ICASSP 2026 - 2026 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) | 979-8-3315-6701-9/26/$31.00 ©2026 IEEE | DOI: 10.1109/ICASSP55912.2026.11463290

A BIMODAL APPROACH FOR DETECTING FATIGUE USING SPEECH AND PERSONAL ASSESSMENTS IN COLLEGE STUDENTS

Kapotaksha Das[⋆] *Mihai Burzo*[•] *John Elson*[§] *Clay Maranville*[§] *Mohamed Abouelenien*[⋆]

[⋆] Computer and Information Science, University of Michigan, USA
[•] Mechanical Engineering, University of Michigan, USA
[§] Ford Motor Company, USA

ABSTRACT

Fatigue detection is crucial in ensuring safety and performance, particularly in contexts such as driving and education, where fatigue can have significant consequences. Traditional methods often rely on visual or physiological cues, which can lack immediacy and objectivity and may be inconvenient. This study introduces a novel approach by integrating audio data with survey-based personal assessments to detect fatigue in college students. Using a multimodal dataset collected under controlled lab conditions, we analyze spontaneous speech to identify vocal characteristics indicative of fatigue. Our findings in this pilot study demonstrate that integrating traditional survey measures with audio features enhances model performance, improving the understanding of individuals' fatigue states. Our approach attains an F1 score of up to 64.62%, suggesting the feasibility of this bimodal framework despite limited sample size. This non-intrusive, scalable method offers a promising avenue for developing personalized and efficient fatigue detection systems with minimal data collection requirements.

Index Terms— Fatigue Detection, Speech Analysis, Multimodal Learning, Survey-based Features, Machine Learning

1. INTRODUCTION

In recent years, the advent of speech technology has opened new avenues for assessing and monitoring human physiological and psychological states such as emotion [1] and stress [2]. Amid growing interest in integrating multimodal data, analyzing human biorhythms through speech offers compelling potential for various applications in healthcare, such as monitoring, diagnosis, and treatment [3]. In 2022, the National Highway Traffic Safety Administration reported 693 fatalities resulting from drowsy driving [4]. However, another study by the AAA indicates that these figures likely underestimate the true impact of driving while mentally or physically fatigued [5]. Moreover, fatigue is a chronic issue amongst college students, with an American College Health Association survey reporting over 70 % of students are tired for three or more days of the week [6]. Accordingly, understanding fatigue, a state with significant implications for performance and health, is a critical challenge that can benefit from innovative approaches using speech data.

Traditional methods of fatigue detection, such as visual cues or physiological measurements [7], often lack immediacy, objectivity, and scalability. In contrast, audio analysis offers a promising alternative, enabling real-time fatigue detection with minimal intrusion. Research has also shown that speech can indicate fatigue in a speaker [8].

Incorporating audio analysis into fatigue detection systems presents numerous benefits, particularly its non-intrusive nature. Audio data can be collected passively without disrupting the user's routine or requiring them to look in specific camera directions, ensuring comfort and compliance. Additionally, this approach is cost-effective due to the widespread availability of built-in microphones, eliminating the need for specialized hardware. The accessibility of audio technology enables its implementation across various platforms, making it a practical choice for diverse applications. Despite the lower detection scores using speech and audio in real-world datasets due to diverse contexts and demographic challenges, as shown by Brueckner et al. [9] and Han et al. [10], audio data integrates seamlessly with existing systems. This allows it to complement other modalities like visual or physiological data for a more comprehensive analysis. This versatility enables widespread application across various environments, from vehicle monitoring to workplace settings, providing efficient and scalable solutions for fatigue management.

This study introduces a novel approach to modeling fatigue using a multimodal dataset collected from college students under controlled lab conditions. By leveraging audio data from spontaneous speech, complemented by survey-based personal assessments, our research aims to develop an effective model for predicting fatigue states. To our knowledge, integrating traditional survey measures with audio features is a novel strategy. This approach enhances model performance and provides a potential method for fine-tuning models for individual subjects without requiring extensive data collection. This innovation could significantly reduce the data demands typically associated with personalized fatigue detection, making the system more efficient and adaptable. While previous studies have explored large-scale speech analysis, they often lack high-density, multi-session data per subject. This study focuses on a deep, within-subject analysis of 12 individuals to validate the specific contribution of personal assessments to speech models.

In particular, we aim to answer the following research questions:

- Can speech-derived audio features alone accurately detect self-reported fatigue states in college students?

- Does augmenting speech features with one-time, survey-based personal assessments (chronotype and sleep quality) improve the accuracy of fatigue detection models?

2. LITERATURE REVIEW

Fatigue detection has been the subject of extensive research across various fields due to its significant implications. Fatigue detection is a critical research area with a broad impact for safety and performance in domains such as driving, aviation, and the workplace. Sikander and Anwar [7] surveyed a wide range of approaches, from rule-based models to advanced machine learning, leveraging phys-

iological signals (e.g., EEG, heart activity, skin conductance) and behavioral cues (e.g., eye or head movements, posture). While EEG can detect mental fatigue within seconds [11], its usage is limited by cumbersome, lab-bound hardware. Wearables, such as accelerometers, photoplethysmography, and skin conductance sensors, offer more practical physiological monitoring [12], but still require dedicated, calibrated equipment [13].

Another significant area of fatigue detection research involves non-contact methods using visual data. Cameras have been employed to detect signs of fatigue through yawning [14] and eye-tracking techniques that monitor extended periods of eye closure [15]. While these methods offer the advantage of contactless monitoring, they are frequently hindered by environmental lighting conditions and the requirement for continuous video observation of subjects, which limits their reliability in diverse real-world settings.

Research into using voice for fatigue detection has produced varying methodologies and results. Gao et al. [16] developed a framework for monitoring fatigue by subjecting participants to 36-hour periods of sleep deprivation, a methodology mirrored by Huang et al. [17] in a study involving air traffic controllers over 48-hour deprivation periods. Although these studies effectively demonstrate voice as a metric for significant fatigue, they reflect extreme conditions not commonly experienced daily.

Efforts to model real-world fatigue conditions include Brueckner et al.'s [9] analysis of phone conversations among elderly Japanese individuals, achieving a 69% accuracy rate. Despite their potential, these studies faced limitations, such as long call segments and relatively small sample sizes. Similarly, Han et al. [10] reported a 42% F1 score in detecting fatigue from speech in COVID-19 patients, further highlighting the challenges of using voice as a reliable indicator of fatigue. Despite their insights, existing research often does not explore various contexts and demographic groups. Our study aims to bridge this gap by concentrating on college students, capturing fatigue levels as they engage in their daily activities. This approach allows us to understand and model real-world fatigue within a diverse, everyday population, providing insights that reflect typical life patterns rather than extreme or isolated conditions.

Research shows a correlation between chronotype and reported fatigue levels [18, 19], particularly in medical contexts where surveys play a crucial role. Traditionally, these surveys have not been integrated with real-time data for the detection of fatigue.

Our work addresses these gaps by studying college students, an under-examined yet high-risk group, using a multimodal dataset that combines spontaneous speech with standardized surveys on chronotype and sleep. By integrating objective vocal features with individual sleep and activity patterns, we aim to enable more nuanced, personalized, and scalable fatigue modeling than previous methods.

3. DATASET DESCRIPTION

We constructed a multimodal dataset to study human biorhythms and classify states among 12 college students, comprising seven males and five females. Data was gathered in a controlled lab environment with consistent temperature, humidity, and lighting, approved by the University's institutional review board. Participants completed six recording sessions over multiple days for diverse data, as shown in Table 1. Table 2 outlines the types of recordings from each session to capture various activity states. Morning sessions were scheduled within 90 minutes of awakening, typically between 7-11 AM. Evening sessions followed post-work, about 8 hours later, between 5-9 PM. Baseline and afternoon sessions were scheduled between 11 AM - 4 PM, according to personal preference. Sessions were spaced out over multiple days. Dual sessions featured morning and evening recordings on the same day, while other sessions happened on different days to prevent data bias. Subjects continued their usual daily routines during the recordings, primarily involving the study and work activities typical of college students.

Table 1. Session Setup

Session Type	**Session Time**
Baseline	Mid-day
Single Sessions (Separate Days)	Morning, Eve, Afternoon
Dual Sessions (Same Day)	Morning, Evening
Total: 6 sessions per subject	

Table 2. Recording Setup

Recording Type	**Description**	**Duration**
Baseline, Silent	Quiet sitting & natural breathing	2.5 min
Active	Free-flowing dialogue	6.0 min
Total: ∼37 minutes of data per subject		

Recordings utilized multiple modalities: physiological sensors to capture heart rate, skin conductance, temperature, and respiration, two RGB cameras (frontal and overhead views), a thermal camera, and audio captured using a Lavalier microphone. Sensors were attached immediately before and detached immediately after each recording session. The audio component was recorded using off-the-shelf equipment not specifically designed for audio pickup, demonstrating the viability of utilizing non-specialized hardware for effective fatigue detection. In particular, for the research presented in this paper, we used audio data from 12 subjects, gathered during their active recording sessions, which totaled approximately 60 sessions and approximately 5.5 hours of data. Each subject spoke about any topic of their choice for six continuous minutes, with no other participants present.

We selected four surveys to complement and potentially augment the audio data: the Karolinska Sleepiness Scale (KSS) [20], the Owl & Lark Questionnaire (OLQ) [21], the Munich Chronotype Questionnaire (MCQ) [22], and the Pittsburgh Sleep Quality Assessment (PSQI) [23]. The KSS was chosen for its ability to capture the subject's immediate level of sleepiness during each session, providing a direct measure of perceived fatigue, which we used as a ground truth for classification. The OLQ identifies a person's chronotype, whether they are a morning or evening type, which can significantly affect energy levels throughout the day and potentially align with detectable changes in speech patterns. The MCQ assesses an individual's natural sleep-wake patterns and circadian preference, offering insights into how these patterns might influence speech and alertness. Lastly, the PSQI evaluates overall sleep quality and disturbances, factors known to impact cognitive and physiological states, thereby enriching the context of the audio data. The KSS survey was administered during every recording session to allow subjects to self-report their perceived levels of tiredness. In contrast, the OLQ, PSQI, and MCQ surveys were administered once during the first baseline session and were used only as features to train the models. Surveys are used as features to supplement speech, but survey information alone is insufficient to predict fatigue. Both fatigue and non-fatigue examples from each subject appear in both training and test sets; thus, survey features do not trivially allow subject identification or label leakage.

4. EXPERIMENTAL SETUP

We use the KSS score to model fatigue, where each subject reported their level of tiredness on a scale of 1 (Very alert) to 9 (Very sleepy)

for every recording session they attended. Data from recording sessions where subjects reported a score of 5 or lower were assigned a "Not Fatigued" label. In contrast, scores of 6 or higher were assigned a "Fatigued" label for classification purposes. The KSS is used exclusively as a self-reported, real-time ground truth for fatigue labeling, in line with prior work [24], not as a detection proxy.

The recordings in the dataset were split into training and testing datasets. All data from a given session (and thus a specific time point and fatigue label) was grouped and either wholly in train or test—never both. Although the same subject has data in both sets, different sessions from different days and fatigue states are split. However, no data from the same recording session would be simultaneously present in both the training and testing splits. Our aim is not person identification, but robust within-subject fatigue prediction relevant for real-world personalization. This meant that, on average, around 15 minutes of audio data was available per subject during training. Two-fold Cross-validation was also employed here by switching the recordings used for training and testing for each subject, allowing for a comprehensive and unbiased analysis of all available data.

We processed the audio data from each recording by segmenting it into 16-second consecutive windows. Sixteen-second segments were selected based on preliminary tests that optimized the trade-off between data volume and the minimum window length required for robust audio feature extraction. This facilitates fair comparisons with other modalities (e.g., physiological and thermal), which also require minimum window lengths for reliable measurement. This resulted in 747 "Not Fatigued" and 515 "Fatigued" segments. To explore higher segment density, we also experimented with 8-second windows, which produced 1,527 "Not Fatigued" segments and 1,054 "Fatigued" segments. The observed class imbalance is common in studies that rely on self-reported scores, as individuals' perceptions of fatigue and alertness vary significantly. Additionally, college students' lifestyle and scheduling patterns may naturally lead to more reports of alertness. While capturing real-world variations, this context presents a classification challenge but offers a realistic depiction of fatigue levels in this population. By incorporating survey data from the subjects, we aim to factor in these individual perceptions, thereby providing a nuanced understanding of how personal sleep and activity patterns influence reported alertness and enhancing the accuracy of our modeling efforts.

We employed Opensmile [25] to extract audio features, using the ComParE_2016 [26] and eGeMAPSv02 [27] feature sets due to their robustness and comprehensive coverage of relevant acoustic properties. The ComParE feature set comprises 6,373 features, providing extensive information on various acoustic aspects and enabling an in-depth analysis of fatigue-related speech characteristics. It encompasses multiple spectral, prosodic, and voice quality attributes, offering a comprehensive view of vocal data. Meanwhile, the eGeMAPS feature set comprises 88 features designed to capture paralinguistic elements, including pitch, formants, loudness, shimmer, speech rate, and articulation speed. These targeted features are well-suited for analyzing subtle changes in speech associated with fatigue. By experimenting with ComParE and eGeMAPS, we aim to identify the most effective features for accurately modeling fatigue-related vocal characteristics.

Survey data required specific processing to transform responses into numerical representations for analysis. Scale-based responses, such as those on a Likert scale, were numerically encoded based on their position (e.g., "Strongly Agree" = 5). Time-based responses, such as reporting typical sleep times, were converted into numerical values as minutes or hours. This preprocessing enabled the inclusion of survey data in feature vectors for machine learning models. When the Survey data was used as part of the personalized feature vector, it was appended to the feature vector by joining on Subject. In this fusion approach, we utilized the MCQ, OLQ, and PSQI surveys to determine if chronotype-focused surveys provided any benefits in augmenting the audio modality for fatigue detection purposes.

Table 3. F1 Scores For Fatigue Detection Using Segmented Speech & Augmented With Surveys

Features Used	**ComParE_2016**		**eGeMAPSv02**	
Window Size	**16s**	**8s**	**16s**	**8s**
Audio Only	60.10%	**60.35%**	59.63%	59.44%
Audio & OLQ	59.99%	58.16%	**63.24%**	61.59%
Audio & PSQI	58.26%	59.30%	**63.66%**	61.05%
Audio & MCQ	**64.07%**	60.16%	64.05%	62.02%
Audio & All Surveys	61.70%	58.44%	**64.62%**	62.38%

Given the dataset size (N=12) and the tabular nature of the extracted feature sets, we opted for a gradient-boosted classifier (XGBoost) [28] rather than deep learning architectures (e.g., Transformers or LSTMs), which are prone to overfitting on smaller cohorts. Similarly, we utilized early fusion (feature concatenation) to directly isolate and evaluate the incremental predictive value of the static survey features against the acoustic baseline. Model performance was evaluated using the F1 score, which is ideal for balancing precision and recall, especially when dealing with imbalanced class distributions, as it considers both false positives and false negatives.

5. EXPERIMENTAL RESULTS & ANALYSIS

Table 3 lists the various F1 scores using the different feature sets for the audio and survey data. Because of the imbalance in the class distributions for the self-reported KSS labels as mentioned in Section 4, the **baseline F1 score is 37.2%**. The table shows that 16-second windows, in general, perform better than 8-second windows. Using 16-second windows, the audio modality alone achieved an F1 score of 60.1% using the ComParE feature set and 59.6% using the eGeMAPS feature set, which is much higher than the baseline score for this task. This performance also surpasses preliminary results obtained when testing other modalities, such as **thermal (44%), physiological (55.6%), and webcam-based visual (59.2%) modalities**. While space constraints preclude a detailed tabular comparison, these audio-based results (F1 $\approx$ 60%) outperform our preliminary benchmarks on the same subjects using thermal (44%) and physiological (55.6%) sensors, validating the efficacy of the non-intrusive audio modality.

Using surveys with the ComParE audio features appears to be less beneficial, except when using the MCQ survey, which increases the F1 score to 64.07%. We observe more significant benefits of using surveys when using the more compact eGeMAPS audio features, where the performance rises to an F1 score of **64.62%**, with the best results achieved by integrating all three surveyed measures. This improvement in prediction performance can be primarily attributed to the additional context provided by the surveys. They offer valuable insights into subjects' typical sleep patterns, daytime alertness, and chronotype, which enrich the interpretation of audio data. These surveys establish personalized baselines, as understanding whether a participant is a morning or evening person and knowing their usual sleep quality can influence expected patterns in audio features such as speech rate and tone.

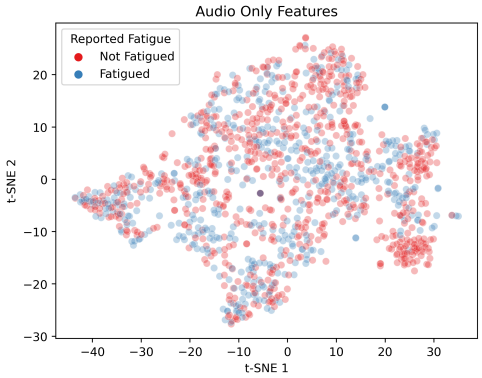

Fig. 1. t-SNE Plot When Using Only eGeMAPS Audio-based Features From 16-second Segments.

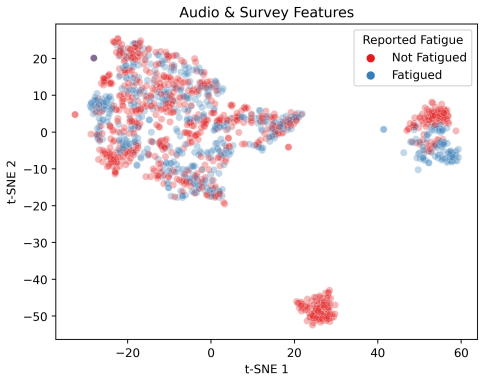

Fig. 2. t-SNE Plot When Using eGeMAPS Audio From 16-second Segments & Survey-based Features.

The t-SNE plots further illustrate the enhanced model performance, a dimensionality reduction technique used to visualize high-dimensional data in a two-dimensional space. These plots differentiate between "Fatigued" and "Not Fatigued" states by representing each data segment as a point on the plot, with the proximity between points indicating similarity in feature space. Fig. 1 shows the t-SNE plot using only eGeMAPS audio-based features, which has limited clustering, reflecting some underlying ability to capture relevant fatigue-associated patterns. However, with the integration of survey data, the plot in Fig. 2 demonstrates an improvement in the separation of data points. This indicates that the surveys provide contextual value, introducing individualized information that complements the real-time nuances captured by audio. While the improvement in separation is modest, it highlights the potential of using surveys to enhance audio-based fatigue detection, enabling a more comprehensive and nuanced analysis that considers both subjective survey responses and objective audio features.

Achieving a 64% F1 score in this study is significant due to the dataset's unique challenges and the task's nature. The self-reported KSS scores introduce variability and subjectivity, as individuals perceive their alertness differently. Moreover, focusing on college students, whose daily routines often vary, adds complexity since their broad range of activities cannot be strictly controlled to ensure explicit fatigue. While this setup enhances the realism of the dataset in capturing natural patterns of fatigue, it also makes the classification task more challenging. Despite these complexities, achieving a 64% F1 score indicates meaningful progress in fatigue detection.

6. CONCLUSION

In this study, we set out to address two primary research questions: (1) whether speech-derived audio features alone can effectively detect self-reported fatigue states in college students, and (2) whether augmenting these features with one-time, survey-based personal assessments reflecting factors such as chronotype and sleep quality could further enhance fatigue detection accuracy.

Our findings demonstrate clear answers to both questions. First, we show that audio analysis serves as a feasible and effective primary modality for detecting fatigue in college students, offering a non-intrusive and accessible alternative to more cumbersome traditional methods. Second, we find that incorporating static survey data does indeed improve model performance, adding critical context from participants' typical sleep patterns and chronotypes that complement the real-time vocal information.

While the subjective nature of self-reported measures and the routine variability among college students present significant modeling challenges, our approach achieved a meaningful increase in F1 score, highlighting both the viability and the benefit of combining multimodal data for this application. Visualization through t-SNE plots further reinforced the value of integrating survey responses with audio features, yielding greater separability of fatigue states.

Ultimately, these results highlight the potential of scalable, multimodal fatigue detection systems that utilize both passive audio data and a limited number of personalized baseline assessments. Future research should explore larger and more diverse populations as well as the integration of additional modalities, continually refining this framework for real-world, customized health monitoring and fatigue management.

7. LIMITATIONS & FUTURE WORK

We acknowledge that the current cohort size (N=12) limits broad generalizability, although the high density of longitudinal data enables effective within-subject modeling. Consequently, we interpret these results as a validation of the *methodology* of augmenting speech with static traits. Future work will focus on scaling the dataset to include diverse populations and additional modalities (thermal, visual, and physiological). Additionally, we aim to refine survey instruments and employ deep learning frameworks to better integrate these multimodal and static features.

8. ACKNOWLEDGMENT

This material is based in part upon work supported by the Ford Motor Company. Any opinions, findings, conclusions, or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of Ford Motor Company or any other Ford entity.

9. REFERENCES

[1] Taiba Majid Wani, Teddy Surya Gunawan, Syed Asif Ahmad Qadri, Mira Kartiwi, and Eliathamby Ambikairajah, "A comprehensive review of speech emotion recognition systems," *IEEE access*, vol. 9, pp. 47795–47814, 2021.

[2] John HL Hansen and Sanjay Patil, "Speech under stress: Analysis, modeling and recognition," *Speaker classification I: Fundamentals, features, and methods*, pp. 108–137, 2007.

[3] Siddique Latif, Junaid Qadir, Adnan Qayyum, Muhammad Usama, and Shahzad Younis, "Speech technology for healthcare: Opportunities, challenges, and state of the art," *IEEE Reviews in Biomedical Engineering*, vol. 14, pp. 342–356, 2020.

[4] National Highway Traffic Safety Administration (NHTSA), USDOT, "Drowsy Driving," `www.nhtsa.gov/risky-driving/drowsy-driving`, 2023.

[5] Brian C Tefft et al., *Prevalence of motor vehicle crashes involving drowsy drivers, United States, 2009-2013*, AAA Foundation for Traffic Safety Washington, DC, 2014.

[6] American College Health Association, "Acha-ncha iii: Reference group executive summary spring 2024," Tech. Rep., ACHA, Silver Spring, MD, 2024.

[7] Gulbadan Sikander and Shahzad Anwar, "Driver fatigue detection systems: A review," *IEEE Transactions on Intelligent Transportation Systems*, vol. 20, no. 6, pp. 2339–2352, 2018.

[8] Jeffrey Whitmore and Stanley Fisher, "Speech during sustained operations," *Speech Communication*, vol. 20, no. 1-2, pp. 55–70, 1996.

[9] Raymond Brueckner, Misa Takegami, Namhee Kwon, Nate Blaylock, Vinod Subramanian, Eri Kiyoshige, Soshiro Ogata, Yuriko Nakaoku, Henry O'Connell, and Kunihiro Nishimura, "Voice technology to identify fatigue from japanese speech," in *Proc. SMM23, Workshop on Speech, Music and Mind 2023*, 2023, pp. 31–35.

[10] Jing Han, Kun Qian, Meishu Song, Zijiang Yang, Zhao Ren, Shuo Liu, Juan Liu, Huaiyuan Zheng, Wei Ji, Tomoya Koike, Xiao Li, Zixing Zhang, Yoshiharu Yamamoto, and Björn W. Schuller, "An early study on intelligent analysis of speech under covid-19: Severity, sleep quality, fatigue, and anxiety," in *Interspeech 2020*, 2020, pp. 4946–4950.

[11] François Laurent, Mario Valderrama, Michel Besserve, Mathias Guillard, Jean-Philippe Lachaux, Jacques Martinerie, and Genevieve Florence, "Multimodal information improves the rapid detection of mental fatigue," *Biomedical Signal Processing and Control*, vol. 8, no. 4, pp. 400–408, 2013.

[12] Hongyu Luo, Pierre-Alexandre Lee, Ieuan Clay, Martin Jaggi, and Valeria De Luca, "Assessment of fatigue using wearable sensors: a pilot study," *Digital biomarkers*, vol. 4, no. Suppl 1, pp. 59–72, 2020.

[13] Aqsa Mehreen, Syed Muhammad Anwar, Muhammad Haseeb, Muhammad Majid, and Muhammad Obaid Ullah, "A hybrid scheme for drowsiness detection using wearable sensors," *IEEE Sensors Journal*, vol. 19, no. 13, pp. 5119–5126, 2019.

[14] Xiao Fan, Bao-Cai Yin, and Yan-Feng Sun, "Yawning detection for monitoring driver fatigue," in *2007 international conference on machine learning and cybernetics*. IEEE, 2007, vol. 2, pp. 664–668.

[15] H Wang, LB Zhou, and Y Ying, "A novel approach for real time eye state detection in fatigue awareness system," in *2010 IEEE Conference on Robotics, Automation and Mechatronics*. IEEE, 2010, pp. 528–532.

[16] Xiujie Gao, Kefeng Ma, Honglian Yang, Kun Wang, Bo Fu, Yingwen Zhu, Xiaojun She, and Bo Cui, "A rapid, non-invasive method for fatigue detection based on voice information," *Frontiers in Cell and Developmental Biology*, vol. 10, pp. 994001, 2022.

[17] Zhousheng Huang, Weizhen Tang, Qiqi Tian, Ting Huang, and Jinze Li, "Air traffic controller fatigue detection based on facial and vocal features using long short-term memory," *IEEE Access*, 2024.

[18] Łukasz Mokros, Joanna Miłkowska-Dymanowska, Łukasz Gwadera, Tadeusz Pietras, and Wojciech Piotrowski, "Chronotype and the big-five personality traits as predictors of chronic fatigue among patients with sarcoidosis. a cross-sectional study," *Sarcoidosis, Vasculitis, and Diffuse Lung Diseases*, vol. 40, no. 2, 2023.

[19] Adrian A Chrobak, Jarosław Nowakowski, Małgorzata Zwolińska-Wcisło, Dorota Cibor, Magdalena Przybylska-Feluś, Katarzyna Ochyra, Monika Rzeźnik, Alicja Dudek, Aleksandra Arciszewska, Marcin Siwek, et al., "Associations between chronotype, sleep disturbances and seasonality with fatigue and inflammatory bowel disease symptoms," *Chronobiology international*, vol. 35, no. 8, pp. 1142–1152, 2018.

[20] Torbjörn Åkerstedt and Mats Gillberg, "Subjective and objective sleepiness in the active individual," *International Journal of Neuroscience*, vol. 52, no. 1-2, pp. 29–37, 1990.

[21] Jim A Horne and Olov Ostberg, "A self-assessment questionnaire to determine morningness-eveningness in human circadian rhythms.," *International journal of chronobiology*, vol. 4, no. 2, pp. 97–110, 1976.

[22] Till Roenneberg, Anna Wirz-Justice, and Martha Merrow, "Life between clocks: daily temporal patterns of human chronotypes," *Journal of biological rhythms*, vol. 18, no. 1, pp. 80–90, 2003.

[23] Daniel J Buysse, Charles F Reynolds III, Timothy H Monk, Susan R Berman, and David J Kupfer, "The pittsburgh sleep quality index: a new instrument for psychiatric practice and research," *Psychiatry research*, vol. 28, no. 2, pp. 193–213, 1989.

[24] Shuyan Hu and Gangtie Zheng, "Driver drowsiness detection with eyelid related parameters by support vector machine," *Expert Systems with Applications*, vol. 36, no. 4, pp. 7651–7658, 2009.

[25] Florian Eyben, Martin Wöllmer, and Björn Schuller, "Opensmile: the munich versatile and fast open-source audio feature extractor," in *Proceedings of the 18th ACM International Conference on Multimedia*, New York, NY, USA, 2010, MM '10, p. 1459–1462, Association for Computing Machinery.

[26] Florian Eyben, Felix Weninger, Florian Gross, and Björn Schuller, "Recent developments in opensmile, the munich open-source multimedia feature extractor," in *Proceedings of the 21st ACM international conference on Multimedia*, 2013, pp. 835–838.

[27] Florian Eyben, Klaus R Scherer, Björn W Schuller, Johan Sundberg, Elisabeth André, Carlos Busso, Laurence Y Devillers, Julien Epps, Petri Laukka, Shrikanth S Narayanan, et al., "The geneva minimalistic acoustic parameter set (gemaps) for voice research and affective computing," *IEEE transactions on affective computing*, vol. 7, no. 2, pp. 190–202, 2015.

[28] Tianqi Chen and Carlos Guestrin, "Xgboost: A scalable tree boosting system," *CoRR*, vol. abs/1603.02754, 2016.