

MDFG: Multi-Dimensional Fine-Grained Modeling for Fatigue Detection

Mei Wang^{1,2*}, Xiaojie Zhu^{4*}, Ruimin Hu^{1,2,3†}, Dongliang Zhu^{1,2}, Liang Liao⁵, Mang Ye^{1,2}

¹National Engineering Research Center for Multimedia Software, School of Computer Science, Wuhan University

²Hubei Key Laboratory of Multimedia and Network Communication Engineering, Wuhan University

³School of Cyber Science and Engineering, Wuhan University, Wuhan, China

⁴Cyberspace Security Laboratory, School of Network and Information Security, Xidian University, Xi'an, China

⁵The School of Computer Science and Engineering, Nanyang Technological University, Singapore

{dr.mei.wang, hrm, zhudongliang, yemang}@whu.edu.cn, zhuxiaojie@stu.xidian.edu.cn, liang.liao@ntu.edu.sg

Abstract

Fatigue is a critical factor contributing to accidents in industries such as safety monitoring and engineering construction. Fatigue exhibits dynamic complexity and non-stationary characteristics, so there are many intermediate states of short-term variation between alert and fatigue. Capturing and learning the signs of these intermediate states is essential for accurate fatigue assessment. However, current fatigue detection methods primarily rely on coarse-grained labels, typically spanning minutes to hours, and commonly treat alert and fatigue as two distinctly separate distributions, overlooking the expression of intermediate states and oversimplifying the rich distribution information of fatigue types and levels, thereby limiting detection effectiveness. To address these, this paper explores a refined representation of fatigue in terms of three dimensions: time, type, and level, and proposes a **Multi-Dimensional Fine-Grained Modeling for Fatigue Detection (MDFG)**. This introduces the SmallLoss to extract trustworthy samples, utilizes clustering to identify diverse subtypes under alert and fatigued states, and establishes base class sets in each state. Subsequently, a complete base class set containing intermediate state bases is constructed using the base class synthesis method, which achieves the expression of intermediate fatigue states from absence to presence. Finally, fatigue levels are quantified based on the matching between samples and the complete base class set. Moreover, to cope with the complex variability of fatigue states, MDFG employs meta-learning for training. MDFG achieves an Average accuracy improvement of 10.0% and 12.1% on two real datasets compared to methods that do not consider fine-grained information. Extensive experiments demonstrate that the MDFG exhibits superior robustness and stability among current fatigue detection methods.

Introduction

Fatigue significantly reduces cognitive task performance, particularly in high-risk fields such as aviation, automotive, and construction machinery operations, which has become a critical factor in causing safety accidents (Tu et al. 2022; Li et al. 2022; Zhang et al. 2020; Dai, Li, and Zhang 2023). Fatigue detection has received a lot of attention in

academia, and researchers have developed various fatigue detection methods, including those based on physiological signals (Zhao et al. 2024; Gong et al. 2024), psychological measurements (Wiehler et al. 2022), behavioral patterns (Dai, Li, and Zhang 2023; Sukumar et al. 2024), and facial features (Ghoddosian, Galib, and Athitsos 2019; Liao et al. 2024). It has become a new research hotspot.

Fatigue does not develop steadily. Transitional states of fatigue often occur when individuals are disturbed by the environment (Yin et al. 2016), i.e., fatigue includes many intermediate states with short-term variations between alert and fatigue, exhibiting varying fatigue levels. Capturing these states is crucial for accurate assessment and management of fatigue. However, fatigue labels typically rely on self-assessment questionnaires (e.g., KSS and SSS (Perkins et al. 2022)), reflecting overall fatigue over several minutes to hours, known as Coarse-Grained Label (CGL), which fail to capture short-term fatigue changes within seconds or milliseconds. Thus, a fine-grained labeling system is urgently needed for more accurate and reliable fatigue detection.

For Fine-Grained Label (FGL) in fatigue detection, cFN-SCL (Wang et al. 2024a) introduces a method that automatically generates FGL from CGL by self-supervised contrastive learning. However, cFN-SCL only categorizes short-term clips as alert or fatigued and fails to capture intermediate states. FatigueView (Yang et al. 2022) proposes a hierarchical fine-grained annotation method, FatigueTree, which refines labels from coarse to fine (elements, sets, and events) by considering the “eyes,” “head,” “mouth,” and their 14 combinations. Take the element “eye” as an example; its set contains various types like eye squinting, closure, blinking, etc., with events categorized into distinct levels: alert, signs, sleepy, and sleep. Similarly, Liu et al. (2021) manually re-labeled single-minute videos. However, these methods primarily rely on manual, it challenging to assess short-term fatigue variations effectively. Nonetheless, these studies underscore the complexity and diversity of fatigue types and levels. Fig.1(a) and Fig.1(b) demonstrate the errors from neglecting these aspects, while Fig.1(c) shows the impact on performance in reRLDD and reDROZY. In conclusion, effectively measuring intermediate states requires more than just a temporal dimension; it must account for the intricate distribution of fatigue types and levels. Therefore, we propose a Multi-Dimensional Fine-Grained Label (MDFGL),

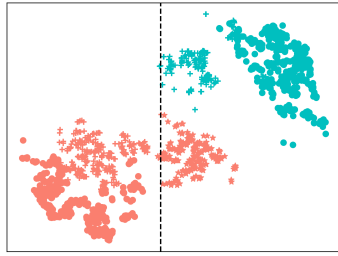
*These authors contributed equally.

†Corresponding author.

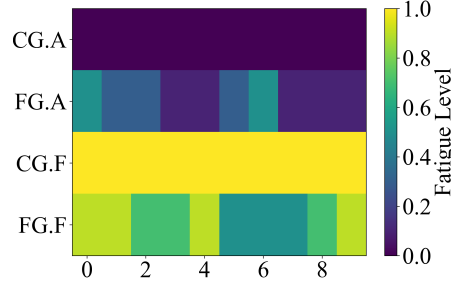
Copyright © 2025, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

Object	Criterion	Temporal Resolution	Complexity	Clarity	Example
CGL	subjective	low, minutes or above	single type	vague, inaccurate	“alert” or “fatigue”
MDFGL	objective	high, up to milliseconds	rich types	accurate, specific	“fatigue score: 0.7”

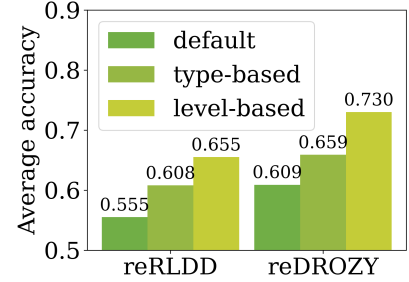
Table 1: Difference between Coarse-Grained Label and Multi-Dimensional Fine-Grained Label



(a) Misclassified sub-types if we treat it as one distribution



(b) Manual fails to capture short-term fatigue changes and quantify fatigue levels



(c) Overlooking fine-grained information substantially reduces performance

Figure 1: (a) If we treat the data as a single type, the sub-types will be misclassified. Different colors and shapes correspond to different classes and types. (b) Demonstrates that the intermediate states have different fatigue levels, which is challenging to annotate manually. Deep purple indicates alert, while yellow indicates fatigue. CG and FG denote coarse-grained and fine-grained perspectives, respectively. (c) The “default” indicates no fine-grained, while “type-based” and “level-based” involve type and level fine-grained, respectively. The latter methods significantly outperform “default,” all using the same network ϕ_F .

defined across time, type, and level three dimensions to advance more refined and accurate fatigue assessments.

MDFGL encompasses Fine-Grained Temporality (FGT), Fine-Grained Categorization (FGC), and Fine-Grained Quantification (FGQ). FGT is to refine the label from coarse, human-perceptible scales like minutes to fine-grained, imperceptible scales like milliseconds, allowing for the precise portrayal of the temporal dimension. FGC is to further subdivide a fatigued state into multiple specific subtypes, each of which exhibits different behavior patterns, demonstrating the type diversity within the fatigue state. FGQ is to partition the range between alert (0) and fatigued (1) into multiple intermediate levels (within fatigue scores 0 to 1), accurately reflecting fatigue level. It imposes higher requirements in temporal resolution, type complexity, and detailed fatigue level. Tab. 1 shows its difference from CGL.

The central challenge of this paper is the automatic generation of MDFGL. Achieving this requires effectively representing the intermediate states. While directly obtaining these states is difficult, we can obtain the distributions for alert and fatigued states, which prompts us to consider constructing intermediate states through synthesis. First, we use clustering analysis to identify the base classes representing the distribution characteristics of alert and fatigue states, which reveal multiple subtypes of the states. Subsequently, employing the Gaussian distribution linear combination synthesizes intermediate state distributions, forming a base class set for these states. With the established intermediate base class set, FGT samples can be automatically assessed by matching them to the nearest synthetic base class,

enabling precise fatigue detection.

Based on the above research, we propose a novel method MDFGL for fatigue detection, which comprises a trustworthy sample and base class extraction module, a complete base class construction module, and a fine-grained label generation module. MDFGL captures the diversity and nuances of fatigue by synthesizing intermediate states, significantly expanding the alert-intermediate-fatigue scale and enhancing fatigue discrimination. Given the variability and complexity of fatigue states, both inter- and intra-individual, we also introduce a meta-learning module into MDFGL, using a multi-task learning mechanism to further improve fatigue discrimination by deeply understanding differences in data distribution. Extensive experiments demonstrate that MDFGL exhibits superior robustness and stability among current methods. In summary, the main contributions of this paper are:

- **Novel perspective:** We propose a new perspective on fatigue detection from MDFGL and provide its complete definition for the first time, including FGT, FGC, and FGQ. This approach is more challenging than traditional fatigue assessment methods, as it involves capturing numerous short-term varying intermediate states.
- **Novel methodology:** MDFGL is a novel method for fatigue detection based on MDFGL learning, which can identify complex fatigue states. Technically, we propose a methodology for synthesizing fatigue intermediate states based on base classes, realizing the expression of these intermediate states from scratch, and a method for quantifying fatigue levels based on these base classes.
- **Superior performance and insightful results:** MDFGL

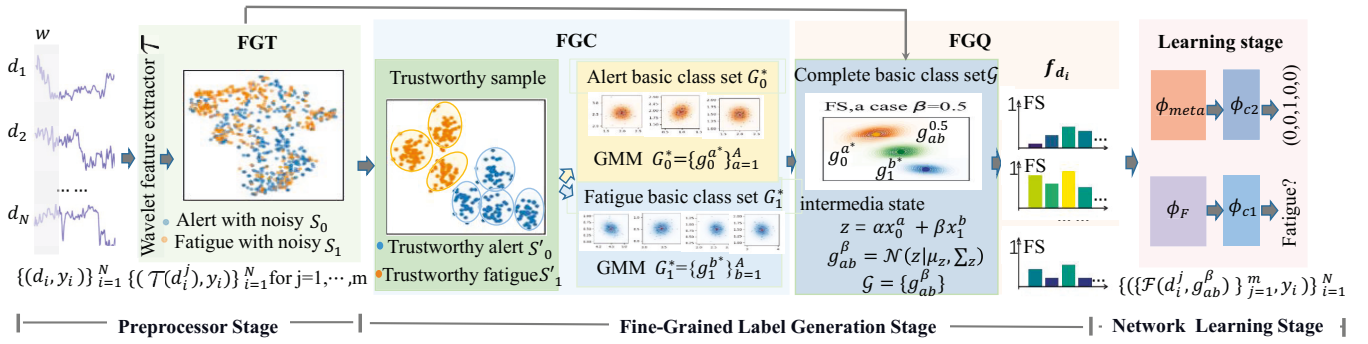


Figure 2: Overview of MDFG. In the preprocessing stage, the wavelet feature extractor \mathcal{T} generates noise-containing FGT samples. In the fine-grained generation stage, we extract base classes from trustworthy samples and then combine the alert base class set G_0^* and the fatigue base class set G_1^* to generate intermediate state base classes. The fatigue score β for $\mathcal{T}(d_i^j)$ is assigned based on the closest base class to that in \mathcal{G} . Finally, in the network learning stage, a fatigue detection model is constructed based on the Fatigue Scores (FS) sequence.

offers a refined expression and analysis of fatigue phenomena across temporal, type, and level dimensions. Compared to the optimal method, it improves the Av acc on reRLDD and reDROZY by 9.1% and 4.9%, respectively, while exhibiting excellent robustness and stability. The code and checkpoints are released at <https://github.com/MeiWang003/MDFG>.

Related Works

Eye movement features are crucial in fatigue detection. Since PERCLOS (Percentage of Eye Closure per Unit Time) (Trutschel et al. 2011) was officially recognized by the U.S. Federal Highway Administration as a viable fatigue detection method in 1999, research has continually advanced. Facial expression analysis methods, such as S-ALE (Liao et al. 2024) and RVT (Wang et al. 2024b), have also validated the importance of the eye region in fatigue detection. In recent years, numerous fatigue detection studies based on eye features have emerged, including Eye16features (Masoz et al. 2016), Eye4features (Byrnes and Sturton 2018), HMLSTM (Ghoddosian, Galib, and Athitsos 2019), FacialUnits (Cheng et al. 2019), Cheng et al. (2019), MPE-blink (Zeng et al. 2023), Lee et al. (2022) and Mehmood et al. (2024). However, these methods typically assume the alert and fatigued states as two distinctly different distributions for learning, overlooking the complex variability of states and the expression of intermediate states. This paper uses eye features and proposes MDFG to improve fatigue recognition by enhancing the fine-grained representation of fatigue states.

Methodology

The framework of the proposed method MDFG is presented in Fig. 2, divided into three stages: data preprocessing, fine-grained label generation, and network learning, which will be developed in detail below.

Preprocessor Stage

Initially, we utilize the robust face analysis toolkit Insight-Face (Deng et al. 2020) to detect faces and extract the eye movement sequences eye aspect ratio (EAR) (Ghoddosian, Galib, and Athitsos 2019) based on key points for each video. Then, each input eye sequence d_i is divided into m segments using a sliding window of length w and step size $w/2$. The wavelet features of each segment are extracted by the wavelet feature extractor \mathcal{T} . In \mathcal{T} , we apply the db4 wavelet with a maximum decomposition level of 3, and the energy of the signal decomposed at the third level as the fatigue feature indicator $\mathcal{T}(d_i^j)$. This process yields the dataset $D = \{(\mathcal{T}(d_i^j), y_i)\}_{i=1}^N$, where N denotes the total number of eye sequences, $j = 1, \dots, m$. Since y_i is a coarse-grained label and cannot accurately describe the fatigue state of each $\mathcal{T}(d_i^j)$, there is noise in the D . Consequently, we feed it into the subsequent stage for further processing.

Fine-grained Label Generation Stage

Fine-Grained Categorization (FGC). This section extracts trustworthy samples from D and then identifies multiple patterns within the alert and fatigue states. Each pattern corresponds to a specific subtype within its respective state. Detailed analysis of these subtypes helps prevent detection errors associated with oversimplified type distributions in traditional methods. Subsequently, the base class set is extracted from the subtypes, forming the foundation for constructing synthetic intermediate states. Each base class is a distribution that best represents its subtype. Specifically, suppose the alert samples containing noise is S_0 , and the fatigue samples containing noise is S_1 . We first uses SmallLoss mechanism (Han et al. 2018; Park et al. 2024) with HMLSTM as the backbone ϕ_{trust} to extract trustworthy samples, resulting in S'_0 for the alert and S'_1 for the fatigue.

$$S'_0 = \{d_i^j | (\ell(\phi_{\text{trust}}(d_i^j)) < \gamma_0) \text{ if } d_i^j \in S_0\} \quad (1)$$

$$S'_1 = \{d_i^j | (\ell(\phi_{\text{trust}}(d_i^j)) < \gamma_1) \text{ if } d_i^j \in S_1\} \quad (2)$$

$\ell()$ represents the sample loss value, and γ_0 and γ_1 denote the trustworthy thresholds for alert and fatigue, respectively. Clustering is then used to classify S'_0 and S'_1 into A groups and B groups, respectively. Each group represents a pattern, i.e., a subtype. To obtain the base class, a GMM is applied to fit each group. The g_0^a represents the GMM-fitted distribution of the a -th group in the alert samples, as follows:

$$g_0^a = \sum_{k=1}^K (\pi_0^a)_k \mathcal{N}(x_0^a | (\mu_0^a)_k, (\Sigma_0^a)_k) \quad (3)$$

where x_0^a denotes samples in the a -th alert group modeled with K Gaussian components, π_0^a , μ_0^a , and Σ_0^a are the mixture coefficient, mean, and covariance matrix, respectively. g_0^a denotes a subtype distribution. This results in an alert subtype distribution set $G_0 = \{g_0^a\}_{a=1}^A$ and for fatigue is $G_1 = \{g_1^b\}_{b=1}^B$. The calculation of g_1^b is same as g_0^a .

Fine-Grained Quantification (FGQ). G_0 and G_1 represent the sets of subtype distributions corresponding to fatigue scores (FS) of 0 and 1, respectively. For the intermediate states between alert and fatigue study, unlike the approach in (Zhou et al. 2020), which defines intermediate states by averaging PERCLOS fatigue features over a 15 seconds window—specifically, where the average in the first 3 seconds is less than 10 and in the last 3 seconds is greater than 10—this study explores a more fine-grained analysis of fatigue states, considering intermediate states with FS of 0.1, 0.3, 0.5, etc. We propose a method that constructs an intermediate state base class set by base class set synthesis and then quantifies the fatigue level by identifying the synthetic base class set closest to the target object. This approach enables automated fatigue assessment without human intervention. Expressly, for each subtype in G_0 and G_1 , we represent their base class by selecting the distribution with the highest mixing coefficient from the K coefficients. In this way, we obtain the base class sets for alert and fatigue, $G_0^* = \{g_0^{a*}\}_{a=1}^A$ and $G_1^* = \{g_1^{b*}\}_{b=1}^B$.

$$g_0^{a*} = \mathcal{N}(x_0^a | (\mu_0^a)_{k_0^*}, (\Sigma_0^a)_{k_0^*}) \quad (4)$$

$$g_1^{b*} = \mathcal{N}(x_1^b | (\mu_1^b)_{k_1^*}, (\Sigma_1^b)_{k_1^*}) \quad (5)$$

$$k_0^* = \arg \max_k (\pi_0^a)_k, \quad k_1^* = \arg \max_k (\pi_1^b)_k \quad (6)$$

Where g_0^{a*} and g_1^{b*} denote the base class corresponding to one subtype in each state. After obtaining the base class set, a Gaussian distribution-based linear combination method is employed to construct the base class set for intermediate states. Taking g_0^{a*} and g_1^{b*} as examples, these base classes follow multivariate Gaussian distributions, respectively, denoted as $x_0^a \sim \mathcal{N}_p((\mu_0^a)_{k_0^*}, (\Sigma_0^a)_{k_0^*})$ and $x_1^b \sim \mathcal{N}_p((\mu_1^b)_{k_1^*}, (\Sigma_1^b)_{k_1^*})$. For the intermediate state with FS β , we introduce a fatigue weight parameter β (with a value range from 0 to 1), and the corresponding alert weight is $\alpha = 1 - \beta$. The new state z distribution is constructed through a linear combination of the alert and fatigued state, represented as: $z = \alpha x_0^a + \beta x_1^b$. According to the properties of Gaussian distribution, the new distribution resulting

from this linear combination is still a multivariate Gaussian distribution, with its mean and covariance matrix given by:

$$\mu_z = \alpha (\mu_0^a)_{k_0^*} + \beta (\mu_1^b)_{k_1^*} \quad (7)$$

$$\Sigma_z = \alpha^2 (\Sigma_0^a)_{k_0^*} + \beta^2 (\Sigma_1^b)_{k_1^*} \quad (8)$$

Therefore, we obtain the base class g_{ab}^β of FS β , $g_{ab}^\beta = \mathcal{N}(z | \mu_z, \Sigma_z)$. To obtain representative base classes, β is set to multi-group FS, specifically $\{0, 0.1, 0.3, 0.5, 0.7, 0.9, 1\}$, each corresponding to a different intermediate state. The set of base classes generated by these β is referred to as the complete basic class set $\mathcal{G} = \{g_{ab}^\beta\}$, where $a = 1, \dots, A$ and $b = 1, \dots, B$. The FS calculation for each segment d_i^j is $f_{d_i^j} = \mathcal{F}(d_i^j, g_{ab}^\beta)$, which is first computing its probability density value $p(d_i^j | g_{ab}^\beta)$ under each distribution g_{ab}^β , and then subsequently obtaining the values of β^* , a^* , $b^* = \arg \max_{\beta, a, b} (p(d_i^j | g_{ab}^\beta))$, which leads to $f_{d_i^j} = \beta^*$. The bar chart in the FGQ of Fig.2 shows the fatigue quantization sequence f_{d_i} corresponding to each sample d_i .

Network Learning Stage

At this stage, for the fatigue-quantized sequence samples $\{(f_{d_i}, y_i)\}_{i=1}^N$, where $f_{d_i} = \{\mathcal{F}(d_i^j, g_{ab}^\beta)\}_{j=1}^m$, we design a Deep Neural Network (DNN) ϕ_F . It consists of three fully connected layers with 128, 64, and 32 neurons, respectively, each followed by batch normalization, ReLU activation, and Dropout (0.5). A final softmax activation layer ϕ_{c1} is used to generate the fatigue classification results.

In the meta-learning process, to effectively capture the complex fatigue information, pairwise comparisons are introduced. Specifically, within each mini-batch (size N_b), after encoding with ϕ_F , the mini-batch samples are split into a meta-training set and a meta-validation set. Given the class imbalance in the reRLDD and reDROZY datasets used in this paper, a sample f_{d_j} from the minority class is randomly selected and concatenated with its own features to create a positive sample pair. Simultaneously, M samples from other classes in the meta-training set are randomly chosen, and each feature $\phi_F(f_{d_i})$ of these samples is concatenated with $\phi_F(f_{d_j})$ to form negative sample pairs $C(\phi_F(f_{d_i}), \phi_F(f_{d_j}))$. Here, $C(\cdot, \cdot)$ denotes concatenation along the depth dimension. These pairs are then fed into the comparison module ϕ_{meta} and classifier ϕ_{c2} for training. To fully learn ϕ_F , ϕ_{meta} leverages the ϕ_F network. While ϕ_{c2} employs a sigmoid activation function. The meta-learning module is trained by fitting the pairwise score $s_{j,l}$ of $\phi_F(f_{d_i})$ and $\phi_F(f_{d_j})$ to the true pairwise similarity label $y'_{j,l}$, with the loss function defined as cross-entropy:

$$s_{j,l} = \phi_{c2}(\phi_{\text{meta}}(C(\phi_F(f_{d_i}), \phi_F(f_{d_j})))) \quad (l \neq j) \quad (9)$$

$$\mathcal{L}_{\text{meta}} = -\frac{1}{N_b} \sum_{l \neq j} y'_{j,l} \log(s_{j,l}) + (1 - y'_{j,l}) \log(1 - s_{j,l}) \quad (10)$$

where $y'_{j,l} = 1$ if d_j and d_l belong to the same class; otherwise, $y'_{j,l} = 0$.

Multi-group fatigue score β	Group count	reRLDD		reDROZY	
		Acc (Av acc)	F1 (Av F1)	Acc (Av acc)	F1 (Av F1)
[0,1]	2	0.664 (0.608)	0.656 (0.588)	0.706 (0.659)	0.714 (0.665)
[0,0.5,1]	3	0.680 (0.644)	0.670 (0.629)	0.744 (0.723)	0.742 (0.717)
[0,0.3,0.5,0.7,1]	5	0.633 (0.611)	0.609 (0.584)	0.769 (0.736)	0.756 (0.724)
[0,0.1,0.3,0.5,0.7,0.9,1]	7	0.695 (0.655)	0.689 (0.639)	0.775 (0.730)	0.772 (0.719)
[0,0.1,0.2,0.3,0.4,0.5,0.6,0.7,0.8,0.9,1]	11	0.648 (0.617)	0.636 (0.605)	0.756 (0.691)	0.750 (0.684)

Table 2: Results of different FS β groups. The optimal results and the average results (shown in parentheses) are reported.

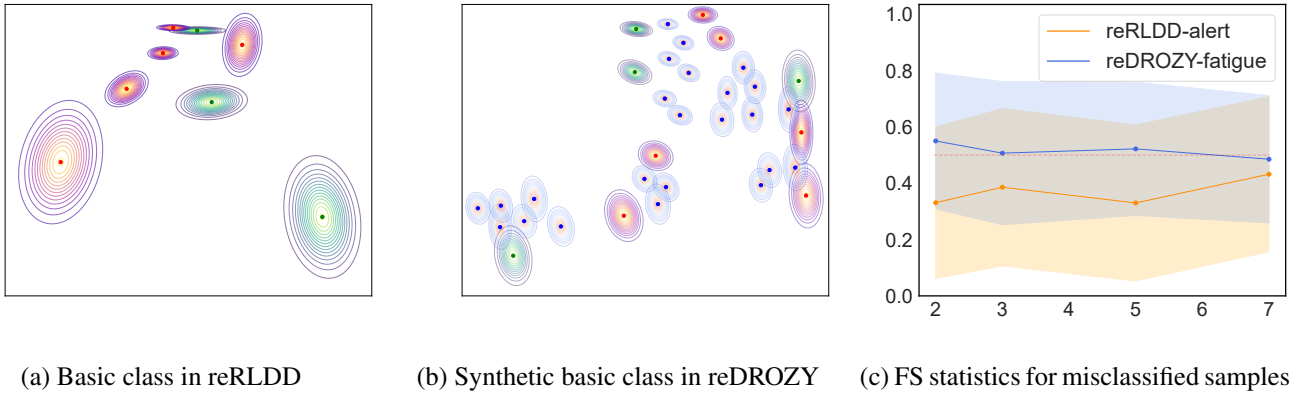


Figure 3: (a) illustrates the base classes extracted from the three alert subtypes and five fatigue subtypes. (b) primarily showcases the TSEN visualization of the 24 synthetic base classes in reDROZY under FS $\beta = 0.5$. These base classes are derived from the base class of four alert and six fatigue. (c) The closer the average FS of misclassified samples is to 0.5, the better the methods' performance. When the Group count is 7, the average FS is closer to 0.5. The blue and orange bands represent the variance.

Experiments

Datasets

This paper conducts experiments on the reRLDD and reDROZY datasets (Wang et al. 2024a), which encompass real fatigue scenarios rather than simulated drowsiness data. The reRLDD dataset comprises 284 alert video instances and 194 fatigue video instances from 12 participants, while the reDROZY dataset consists of 183 alert video instances and 322 fatigue video instances from 14 participants. A notable feature of reDROZY compared to reRLDD is that subjects experienced acute, prolonged wakefulness, leading to gradual sleep deprivation, resulting in a more significant difference between alert and fatigue states.

Baselines

Eye Movement-Based Fatigue Detection Baselines. For eye movement-based fatigue detection methods, We compare Eye4feature (Byrnes and Sturton 2018) and Eye16feature (Massoz et al. 2016), which integrate 4 and 16 eye movement features, respectively, alongside HMLSTM (Ghoddosian, Galib, and Athitsos 2019) and FacialUnits (Cheng et al. 2019). Furthermore, we conduct a comparative analysis of the eye movement feature components in the studies by Lee et al. (2022), Mehmood et al. (2024), and Khandare, Rajak, and Tripathi (2023), despite these studies initially incorporating multiple modalities. Lastly, to

highlight the validity of the multi-dimensional fine-grained approach to fatigue detection, we compare the extracted eye movement wavelet features using temporal classification SOTA methods, including iTransformer (Liu et al. 2023), TimesNet (Wu et al. 2023), TimesURL (Liu and Chen 2024), and MILLET (Early et al. 2023).

Other Fatigue Detection Baselines To comprehensively evaluate the performance of the multi-dimensional fine-grained approach to fatigue detection, we compare it with the methods of Lee et al. (2022) and Mehmood et al. (2024), which utilize multiple features for fatigue detection. Additionally, we compare our method with CNN-LSTM (Liu et al. 2021) and the latest S-ALE (Liao et al. 2024), which detect fatigue based on facial expression features. It is noteworthy that Mehmood et al.'s study primarily focused on extracting geometric features such as mouth corners, head motion, and eye area from each frame, and conducted a statistical significance analysis of these features under alert and fatigued states. There is no classification model. Therefore, to ensure a fair comparison, we extract features at sampling frequencies of 5fps, 10fps, and 15fps while maintaining the learning network consistent with the DNN used in this paper.

Generated FGL verification standard Given the lack of fine-grained objective labels in this study, we establish an indirect validation criterion: for samples that outside the trust-

Method	reRLDD		reDROZY	
	Acc (Av acc)	F1 (Av F1)	Acc (Av acc)	F1 (Av F1)
Eye4feature ¹⁶	0.530 (-)	0.448 (-)	0.538 (-)	0.530 (-)
Eye16feature ¹⁸	0.604 (-)	0.585 (-)	0.624 (-)	0.611 (-)
FacialUints ¹⁹	0.604 (-)	0.585 (-)	0.645 (-)	0.636 (-)
HMLSTM ¹⁹	0.567 (0.539)	0.443 (0.433)	0.684 (0.622)	0.757 (0.709)
Lee et al. (eye) ²²	0.552 (0.552)	0.343 (0.343)	0.760 (0.697)	0.838 (0.795)
Khandare et al. ²³	0.523 (0.477)	0.412 (0.123)	0.678 (0.613)	<u>0.806 (0.749)</u>
Mehmood et al. (eye) _{5fps} ²⁴	0.547 (0.503)	0.442 (0.386)	0.744 (0.668)	0.724 (0.659)
Mehmood et al. (eye) _{10fps}	0.563 (0.525)	0.477 (0.424)	0.744 (0.670)	0.722 (0.663)
Mehmood et al. (eye) _{15fps}	0.563 (0.514)	0.485 (0.398)	0.719 (0.646)	0.705 (0.645)
iTransformer ²⁴	0.573 (0.561)	0.552 (0.539)	0.754 (0.692)	0.750 (0.696)
TimesNet ²³	0.631 (0.498)	0.627 (0.462)	0.754 (0.690)	0.739 (0.684)
TimesURL ²⁴	0.529 (0.511)	0.510 (0.481)	0.684 (0.673)	0.695 (0.684)
MILLET ²⁴	0.573 (0.489)	0.572 (0.456)	0.754 (0.694)	0.731 (0.688)
MDFG (w/o meta)	<u>0.695 (0.655)</u>	<u>0.689 (0.639)</u>	<u>0.775 (0.730)</u>	<u>0.772 (0.719)</u>
MDFG	0.719 (0.688)	0.716 (0.681)	0.781 (0.746)	0.777 (0.737)

Note: (-) indicates the use of SVM in the text without the final average value.

Table 3: Comparison with eye movement-related methods. The best results and the final five-epoch average results (shown in parentheses) are reported. Optimal results are bolded and suboptimal results are underlined.

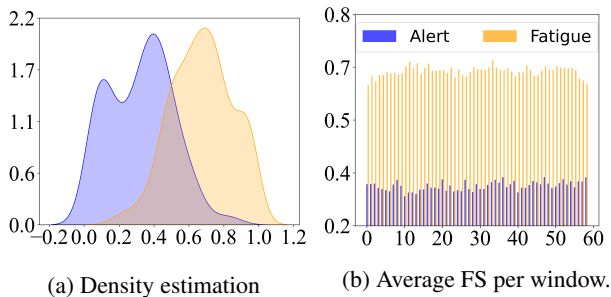


Figure 4: FGQ Related Results. (a) Density estimation of reDROZY. The x-axis is fatigue score. (b) Average FS for segments in sliding windows across different states.

worthy samples, their FS should primarily fall within the interval (0, 1). To validate this criterion, we match these samples to the generated intermediate states and analyze the distribution of FS.

Experiments Settings

The experiment is implemented using PyTorch 1.13.1 on two 3090 GPUs. The ϕ_F network is trained with Adam, using a weight decay of $1e-3$, a learning rate of 0.001, and for 200 epochs. In the experiment, each instance is 30 seconds, with a 1-second sliding window and 0.5-second step size, all sampled at the video’s frame rate. Both datasets are trained using data from the first eight individuals, while the remaining is for testing. In this paper, performance is evaluated using the optimal results accuracy (Acc), F1 score (F1), and the average of the last 5 epochs’ average accuracy (Av Acc), average F1 score (Av F1).

Results

Base Class Related Results. We first present the results related to the base classes. Through kmeans clustering, the subtype counts of alert and fatigue in reRLDD are determined to be 3 and 5, respectively. Fig. 3 (a) presents their base classes, and the significant differences between these base classes indicate that they are excellent representative. Fig. 3 (b) shows the synthetic base classes in reDROZY. When the group count in Tab. 2 is set to 7, the scale of the base classes for alert, intermediate, and fatigue states increases from 10 (4+6), considering only FS 0 and 1, to 130 (4×6×5+10). Notably, the intermediate state base class size increases from 0 to 120, markedly enhancing its expressiveness. In the result, Av acc and Av F1 are improved by 4.7% and 5.1% on reRLDD and by 7.1% and 5.4% on reDROZY, respectively. This further validates the effectiveness of enhancing the intermediate state representation. Here, we did not set β to slightly alert 0.4 and slightly fatigued 0.6 due to their proximity to 0.5, which can lead to misclassification near the middle boundary. In fact, when set to [0, 0.4, 0.6, 1], Av acc and Av F1 on the reDROZY decreased by 2.3% and 2.2% compared to [0, 0.5, 1].

When the average FS of the samples is close to 0.5, classifying these samples is challenging. For misclassified samples, the further the average FS deviates from 0.5, the poorer the method’s performance. Fig. 3 (c) shows that When the group count is set to 7, the average FS of misclassified samples is close to 0.5, indicating optimal performance. In contrast, when set to 2, the two average FS values are 0.550 and 0.381, respectively. This result quantitatively validates the effectiveness of our proposed method in enhancing fine-grained representation. We analyze that when set to 7, not only is the representativeness of the intermediate base class ensured, but the settings of 0.1 and 0.9 play a positive role in guiding the evaluation of data with FS close to 0 and 1.

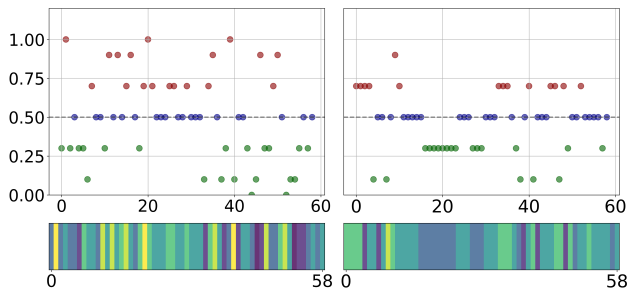


Figure 5: In the case of fatigue, Av FS = 0.49 (left); Alert, Av FS = 0.49 (right). Red, green, and blue represent FS > 0.5, < 0.5, = 0.5, respectively.

FGQ Related Results. Fig. 4 (a) and (b) show the FS statistics at the sample level and segment sample level. In reDROZY, the overall average FS for alert samples is 0.322, while for fatigue is 0.675. In reRLDD, it is 0.406 and 0.662. The significant differences in FS between states indicate that our method can effectively quantify intermediate states.

Regarding the FGL validation criteria, taking the fatigue samples in reDROZY as an example, we set the trusted threshold $\gamma_1 = 0.5$ and $\beta = [0, 0.1, 0.3, 0.7, 0.9, 1]$. The results showed that 84.4% of the samples had FS in the interval (0, 1), with 54.8% of them having $FS > 0.5$. When $\gamma_1 = 0.1$, more real fatigue samples were included in the test, and the results show that 74.3% of samples had $FS > 0.5$. These results support the credibility of the FGL.

Compared to Eye Movement-related Methods. We compare representative eye movement-related methods, and as shown in Tab. 3, the proposed method shows significant improvement in overall performance. We illustrate the following three aspects: 1) **Improvement in Top-1 Test Accuracy.** Without incorporating meta-learning, the proposed method outperforms the TimesNet by 6.4% (15.7%) on reRLDD, and the best-performing Lee et al. (eye) by 1.5% (3.3%) on reDROZY. After the introduction, the performance of MDFG is further improved to 8.8% (19%) on reRLDD and 2.1% (4.9%) on reDROZY, respectively. 2) **Robustness.** The Lee et al. (eye) achieves an F1 of 0.838 on reDROZY but drops to 0.343 in reRLDD, a difference of nearly 50%; TimesNet’s F1 differs by 11.2% between the two datasets, while iTransformer’s difference is nearly 20%. In contrast, the MDFG achieves the best performance on both datasets, with an F1 difference of only 6.1%. 3) **Stability.** The smaller the difference between Acc and Av acc, as well as F1 and Av F1, the better the stability of the model. On reDROZY, the differences for the Lee et al. (eye) study are 6.3% and 4.3% in both terms, while MDFG has differences of 3.5% and 4%. In reRLDD, TimesNet, although excellent on Acc, has a more pronounced difference between the two items, 13.3%, and 16.5%, respectively. In contrast, the differences for MDFG are 3.1% and 3.5%, respectively. Overall, MDFG has excellent stability while maintaining high accuracy, and the detection results are more reliable.

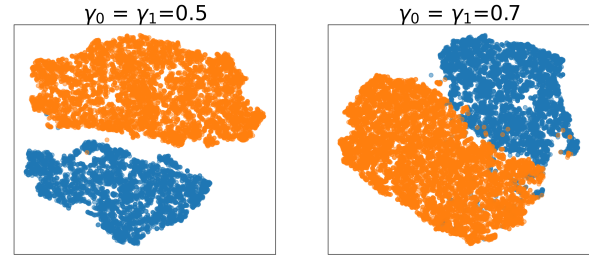


Figure 6: 0.5 is a reliable threshold, with increased alert-fatigue overlap at 0.7. Blue is alert, orange is fatigue.

Method	Acc (Av acc)	F1 (Av F1)
MAML	0.681 (0.637)	0.681 (0.627)
Meta Networks	0.707 (0.664)	0.707 (0.656)
Relation Networks	0.719 (0.659)	0.719 (0.659)
Ours	0.719 (0.688)	<u>0.716 (0.681)</u>

Table 4: Comparison with meta-learning strategies.

Meta-learning Results. In meta-learning, we evaluate three classic methods: model-based MAML (Finn, Abbeel, and Levine 2017), optimization-based Meta Networks (Munkhdalai and Yu 2017), and metric-based Relation Networks (Sung et al. 2018). The results are presented in Tab. 4. It can be observed that the pairwise sample strategy facilitates the learning fatigue knowledge (a strategy adopted in both Relation Networks and our method). However, Relation Networks tends to focus on the relationships of majority classes samples when the data is imbalanced. In this paper, we draw on the paired-sample strategy to transform the binary classification task (fatigue/alert) into a multiple-case classification problem, making more effective use of the limited training data. Compared to Relation Networks, our method achieved an increase of 2.9% and 2.2% in Av acc and Av F1, effectively alleviating this issue.

Fine-Grained Analysis. Fig. 5 presents two challenging cases with average FS close to 0.5. In the fatigue case 1, despite the presence of many alert segments, MDFG correctly identified critical sleepy blink segments with FS of 0.9 and 1. In the alert case 2, MDFG detects almost no segments with FS reaching 0.9, which led to accurate classifications in both cases. In contrast, iTransformers, relying on data-driven approaches, struggled to capture fine-grained fatigue information and adapt to these subtle changes, resulting in misclassification. Further, We select the edge samples with an Average FS between [0.4, 0.6] from the test set for the experiment, as shown in Tab. 5, which demonstrates that MDFG has a significant advantage.

Parameter settings for MDFG. In the trustworthy sample selection, the threshold for alert γ_0 and γ_1 for fatigue are set to 0.5. When the threshold is set to 0.7, the sample mixing, as shown in Fig. 6, can affect the quality of base

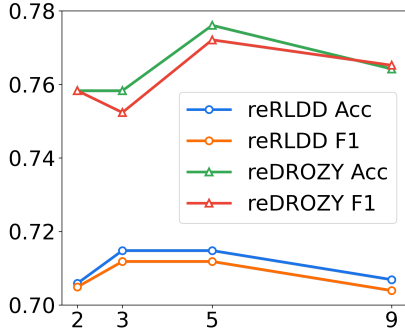


Figure 7: Acc for meta-learning pairwise numbers.

Method	reRLDD		reDROZY	
	Av acc	Av F1	Av acc	Av F1
iTransformers	0.500	0.484	0.672	0.614
MDFG	0.575	0.547	0.678	0.625

Table 5: Results on edge samples.

class generation. In reDROZY, compared to the 0.5, the Acc (Av acc) and F1 (Av F1) decrease by 3.1% (2.6%) and 3.6% (2.4%), respectively. In reRLDD, the decreases are 2.3% (1.0%) and 2.4% (0.5%). In meta-learning, the number of sample pairs for each task (as shown in Fig. 7) is empirically set to 5.

Discussion and Conclusion

Tab. 6 compares the performance of MDFG with other fatigue detection methods that are not limited to eye features. Compared to the superior method by Mehmood et al. (2024), MDFG improves Av acc by 12.2% and Av F1 by 13% on reRLDD. Compared to CNN-LSTM and S-ALE, MDFG exhibits significantly better robustness.

The results in Tab. 3 that only focus on eye features are similar to those in Tab. 4 that include non-eye features, which suggests that eye features are representative in fatigue detection and also validates the rationality of our approach. Given that the performance of Mehmood et al. (2024) improved after introducing multiple features and the effective-

Method	reRLDD		reDROZY	
	Av acc	Av F1	Av acc	Av F1
Lee et al.'22	0.520	0.360	0.620	0.709
Khandare et al.'23	0.499	0.194	0.647	<u>0.742</u>
Mehmood et al.'10fps'24	<u>0.566</u>	<u>0.551</u>	0.728	0.729
CNN-LSTM'21	0.511	0.497	0.714	0.722
S-ALE'24	0.448	0.422	0.783	0.767
MDFG	0.688	0.681	<u>0.746</u>	0.737

Table 6: Comparison of fatigue detection methods. Av acc and Av F1 are reported.

ness of facial features in detecting fatigue, integrating a multimodal method into MDFG could be an interesting and promising avenue for future research.

This paper innovatively proposes to address fatigue detection from the perspective of multi-dimensional fine-grained and firstly defined the concept of MDFGL, including FGT, FGC, and FGQ. Based on the study of MDFGL, we propose a new fatigue detection method, MDFG, which has excellent robustness and stability. Due to technical and cost constraints, manually annotating MDFGL poses a challenge, which is a common issue in tasks such as sentiment analysis and anomaly detection in security monitoring. Our research offers valuable insights into these tasks. In the future, we will integrate multimodal data to enhance the accuracy and generalizability of MDFG, promoting its further development in fatigue detection and various other fields.

Acknowledgments

This work is partially supported by the National Nature Science Foundation of China (No. U22A2035, U1736206, U1803262), and the National Social Science Fund of China (No. 19ZDA113).

References

- Byrnes, A.; and Sturton, C. 2018. On Using Drivers' Eyes to Predict Accident-Causing Drowsiness Levels. In *2018 21st International Conference on Intelligent Transportation Systems (ITSC)*, 2092–2097. IEEE.
- Cheng, Q.; Wang, W.; Jiang, X.; Hou, S.; and Qin, Y. 2019. Assessment of driver mental fatigue using facial landmarks. *IEEE Access*, 7: 150423–150434.
- Dai, L.; Li, Y.; and Zhang, M. 2023. Detection of Operator Fatigue in the Main Control Room of a Nuclear Power Plant Based on Eye Blink Rate, PERCLOS and Mouse Velocity. *Applied Sciences*, 13(4): 2718.
- Deng, J.; Guo, J.; Ververas, E.; Kotsia, I.; and Zafeiriou, S. 2020. Retinaface: Single-shot multi-level face localisation in the wild. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 5203–5212.
- Early, J.; Cheung, G. K.; Cutajar, K.; Xie, H.; Kandola, J.; and Twomey, N. 2023. Inherently interpretable time series classification via multiple instance learning. *arXiv preprint arXiv:2311.10049*.
- Finn, C.; Abbeel, P.; and Levine, S. 2017. Model-agnostic meta-learning for fast adaptation of deep networks. In *International conference on machine learning*, 1126–1135. PMLR.
- Ghodoosian, R.; Galib, M.; and Athitsos, V. 2019. A realistic dataset and baseline temporal model for early drowsiness detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 0–0.
- Gong, P.; Wang, P.; Zhou, Y.; Wen, X.; and Zhang, D. 2024. TFAC-Net: A Temporal-Frequent Attentional Convolutional Network for Driver Drowsiness Recognition With Single-Channel EEG. *IEEE Transactions on Intelligent Transportation Systems*.

- Han, B.; Yao, Q.; Yu, X.; Niu, G.; Xu, M.; Hu, W.; Tsang, I.; and Sugiyama, M. 2018. Co-teaching: Robust training of deep neural networks with extremely noisy labels. *Advances in neural information processing systems*, 31.
- Khandare, H.; Rajak, A.; and Tripathi, R. 2023. A deep learning framework for non-intrusive driver drowsiness detection. In *2023 14th International Conference on Computing Communication and Networking Technologies (ICCCNT)*, 1–5. IEEE.
- Lee, T.; Kim, D.; Park, S.; Kim, D.; and Lee, S.-J. 2022. Predicting mind-wandering with facial videos in online lectures. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2104–2113.
- Li, F.; Chen, C.-H.; Lee, C.-H.; and Feng, S. 2022. Artificial intelligence-enabled non-intrusive vigilance assessment approach to reducing traffic controller’s human errors. *Knowledge-Based Systems*, 239: 108047.
- Liao, J.; Hao, Y.; Zhou, Z.; Pan, J.; and Liang, Y. 2024. Sequence-level affective level estimation based on pyramidal facial expression features. *Pattern Recognition*, 145: 109958.
- Liu, J.; and Chen, S. 2024. Timesurl: Self-supervised contrastive learning for universal time series representation learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, 13918–13926.
- Liu, P.; Chi, H.-L.; Li, X.; and Guo, J. 2021. Effects of dataset characteristics on the performance of fatigue detection for crane operators using hybrid deep neural networks. *Automation in Construction*, 132: 103901.
- Liu, Y.; Hu, T.; Zhang, H.; Wu, H.; Wang, S.; Ma, L.; and Long, M. 2023. iTransformer: Inverted Transformers Are Effective for Time Series Forecasting. *arXiv preprint arXiv:2310.06625*.
- Massoz, Q.; Langohr, T.; François, C.; and Verly, J. G. 2016. The ULg multimodality drowsiness database (called DROZY) and examples of use. In *2016 IEEE Winter Conference on Applications of Computer Vision (WACV)*, 1–7. IEEE.
- Mehmood, I.; Li, H.; Umer, W.; Ma, J.; Shakeel, M. S.; Anwer, S.; Antwi-Afari, M. F.; Tariq, S.; and Wu, H. 2024. Non-invasive detection of mental fatigue in construction equipment operators through geometric measurements of facial features. *Journal of safety research*, 89: 234–250.
- Munkhdalai, T.; and Yu, H. 2017. Meta networks. In *International conference on machine learning*, 2554–2563. PMLR.
- Park, D.; Choi, S.; Kim, D.; Song, H.; and Lee, J.-G. 2024. Robust data pruning under label noise via maximizing re-labeling accuracy. *Advances in Neural Information Processing Systems*, 36.
- Perkins, E.; Sitaula, C.; Burke, M.; and Marzbanrad, F. 2022. Challenges of driver drowsiness prediction: The remaining steps to implementation. *IEEE Transactions on Intelligent Vehicles*, 8(2): 1319–1338.
- Sukumar, N.; Sivashankar, B.; Sumathi, P.; and Maurya, O. P. 2024. Physiological and Physical Sensors for Stress Level, Drowsiness Detection, and Behaviour Analysis. *IEEE Transactions on Consumer Electronics*.
- Sung, F.; Yang, Y.; Zhang, L.; Xiang, T.; Torr, P. H.; and Hospedales, T. M. 2018. Learning to compare: Relation network for few-shot learning. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 1199–1208.
- Trutschel, U.; Sirois, B.; Sommer, D.; Golz, M.; and Edwards, D. 2011. PERCLOS: An alertness measure of the past. In *Driving Assessment Conference*, volume 6. University of Iowa.
- Tu, D.; Basner, M.; Smith, M. G.; Williams, E. S.; Ryder, V. E.; Romoser, A. A.; Ecker, A.; Aeschbach, D.; Stahn, A. C.; Jones, C. W.; et al. 2022. Dynamic ensemble prediction of cognitive performance in spaceflight. *Scientific Reports*, 12(1): 11032.
- Wang, M.; Hu, R.; Zhu, X.; Zhu, D.; and Wang, X. 2024a. Learning with noisy labels for robust fatigue detection. *Knowledge-Based Systems*, 112199.
- Wang, Y.; Turnbull, A.; Xu, Y.; Heffner, K.; Lin, F. V.; and Adeli, E. 2024b. Vision-based estimation of fatigue and engagement in cognitive training sessions. *Artificial Intelligence in Medicine*, 102923.
- Wiehler, A.; Branzoli, F.; Adanyeguh, I.; Mochel, F.; and Pessiglione, M. 2022. A neuro-metabolic account of why daylong cognitive work alters the control of economic decisions. *Current Biology*, 32(16): 3564–3575.
- Wu, H.; Hu, T.; Liu, Y.; Zhou, H.; Wang, J.; and Long, M. 2023. TimesNet: Temporal 2D-Variation Modeling for General Time Series Analysis. In *International Conference on Learning Representations*.
- Yang, C.; Yang, Z.; Li, W.; and See, J. 2022. FatigueView: A Multi-Camera Video Dataset for Vision-Based Drowsiness Detection. *IEEE Transactions on Intelligent Transportation Systems*.
- Yin, H.; Su, Y.; Liu, Y.; and Zhao, D. 2016. A driver fatigue detection method based on multi-sensor signals. In *2016 IEEE winter conference on applications of computer vision (WACV)*, 1–7. IEEE.
- Zeng, W.; Xiao, Y.; Wei, S.; Gan, J.; Zhang, X.; Cao, Z.; Fang, Z.; and Zhou, J. T. 2023. Real-time multi-person eye-blink detection in the wild for untrimmed video. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 13854–13863.
- Zhang, X.; Lu, D.; Pan, J.; Shen, J.; Wu, M.; Hu, X.; and Hu, B. 2020. Fatigue detection with covariance manifolds of electroencephalography in transportation industry. *IEEE Transactions on Industrial Informatics*, 17(5): 3497–3507.
- Zhao, P.; Lian, C.; Xu, B.; and Zeng, Z. 2024. Multiscale Global Prompt Transformer for EEG-Based Driver Fatigue Recognition. *IEEE Transactions on Automation Science and Engineering*.
- Zhou, F.; Alsaid, A.; Blommer, M.; Curry, R.; Swaminathan, R.; Kochhar, D.; Talamonti, W.; Tijerina, L.; and Lei, B. 2020. Driver fatigue transition prediction in highly automated driving using physiological features. *Expert Systems with Applications*, 147: 113204.