Explanations), Permutation Importance, Mutual Information (MI), Mean Decrease in Impurity (MDI), and Maximal Information Coefficient (MIC).

1) SHAP (Shapley Additive Explanations): SHAP [49] is based on game theory and provides a globally consistent way to compute feature importance. SHAP calculates feature importance by considering all possible combinations (subsets) of features and measuring how the prediction changes when a feature is added. Mathematically, for each feature, it computes the average marginal contribution of that feature across all possible feature subsets.

The SHAP value for a feature is computed as:

$$\phi_i = \sum_{S \subseteq F \setminus \{i\}} \frac{|S|!(|F|-|S|-1)!}{|F|!} [f(S \cup \{i\}) - f(S)] \quad (1)$$

where:

- F is the full feature set.
- S is a subset of features excluding i
- $f(S)$ is the model's prediction when only features in S are used.
- ϕ_i is the Shapley value representing feature i's contribution.

SHAP has several advantages:

1) Local and Global Interpretability: Individual predictions can be explained locally and it is possible to define features that are generally most important across the dataset.
2) Model-Agnostic: Works with any machine learning model (though optimized for tree-based models via TreeSHAP).
3) Fair: Based on a solid mathematical foundation from game theory.

2) Permutation Importance: Permutation importance [50] s a straightforward and model-agnostic technique used to measure how much each feature contributes to a model's predictive performance. It measures how shuffling a feature's values affects a model's performance. If a feature is important, randomizing its values should lead to a significant performance drop.

The importance score of feature j is:

$$I_j = \frac{1}{M} \sum_{m=1}^{M} \left(\text{Perf} - \text{Perf}_{\text{perm},j}^{(m)} \right) \quad (2)$$

where:

- $Perf$ is the original model performance.
- $Perf_{\text{perm},j}^{(m)}$ is the model performance after shuffling feature
- M is the number of permutation rounds.

Permutation Importance is a valuable technique due to its versatility and simplicity. It is model-agnostic, meaning it can be applied to any type of machine learning model. The method is intuitive and easy to implement. Moreover, it directly measures the real impact of each feature on the model's output by observing how performance changes when a feature is disrupted. However, it has limitations, such as being less effective when features are highly correlated and requiring significant computational resources for large datasets or complex models.

3) Mutual Information: Mutual information [51] measures the dependency between a feature and the target variable using information theory. It quantifies how much knowing a feature reduces uncertainty about the target.

Mutual Information (MI) is given by:

$$I(X,Y) = \sum_{x \in X} \sum_{y \in Y} p(x,y) \log \frac{p(x,y)}{p(x)p(y)} \quad (3)$$

where:

- $p(x,y)$ is the joint probability distribution of feature X and target Y

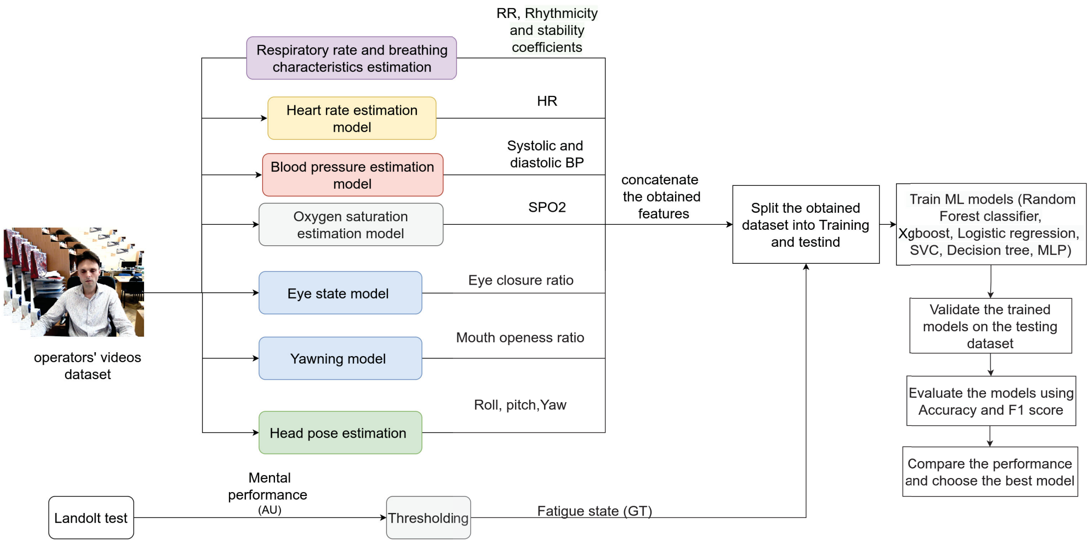

Fig. 1. Fatigue detection approach proposed in [5] .

- $p(x)$ and $p(y)$ are the marginal probabilities.

Mutual Information is a powerful method for capturing both linear and non-linear dependencies between variables, unlike simple correlations. It is model-agnostic and useful in feature selection, as it identifies the most informative features based on their dependency with the target variable. However, it has some limitations, such as being sensitive to how data is discretized, which can affect the results, especially for continuous features.

4) Mean Decrease in Impurity (MDI, Gini Importance): MDI [52]is a feature importance metric used primarily with tree-based models, such as Decision Trees and Random Forests. It measures the contribution of each feature to reducing the impurity of the nodes in a decision tree. Impurity is a measure of how mixed the target variable is in the node.

The MDI importance of a feature computed with an infinite ensemble of fully developed totally randomized trees and an infinitely large training sample is:

$$\operatorname{Imp}\left(X_m\right)=\sum_{k=0}^{p-1} \frac{1}{C_p^k} \frac{1}{p-k} \sum_{B \in \mathcal{P}_k\left(V^{-m}\right)} I\left(X_m ; Y \mid B\right) \quad (4)$$

where:

- $I(X_m; Y \mid B$ is the conditional mutual information of X_m and Y given the variables in B.
- $P_k(V^{-m})$ is the set of subsets of V^{-m} of cardinality k.
- V^{-m} denotes the subset $V \backslash \{X_m\}$.

MDI is a computationally efficient method well-suited for large datasets and tree-based models like Random Forests, Gradient Boosting Machines, and Decision Trees. It provides a global feature importance ranking, helping with feature selection and model interpretation. However, MDI has limitations, such as a bias towards numerical features, difficulty distinguishing between correlated features, and being specific to tree-based models, making it unsuitable for other model types like neural networks or support vector machines.

E. Maximal Information Coefficient (MIC)

MIC [53] detects both linear and nonlinear relationships between features and the target by adapting mutual information to large datasets. It is part of the Maximal Information-based Nonparametric Exploration (MINE) framework, designed to capture complex dependencies in data that other traditional methods like correlation might miss.

MIC equation is:

$$\operatorname{MIC}(X, Y)=\max_{(x, y) \in G(n)} \frac{I(X, Y)}{\log_2(\min(x, y))} \quad (5)$$

where:

- $G(n)$ is the set of all possible grid partitions of the data.
- $I(X, Y)$ is the mutual information.
- x and y represent bin sizes in the partition.

MIC is a powerful tool for detecting relationships between variables, It is model-agnostic, meaning it can be used with any type of data and does not require assumptions about data distribution or model type. MIC is particularly useful for datasets with complex structures and dependencies. However, it has some limitations, such as being sensitive to data discretization, which can impact results, and being computationally intensive, especially for large datasets

F. Feature Selection

To derive a robust feature importance ranking, we integrate the results from the five complementary methods: SHAP, Permutation Importance, Mutual Information, MDI, and MI. Each offering distinct perspectives on feature relevance. SHAP and Permutation Importance provide model-specific insights: SHAP quantifies per-feature contributions to individual predictions, while Permutation Importance assesses performance degradation upon feature randomization. In contrast, MI and MIC are statistical measures, with MI estimating general dependence between features and the target, and MIC emphasizing non-linear associations. MDI, intrinsic to tree-based models, reflects impurity reduction during splits. The challenge here is that each method has different scales and rankings. Directly averaging them isn't feasible because of the scale differences. So, normalization is reasonable to bring all values to a comparable scale. The importance scores within each method should be scaled to a [0,1] range relative to the method-specific maximum. Features are then ranked within each method (1 = most important). Finally, we compute aggregated scores and ranks by averaging normalized values and ranks across all methods. This approach synthesizes model-specific, statistical, and tree-based perspectives into a unified ranking, mitigating biases inherent to individual techniques while preserving their complementary strengths.

G. Used Model to Predict the Fatigue State

For fatigue state classification, we implemented the tabular transformer. The architecture is designed to leverage transformer-based attention mechanisms for capturing complex interactions in tabular physiological and behavioral data. Figure 2 shows the architecture of the used model. The model

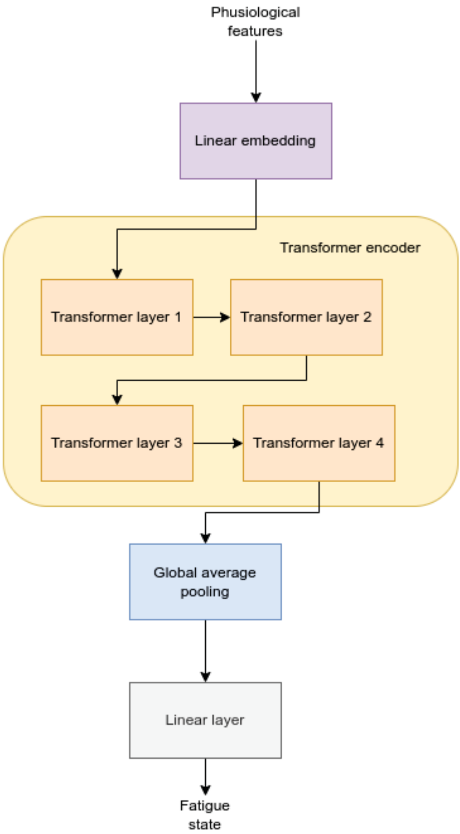

Fig. 2. The model used to predict the fatigue state

processes standardized input features through an embedding layer that projects the features into a 64 dimensional latent space. This embedding step transforms the physiological features into dense representations suitable for capturing non-linear relationships. The core of the architecture consists of a transformer encoder with four stacked layers. Each one employs a multi-head self-attention mechanisms with four heads per layer. The transformer layers process the embedded features as a sequence, which allows the model to relate features through attention scores. The output of the transformer blocks pass to a global average pooling layer. This layer reduces the dimension by averaging the input across the sequence length to produce a single 64-dimensional vector that encapsulates the aggregated feature interactions. The layer output is then passed to a linear layer that maps the latent vector to logits for the two output classes. The input physiological features are standardized using StandardScaler to ensure zero mean and unit variance.

IV. Experiments

A. Performance of the Original Model

As we mentioned earlier, our original random forest model performed quite well, achieving an accuracy of 98% and an F1 score of 94% which indicates that the issue of the imbalanced data that we suffered from did not seem to hinder our model's ability to maintain consistent performance across different classes. It is worth noting that we achieved the best results when we included the participant ID as one of the features. This addition positively affected the model's performance, suggesting that individuals experience fatigue can vary significantly. The positive effect of participant identification was observed in the F1 Score, which increased from 81% when only physiological indicators were utilized. This F1 Score of 81% still demonstrates the good performance achieved through the application of these indicators, as estimated using deep learning models. However, for our feature importance analysis, we will leave out the participant ID so we can focus on the other continuous numerical features (the physiological indicators) used.

B. Feature Importance Analysis

As explained before we used the features importance analyzing techniques listed in Section III-D . Using multiple feature importance techniques enhances the reliability and robustness of the analysis by providing comprehensive insights. This approach helps mitigate model-specific biases and improves explainability. Additionally, it aids in better decision-making for feature selection by identifying consistently important features across methods.

Starting with the first technique, SHAP values indicate the impact of individual features on model predictions, with positive values increasing fatigue predictions and negative values decreasing them. The summary plot in Figure 3 presents feature importance ranked by their impact, with colors representing feature values (high in red, low in blue).

As illustrated in Figure 3, heart rate is the strongest predictor, with higher values increasing fatigue predictions. Respiratory Rate (Average RR), however, shows an inverse relationship — higher RR is associated with lower fatigue, suggesting a more alert physiological state. Head movements (average Pitch, average Roll, average Yaw) contribute notably,

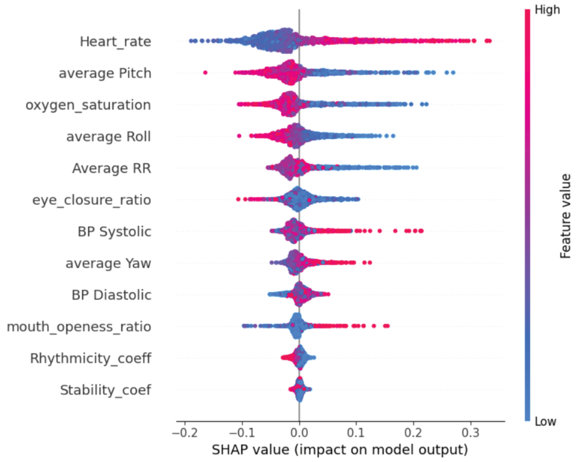

Fig. 3. Results of SHAP

with increased movements linked to fatigue, possibly due to postural instability. Oxygen saturation follows a similar pattern, where lower levels increase fatigue predictions. Eye Closure Ratio and Mouth Openness Ratio are also impactful. Increased eye closure correlates with fatigue, while greater mouth openness (yawning) is a known fatigue indicator. Blood Pressure (BP Systolic, BP Diastolic) plays a secondary role, and Rhythmicity/Stability Coefficients (Rhythmicity Coeff, Stability Coeff) have minimal impact.

Table I presents the average absolute SHAP values for each feature. We believe that this representation of SHAP importance is more interpretable for our objectives, as it provides numerical values that enable comparisons of feature importance across the five techniques within the same framework, thereby enhancing the reliability of the comparisons.

TABLE I. FEATURE IMPORTANCE SCORES FROM SHAP ANALYSIS

Feature	SHAP Importance
Heart Rate	0.057
Average Pitch	0.033
Oxygen Saturation	0.031
Average Roll	0.024
Average RR	0.024
Eye Closure Ratio	0.015
BP Systolic	0.015
Average Yaw	0.012
BP Diastolic	0.010
Mouth Openess Ratio	0.009
Rhythmicity Coefficient	0.005
Stability Coefficient	0.003

Subsequently, the results obtained from the other techniques are presented in the following tables: Permutation Importance (Table II), Mutual Information (Table III), Mean Decrease in Impurity (Table IV), and Maximal Information Coefficient (Table V). The features are organized in descending order of impact, from the most significant to the least, based on the coefficients and outcomes of these techniques.

TABLE II. PERMUTATION IMPORTANCE RESULTS

Features	Permutation Importance Score
Heart Rate	0.034
Average Pitch	0.032
Oxygen Saturation	0.024
Average Roll	0.017
Average RR	0.012
Eye Closure Ratio	0.010
BP Systolic	0.010
Mouth Openess Ratio	0.008
BP Diastolic	0.006
Average Yaw	0.004
Stability Coefficient	0.001
Rhythmicity Coefficient	0.001

TABLE III. MUTUAL INFORMATION RESULTS

Features	MI Score
BP Diastolic	0.142
BP Systolic	0.136
Oxygen Saturation	0.104
Average Pitch	0.039
Average RR	0.035
Average Yaw	0.021
Heart Rate	0.020
Eye Closure Ratio	0.017
Mouth Openess Ratio	0.017
Average Roll	0.012
Stability Coefficient	0.005
Rhythmicity Coefficient	0.000

The results of our analysis reveal notable variations across different techniques. While SHAP analysis, Permutation importance scores and MDI agree on the importance ranking of many features, we observe that Mutual Information scores and Maxiaml Information Coefficient provide different results. This emphasizes the importance of using multiple test and analysis techniques to determine the most critical features. By finding a common set of important features that have significant ranks across the different analytical approaches, we can establish a more robust and reliable foundation for our fatigue detection model.

Next we aggregated the results as explained in section III-F from the five different methods to determine the most important features. Table VI shows the average normalized feature importance scores for the physiological features, which were determined by aggregating the findings from five different feature importance methods (SHAP, Permutation Importance, Mutual Information, MDI, and MIC). By combining the insights from these diverse methods, we aim to capture a comprehensive and balanced view of which features most significantly contribute to predicting mental fatigue.

These results indicate that heart rate, oxygen saturation, average pitch, systolic and diastolic blood pressure are the most influential features for fatigue detection, as they consistently appeared as the most important across the different feature importance techniques. This suggests that these physiological

TABLE IV. MEAN DECREASE IN IMPURITY RESULTS

Features	MDI Score
Heart Rate	0.156
Average Pitch	0.136
Oxygen Saturation	0.126
BP Systolic	0.091
Average Roll	0.090
Average RR	0.082
Eye Closure Ratio	0.080
BP Diastolic	0.076
Average Yaw	0.073
Mouth Openess Ratio	0.043
Rhythmicity Coefficient	0.022
Stability Coefficient	0.019

TABLE V. MAXIMAL INFORMATION COEFFICIENT RESULTS

Features	MIC Score
BP Systolic	0.281
BP Diastolic	0.260
Oxygen Saturation	0.215
Average Pitch	0.158
Average Yaw	0.132
Average RR	0.131
Heart Rate	0.131
Eye Closure Ratio	0.126
Average Roll	0.123
Mouth Openess Ratio	0.093
Stability Coefficient	0.029
Rhythmicity Coefficient	0.026

TABLE VI. AVERAGE NORMALIZED FEATURE IMPORTANCE SCORES

Feature	Average Normalized Score	Final Rank
Heart Rate	0.722	1
Oxygen Saturation	0.719	2
Average Pitch	0.652	3
BP Systolic	0.621	4
BP Diastolic	0.558	5
Average RR	0.409	6
average Roll	0.405	7
Eye Closure Ratio	0.335	8
Average Yaw	0.285	9
Mouth Openess Ratio	0.226	10
Stability Coefficient	0.074	11
Rhythmicity Coefficient	0.072	12

TABLE VII. CROSS-VALIDATION PERFORMANCE OF RANDOM FOREST AND TABULAR TRANSFORMER MODELS

	Fold					Mean
	1	2	3	4	5	
F1 score (RF)	0.69	0.75	0.70	0.70	0.71	0.71
Accuracy (RF)	86.8	89.9	89.8	88.6	89.5	88.9
Precision (RF)	0.64	0.81	0.696	0.77	0.77	0.74
Recall (RF)	0.76	0.70	0.70	0.64	0.66	0.69
F1 score (TT)	0.70	0.75	0.76	0.73	0.76	0.74
Accuracy (TT)	86.6	90.2	91.9	89.2	91.48	89.9
Precision (TT)	0.72	0.86	0.702	0.78	0.83	0.78
Recall (TT)	0.69	0.65	0.83	0.69	0.70	0.71

indicators are crucial for accurate mental fatigue prediction.

The used model is illustrated in Section III-G. It was optimized using the Adam optimizer with a learning rate of 0.001 and trained for 150 epochs, the cross-entropy loss was used. The architecture operates on CPU-based computation and leverages PyTorch's automatic differentiation for efficient gradient calculation. The used hyperparameters include the embedding dimension which was experimentally chosen as 64, the number of transformer layers and attention heads were set to 4. The final classification probabilities are derived by applying a SoftMax function to the logits, though this is implicitly handled by the Cross-Entropy loss during training.

The data was split into 5 folds. The result of the model using the top five features (BP Systolic and Diastolic, Heart Rate, Oxygen Saturation and Average Pitch Angle) is introduced in Table VII in addition to the result of the Random Forest model used previously for comparison.

Both the Random Forest (RF) and Tabular Transformer models delivered consistent performance across cross-validation folds with the Tabular Transformer achieving slightly stronger results overall. The Tabular Transformer achieves a mean F1-Score of 0.741 and Accuracy of 89.90% compared to RF's F1 score of 0.714 and accuracy of 88.97%. Notably, the Tabular Transformer exhibits greater stability in F1 scores across folds what indicates a better robustness to dataset variations. The RF's lower variance in accuracy suggests simpler interpretability at the cost of slightly reduced performance. These results align with the hypothesis that transformer-based architectures with their self-attention mechanisms better model intricate dependencies in tabular data compared to traditional tree-based methods. However, the modest performance gap implies that dataset characteristics, such as feature sparsity or non-linearity, may influence the relative advantage of transformers over ensemble methods.

C. Enhancement of the Computational Cost

The computational cost required with our approach when using all twelve features, is approximately 32 minutes and 28 seconds for processing a one-minute video. This duration was determined while executing all models on a Central Processing Unit (CPU), which explains the high computational cost. In contrast, when employing only the five most significant features on the same CPU, the processing time was reduced to 26 minutes and 12 seconds, indicating a cost reduction of approximately 20%.

Notably, the majority of the processing time is attributed to the calculation of heart rate, which requires around 22 minutes to analyze one-minute video on CPU. This long duration is primarily due to the use of a vision transformer model, which necessitates extensive processing time on the CPU. However, we anticipate that utilizing a Graphics Processing Units (GPU) will significantly decrease this computational cost, since they are specifically designed to handle parallel processing tasks, making them highly efficient for the types of computations required by Vision Transformers (ViTs) and Convolutional Neural Networks (CNNs), thereby facilitating the application of our approach in real-time scenarios.

V. Conclusion

This study builds on our earlier work in contactless mental fatigue detection, with a focus on enhancing computational efficiency while maintaining high performance. A central contribution of this research is the identification of the most informative physiological features through a comparative analysis of five feature importance techniques. Despite some discrepancies among methods, Heart Rate, Oxygen Saturation, Blood Pressure (Systolic and Diastolic), and Average Pitch consistently emerged as the most critical indicators.

Using only these five features, we achieved 88% accuracy and a 71% F1 score with a Random Forest model—demonstrating that significant dimensionality reduction is possible with minimal performance loss. Notably, this refined feature set maintained strong results, despite excluding seven other features used in our earlier model.

We further validated these findings with a Tabular Transformer model, which achieved 89% accuracy and a 74% F1 score, indicating enhanced robustness and stability across varied data. These results highlight both the effectiveness of the selected physiological indicators and the potential of transformer-based architectures for fatigue detection in tabular data.

This research faced challenges that hindered performance in assessing mental fatigue. Individual variability complicates model generalization, while feature selection requires careful analysis to identify relevant variables. Additionally, dataset constraints, such as small sizes and unbalanced targets, negatively impacted model stability.

This work has broad implications to develop effective and non-invasive monitoring systems of mental fatigue across different real-life applications. By identifying key physiological indicators like heart rate and oxygen saturation, the study encourages the utilization of lightweight models for transportation, healthcare, and workplace safety technologies. It also opens opportunities for personalized health tools and devices that adapt to users' cognitive states. Additionally, the findings provide a foundation for future research on scalable and generalizable fatigue detection systems that can enhance human performance and health.

Our future research in mental fatigue detection will focus on exploring advanced transformer architectures customized for tabular data may enhance model performance and robustness. Addressing dataset limitations by collecting larger and more diverse datasets will also be essential for generalizability.

Acknowledgment

The research is funded by the Russian Science Foundation (project 24-21-00300).

References

[1] G. Zhang, K. K. Yau, X. Zhang, and Y. Li, "Traffic accidents involving fatigue driving and their extent of casualties," *Accident Analysis Prevention*, vol. 87, pp. 34–42, 2016. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0001457515301159

[2] R. Alharbey, M. M. Dessouky, A. Sedik, A. I. Siam, and M. A. Elaskily, "Fatigue state detection for tired persons in presence of driving periods," *IEEE Access*, vol. 10, pp. 79 403–79 418, 2022.

[3] L. L. Di Stasi, R. Renner, A. Catena, J. J. Cañas, B. M. Velichkovsky, and S. Pannasch, "Towards a driver fatigue test based on the saccadic main sequence: A partial validation by subjective report data," *Transportation Research Part C: Emerging Technologies*, vol. 21, no. 1, pp. 122–133, 2012. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0968090X1100101X

[4] Y. Yamada and M. Kobayashi, "Fatigue detection model for older adults using eye-tracking data gathered while watching video: Evaluation against diverse fatiguing tasks," in *2017 IEEE International Conference on Healthcare Informatics (ICHI)*, 2017, pp. 275–284.

[5] W. Othman, B. Hamoud, N. Shilov, and A. Kashevnik, "Human operator mental fatigue assessment based on video: Ml-driven approach and its application to hfavd dataset," *Applied Sciences*, vol. 14, no. 22, 2024. [Online]. Available: https://www.mdpi.com/2076-3417/14/22/10510

[6] J. Meng, B. Zhao, Y. Ma, J. Yiyu, and B. Nie, "Effects of fatigue on the physiological parameters of labor employees," *Natural Hazards*, vol. 74, 11 2014.

[7] H. Luo, P.-A. Lee, I. Clay, M. Jaggi, and V. De Luca, "Assessment of Fatigue Using Wearable Sensors: A Pilot Study," *Digital Biomarkers*, vol. 4, no. Suppl. 1, pp. 59–72, 11 2020. [Online]. Available: https://doi.org/10.1159/000512166

[8] G. Louppe, "Understanding random forests: From theory to practice," 2015.

[9] J.-Y. Franceschi, A. Dieuleveut, and M. Jaggi, "Unsupervised scalable representation learning for multivariate time series," 2020.

[10] J. Veltman and A. Gaillard, "Physiological indices of workload in a simulated flight task," *Biological Psychology*, vol. 42, no. 3, pp. 323–342, 1996, psychophysiology of Workload. [Online]. Available: https://www.sciencedirect.com/science/article/pii/0301051195051651

[11] N. Hjortskov, D. Rissén, A. Blangsted, N. Fallentin, U. Lundberg, and K. Søgaard, "The effect of mental stress on heart rate variability and blood pressure during computer work," *European journal of applied physiology*, vol. 92, pp. 84–9, 07 2004.

[12] M. Ye, W. Zhang, P. Cao, and K. Liu, "Driver fatigue detection based on residual channel attention network and head pose estimation," *Applied Sciences*, vol. 11, no. 19, 2021. [Online]. Available: https://www.mdpi.com/2076-3417/11/19/9195

[13] A. Hussain, F. Saharil, R. Mokri, and B. Y. Majlis, "On the use of mems accelerometer to detect fatigue department," *2004 IEEE International Conference on Semiconductor Electronics*, pp. 5 pp.–, 2004. [Online]. Available: https://api.semanticscholar.org/CorpusID:34128395

[14] S. Ansari, F. Naghdy, H. Du, and Y. N. Pahnwar, "Driver mental fatigue detection based on head posture using new modified relu-bilstm deep neural network," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 8, pp. 10 957–10 969, 2022.

[15] T. Mori, K. Jokinen, and Y. Den, "Cognitive states and types of nods," in *PVLAM*, 2022. [Online]. Available: https://api.semanticscholar.org/CorpusID:252624596

[16] Q. Zhao, B. Nie, T. Bian, X. Ma, L. Sha, K. Wang, and J. Meng, "Experimental study on eye movement characteristics of fatigue of selected college students," 2023.

[17] S. Dey, S. A. Chowdhury, S. Sultana, M. A. Hossain, M. Dey, and S. K. Das, "Real time driver fatigue detection based on facial behaviour along with machine learning approaches," in *2019 IEEE International Conference on Signal Processing, Information, Communication Systems (SPICSCON)*, 2019, pp. 135–140.

[18] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, vol. 1, 2005, pp. 886–893 vol. 1.

[19] A. Kashevnik, R. Shchedrin, C. Kaiser, and A. Stocker, "Driver distraction detection methods: A literature review and framework," *IEEE Access*, vol. PP, pp. 1–1, 04 2021.

[20] M. A. Kamran, M. M. N. Mannan, and M. Y. Jeong, “Drowsiness, fatigue and poor sleep’s causes and detection: A comprehensive study,” *IEEE Access*, vol. 7, pp. 167 172–167 186, 2019.

[21] J. Li, “Heart rate estimation using deep learning and hybrid artifact removal for optical volume tracing signals in motion states,” *International Journal of High Speed Electronics and Systems*, vol. 0, no. 0, p. 2540321, 0. [Online]. Available: https://doi.org/10.1142/S0129156425403213

[22] R. Song, Z. Du, J. Cheng, C. Li, and X. Yang, “Video-based heart rate estimation with spectrogram signal quality ranking and fusion,” *Biomedical Signal Processing and Control*, vol. 100, p. 107094, 2025. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1746809424011522

[23] A. Dasari, L. A. Jeni, and C. S. Tucker, “Video-based estimation of blood pressure,” *PLOS ONE*, vol. 20, no. 1, pp. 1–31, 01 2025. [Online]. Available: https://doi.org/10.1371/journal.pone.0311654

[24] W. Zhuo, J. Qian, H. Shao, L. Luo, and J. Yang, “Video-based remote blood pressure measurement using convolutional networks and random forest,” in *IJCAI 2024: International Joint Conference on Artificial Intelligence*, 2024.

[25] H. Chen, X. Zhang, Z. Guo, N. Ying, M. Yang, and C. Guo, “Actnet: Attention based cnn and transformer network for respiratory rate estimation,” *Biomedical Signal Processing and Control*, vol. 96, p. 106497, 2024. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S174680942400555X

[26] M. Mozafari, A. J. Law, R. A. Goubran, and J. R. Green, “Respiratory rate estimation from thermal video data using spatio-temporal deep learning,” *Sensors*, vol. 24, no. 19, 2024. [Online]. Available: https://www.mdpi.com/1424-8220/24/19/6386

[27] X. Sun, T. Wen, W. Chen, and B. Huang, “Ccspo2net: Camera-based contactless oxygen saturation measurement foundation model in clinical settings,” *IEEE Transactions on Instrumentation and Measurement*, vol. 73, pp. 1–11, 2024.

[28] K. Z. Arefin, K. S. Alam, S. M. Mamun, N. U. S. Sabith, M. Rabbani, P. Sridevi, and S. I. Ahamed, “Pulsesight : A novel method for contactless oxygen saturation (spo2) monitoring using smartphone cameras, remote photoplethysmography and machine learning,” *Smart Health*, vol. 36, p. 100542, 2025. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2352648325000030

[29] N. Ruiz, E. Chong, and J. M. Rehg, “Fine-grained head pose estimation without keypoints,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, June 2018.

[30] R. Ranjan, V. M. Patel, and R. Chellappa, “Hyperface: A deep multi-task learning framework for face detection, landmark localization, pose estimation, and gender recognition,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 41, no. 1, pp. 121–135, 2019.

[31] G. P. Meyer, S. Gupta, I. Frosio, D. Reddy, and J. Kautz, “Robust model-based 3d head pose estimation,” in *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, December 2015.

[32] H. Yang, L. Liu, W. Min, X. Yang, and X. Xiong, “Driver yawning detection based on subtle facial action recognition,” *IEEE Transactions on Multimedia*, vol. 23, pp. 572–583, 2021. [Online]. Available: https://api.semanticscholar.org/CorpusID:216334697

[33] S. Saurav, S. Mathur, I. Sang, S. S. Prasad, and S. Singh, “Yawn detection for driver’s drowsiness prediction using bi-directional lstm with cnn features,” in *Intelligent Human Computer Interaction*, U. S. Tiwary and S. Chaudhury, Eds. Cham: Springer International Publishing, 2020, pp. 189–200.

[34] T. Soukupová and J. Cech, “Eye-blink detection using facial landmarks,” 2016. [Online]. Available: https://api.semanticscholar.org/CorpusID:35923299

[35] M. A. Aqif, N. Rahaman, and M. A. R. Ahad, “A study on tiredness assessment by using eye blink detection,” *Jurnal Kejuruteraan*, vol. 31, pp. 209–214, 10 2019.

[36] S. Abtahi, M. Omidyeganeh, S. Shirmohammadi, and B. Hariri, “Yawdd: A yawning detection dataset,” 03 2014, pp. 24–28.

[37] Q. Massoz, T. Langohr, C. François, and J. G. Verly, “The ulg multimodality drowsiness database (called drozy) and examples of use,” in *2016 IEEE Winter Conference on Applications of Computer Vision (WACV)*, 2016, pp. 1–7.

[38] M. Basner and D. F. Dinges, “Maximizing Sensitivity of the Psychomotor Vigilance Test (PVT) to Sleep Loss,” *Sleep*, vol. 34, no. 5, pp. 581–591, 05 2011. [Online]. Available: https://doi.org/10.1093/sleep/34.5.581

[39] A. Shahid, K. Wilkinson, S. Marcu, and C. M. Shapiro, *Karolinska Sleepiness Scale (KSS)*. New York, NY: Springer New York, 2012, pp. 209–210. [Online]. Available: https://doi.org/10.1007/978-1-4419-9893-4_47

[40] E. Kavalci Yilmaz and M. A. Akcayol, “SUST-DDD: A Real-Drive Dataset for Driver Drowsiness Detection.” The 31st IEEE Conference of Open Innovations Association FRUCT, 2022. [Online]. Available: https://doi.org/10.5281/zenodo.6519933

[41] P. Liu, H.-L. Chi, X. Li, and J. Guo, “Effects of dataset characteristics on the performance of fatigue detection for crane operators using hybrid deep neural networks,” *Automation in Construction*, vol. 132, p. 103901, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0926580521003526

[42] W. Othman, A. Kashevnik, I. Ryabchikov, and N. Shilov, “Contactless camera-based approach for driver respiratory rate estimation in vehicle cabin,” *Lecture Notes in Networks and Systems*, vol. 5431, p. 429–442, 2022.

[43] W. Othman, A. Kashevnik, A. Ali, N. Shilov, and D. Ryumin, “Remote heart rate estimation based on transformer with multi-skip connection decoder: Method and evaluation in the wild,” *Sensors*, vol. 24, no. 3, 2024. [Online]. Available: https://www.mdpi.com/1424-8220/24/3/775

[44] B. Hamoud, A. Kashevnik, W. Othman, and N. Shilov, “Neural network model combination for video-based blood pressure estimation: New approach and evaluation,” *Sensors*, vol. 23, no. 4, 2023. [Online]. Available: https://www.mdpi.com/1424-8220/23/4/1753

[45] B. Hamoud, W. Othman, N. Shilov, and A. Kashevnik, “Contactless oxygen saturation detection based on face analysis: An approach and case study,” in *2023 33rd Conference of Open Innovations Association (FRUCT)*, 2023, pp. 54–62.

[46] A. Kashevnik, A. Ali, I. Lashkov, and D. Zubok, “Human head angle detection based on image analysis,” in *Proceedings of the Future Technologies Conference (FTC) 2020, Volume 1*, K. Arai, S. Kapoor, and R. Bhatia, Eds. Cham: Springer International Publishing, 2021, pp. 233–242.

[47] F. Hasan and A. Kashevnik, “State-of-the-art analysis of modern drowsiness detection algorithms based on computer vision,” in *2021 29th Conference of Open Innovations Association (FRUCT)*, 2021, pp. 141–149.

[48] S. Kovalenko, A. Mamonov, V. Kuznetsov, A. Bulygin, I. Shoshina, I. Brak, and A. Kashevnik, “Operatoreyevp: Operator dataset for fatigue detection based on eye movements, heart rate data, and video information,” *Sensors*, vol. 23, no. 13, 2023. [Online]. Available: https://www.mdpi.com/1424-8220/23/13/6197

[49] S. M. Lundberg and S.-I. Lee, “A unified approach to interpreting model predictions,” in *Advances in Neural Information Processing Systems*, I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, Eds., vol. 30. Curran Associates, Inc., 2017. [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2017/file/8a20a8621978632d76c43dfd28b67767-Paper.pdf

[50] A. Altmann, L. Tolosi, O. Sander, and T. Lengauer, “Permutation importance: A corrected feature importance measure,” *Bioinformatics (Oxford, England)*, vol. 26, pp. 1340–7, 04 2010.

[51] *Entropy, Relative Entropy, and Mutual Information*. John Wiley Sons, Ltd, 2005, ch. 2, pp. 13–55. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/047174882X.ch2

[52] G. Louppe, L. Wehenkel, A. Sutera, and P. Geurts, “Understanding variable importances in forests of randomized trees,” in *Advances in Neural Information Processing Systems*, C. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K. Weinberger, Eds., vol. 26. Curran Associates, Inc., 2013. [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2013/file/e3796ae838835da0b6f6ea37bcf8bcb7-Paper.pdf

[53] D. N. Reshef, Y. A. Reshef, H. K. Finucane, S. R. Grossman, G. McVean, P. J. Turnbaugh, E. S. Lander, M. Mitzenmacher, and P. C. Sabeti, “Detecting novel associations in large data sets,” *Science*, vol. 334, no. 6062, pp. 1518–1524, 2011. [Online]. Available: https://www.science.org/doi/abs/10.1126/science.1205438