Kaveti Pavan* and Nagarajan Ganapathy

Assessment of Driver Stress using Multimodal wereable Signals and Self-Attention Networks

https://doi.org/10.1515/cdbme-2024-2090

Abstract: Assessment of driver stress, crucial for road safety, can greatly benefit from the analysis of multimodal physiological signals. However, fusing such heterogeneous data poses significant challenges, particularly in intermediate fusion where noise can also be fused. In this study, we address this challenge by exploring a 1D convolutional neural network (CNN) with self-attention mechanisms on multimodal data. Electrocardiogram (ECG) signals (256 Hz) and respiration (RESP) signals (128 Hz) were obtained from ten subjects using textile electrodes while driving in different scenarios, namely normal driving and phone usage (calling). The obtained multimodal data is preprocessed and then applied to a self-attention mechanism (SAM) CNN (SAMcNN) to identify driver stress. Experiments are validated using Leave-one-out-subject cross validation. The proposed approach is capable of classifying driver stress. It is observed that shorter segments yield an accuracy of 64.16% compared to longer segment lengths. Thus, exploring self-attention mechanisms for multimodal signals using wearable shirts facilitates non-intrusive monitoring in real-world driving scenarios

Keywords: Driver stress, Multimodal signals, CNN, SAM, Real-world driving, Non-intrusive monitoring

1 Introduction

One of the significant contributors to road traffic accidents, which result in numerous injuries and fatalities, is the presence of stress while driving. Stress can be defined as a nonspecific physiological response to a combination of external demands and internal concerns [1].

It is important to distinguish between two types of stress: eustress, which is positive and correlates with life satisfaction, and distress, which is negative and opposite in mental state [12].

In this study, the term "stress" typically refers to negative stress. According to the World Health Organization (WHO), traffic accidents cause approximately 1.3 million deaths annually [2]. The European Commission estimates that the cost of car accidents in Europe amounts to 160 billion euros, with 60%–80% attributed to drivers' psychophysical condition [3].

Stress significantly contributes to poor psychophysical condition, increasing the risk of accidents almost tenfold.Stress undermines drivers' cognitive abilities, leading to compromised driving performance. Therefore, to mitigate the risk of accidents and enhance driving safety, it is necessary to develop a system capable of accurately detecting drivers' stress levels

The literature contains numerous stress monitoring systems that analyze various physiological signals, including heart rate, heart rate variability, pupil dilation, blood pressure, respiration rate, and Galvanic Skin Response [4–6]. These systems investigate the complex relationship between these physiological indicators and individual stress levels.

Extensive research has established a strong correlation between physiological measures like electrocardiography (ECG), electrodermal activity (EDA), and respiration rate, and driver stress [7]. These studies highlight the connection between physiological responses and stress in driving scenarios, emphasizing the importance of leveraging these signals for effective stress detection.

Utilizing these physiological markers, stress can be identified to mitigate stress-related risks on the road, contributing to driving safety. By understanding the relationship between physiological signals and stress levels, interventions can be addresed to prevent accidents, monitoring drivers well-being and reducing the likelihood of road incidents

In this paper, we introduce a Self-Attention Multimodal Convolutional Neural Network (SAMcNN), for driver stress detection, integrating a self-attention mechanism with convolution neural network (CNN). Previous studies focusing on multimodal signal fusion, our approach extends this by incorporating attentive mechanisms directly into multimodal physiological signals. Our research address the following questions:
RQ1) How does the integration of self-attention mechanisms in combining multiple physiological signals across various segments improve the assessment of driver stress?
RQ2) How do self-attention mechanisms perform compared to single-modality attentive mechanisms in multimodal signal processing for driver stress detection?

***Corresponding author: Kaveti Pavan,** Indian institute of technology, Hyderabad, India, e-mail: bm23resch01001@iith.ac.in
Nagarajan Ganapathy, Indian institute of technology, Hyderabad, India, e-mail: gnagarajan@bme.iith.ac.in

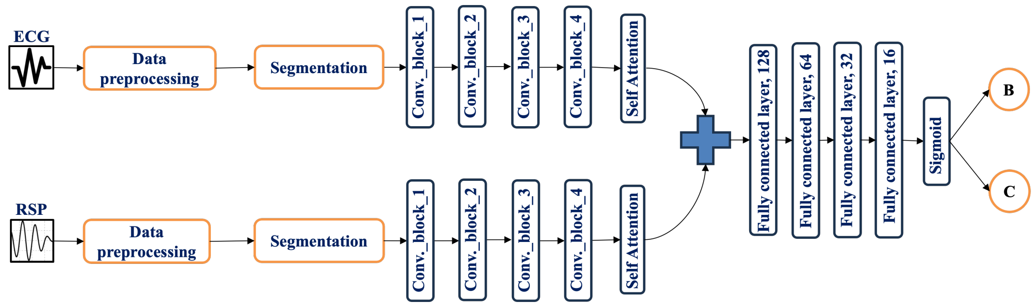

Fig. 1: Proposed SAMcNN-based Deep Learning Framework for Driver stress

2 Methodolgy

2.1 Framework

The proposed end-to-end methodology see Figure .1. The input consists of ECG and RESP signals, segmented into non-overlapping 30, 10 and 15 seconds (sec) windows, respectively. These segments undergo processing by a four-layer 1DCNN network, incorporating attentive information through a self-attention network. The resulting features are then concatenated and inputted into the final classifier

2.2 Dataset details

In this study, 10 male volunteers (average age: 25.1±0.9 years, weight: 73±11.54 kg, height: 173±5.48 cm) wore Hexoskin Pro Kit smart shirts (Carre Technology, Canada) [8] equipped with sensors recording single-lead ECG data at 256 Hz and dual-channel 128 Hz breathing data. Data collection occurred at IIT Hyderabad using a Renault TRIBER car, where volunteers drove in Baseline (sitting) and Calling scenarios. Prior to participation, volunteers provided written consent and received instructions on using the smart shirts.

2.3 Self attention based driver stress

2.3.1 Self attention mechanism

An attention mechanism in multimodal physiological signals refers to a mechanism that allows a model to focus on relevant parts of the input data while disregarding irrelevant parts. It assigns weights to different parts of the input data based on their importance for the task at hand. Self-attention mechanism, specifically, calculates attention weights based solely on the input data itself, without any external context. This is achieved by computing attention scores between each pair of elements in the input sequence and then aggregating them to obtain the final attention weights. The equation for self-attention mechanism [11] can be represented as follows:

$$\text{Attention}(Q,K,V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{1}$$

The input data is transformed into query (Q), key (K), and value (V) matrices.The softmax function computes attention weights by applying the dot product of query and key vectors, divided by $\sqrt{d_k}$, where d_k is the dimensionality of the key vectors

2.4 Data preprocessing

The acquired signals were preprocessed using the Neurokit2 toolbox [9]. ECG signals were filtered using a fifth-order Butterworth filter with a low-pass cutoff frequency of 0.5 Hz, while RESP signals underwent filtering with a second-order Butterworth filter with high-pass cutoff frequency of 3 Hz and low-pass cutoff frequency of 0.05 Hz.

2.4.1 Architecture details

The analysis utilizes an SAMcNN framework, see Figure. 1, incorporating self-attention network (SA) layers. CNNs are chosen for their effectiveness in capturing temporal patterns, while SA is employed to extract attentive information from multimodal physiological signals. In the proposed framework, the CNN consists of four convolutional blocks with a kernel size of 9 and filter numbers of (16, 32, 64, 128), followed by batch normalization, average pooling layers of size 2, and a dropout rate of 0.3. Fully connected layers with sizes of 128, 64, 32, and 16 follow, utilizing ReLU activation. The softmax output layer identifies two classes using categorical cross-entropy loss and the RMSPROP optimizer. Training is done for 30 epochs with a batch size of 9 and a learning rate of 0.001. Model is evaluated using Leave-One-Subject-Out-Cross-Validation (LOSOCV) [10].

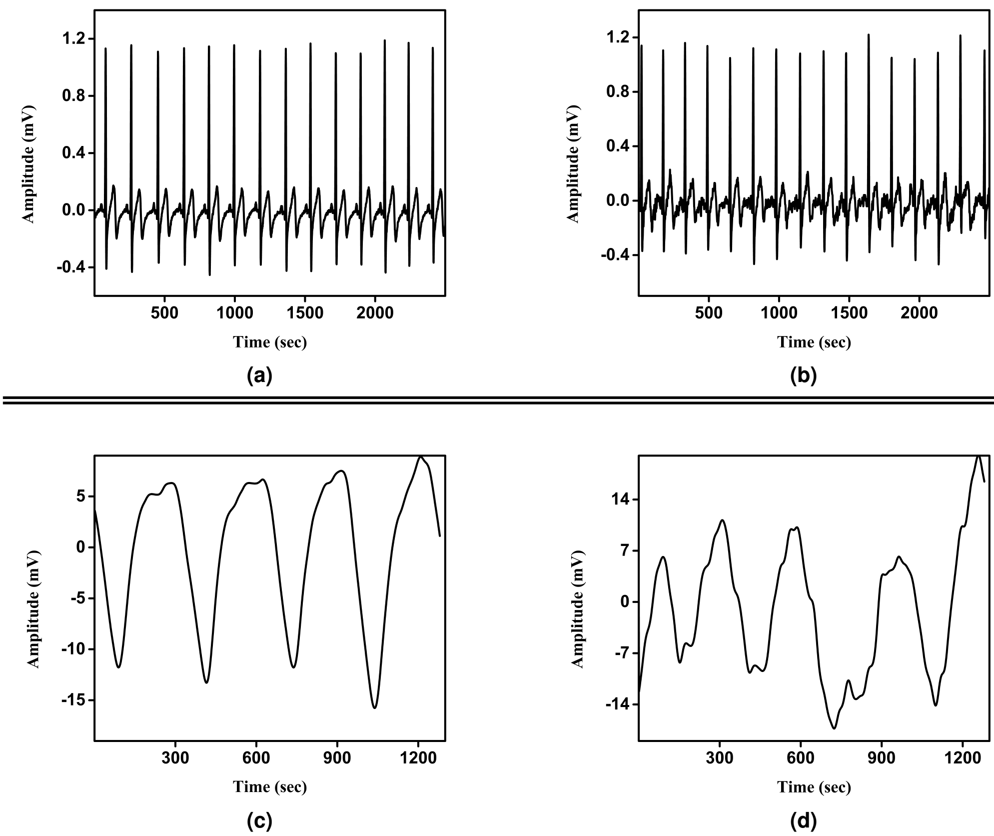

Fig. 2: Raw ECG and RESP signals with observed artifacts in ECG signals: **(a)** ECG of Baseline class, **(b)** ECG of Calling class, **(c)** RESP of Baseline class, **(d)** RESP of Calling class.

3 Results

See Figure .2 for raw ECG and RESP signals, categorized into baseline and calling classes, where ECG signals reveal characteristic waves (P, QRS, T) and RESP signals exhibit rhythmic breathing patterns. During our analysis, we identified artifacts such as motion, loose contact, and muscle interference, which are crucial considerations for signal processing. To mitigate class data imbalances, we uniformly sampled data across subjects and tasks, aligning with the shortest time interval of 3 minutes.

Tab. 1: Proposed framework performance on different segment lengths by using LOSOCV validation

Segment length	Precision	Recall	F-score	Accuracy (%)
10s	0.70	0.64	0.57	64.16
15s	0.58	0.62	0.56	62.91
30s	0.30	0.51	0.36	51.66

Our proposed framework's performance (see Table .1) across various segment lengths, employing LOSOCV. Utilizing multimodal ECG and RESP signals, processed by SAMcNN, we classify driver stress during for baseline and calling classes. Results indicate the highest performance with a 10-second segment, achieving an average precision of 0.70, recall of 0.64, F-score of 0.57, and accuracy of 64.16%. Performance slightly declines with longer segment lengths

Tab. 2: Model performance for different SA connection for 10 sec model

SA connections	Precision	Recall	F-score	Accuracy (%)
ECG with SA	0.49	0.58	0.50	59.33
RESP with SA	0.57	0.60	0.54	60.00
ECG,RESP with SA	0.70	0.64	0.57	64.16
ECG,RESP without SA	0.58	0.60	0.54	60.27

The performance of different self-attention (SA) connections within the SAMcNN framework for 10 sec model (see Table .2). It assesses ECG with SA, RESP with SA, and combined ECG and RESP with SA, alongside a scenario without SA. Notably, combining ECG and RESP with SA yields the highest with an average precision (0.70), recall (0.64), F-score (0.57), and accuracy (64.16%).

4 Discussion

In this study, during the sessions stress is induced through phone conversations while driving, with each session lasting approximately an hour from overall data.

In response to **RQ1**, integrating self-attention mechanisms to combine multiple physiological signals across shorter segments significantly enhances the assessment of driver stress. Results in Table 1 indicate that incorporating self-attention mechanisms within the SAMcNN framework substantially improves stress detection accuracy. Specifically, combining ECG and RESP signals with self-attention achieves an average precision, recall, F-score, and accuracy, indicating the effectiveness of self-attention in capturing relevant features for stress detection during driving scenarios, such as phone usage while driving.

In response to **RQ2**, self-attention mechanisms offer advantages over single-modality attentive mechanisms in multimodal signal processing for driver stress detection. These mechanisms enable dynamic weighing of input signals, focusing on the most relevant information for stress detection, which is particularly beneficial in the context of monitoring multiple physiological signals for stress assessment during driving.

Results in Table 2 show that integrating information from diverse sources yields the highest with an average precision, recall, F-score, and accuracy, underscoring the superiority of self-attention mechanisms over single-modality attentive mechanisms in improving stress assessment performance during driving scenarios.

Our findings shows the importance of integrating self-attention mechanisms in multimodal signal processing for driver stress detection in real-world driving scenarios. Despite encountering artifacts during data collection, such as motion and muscle interference see Figure .2. The integration of self-attention mechanisms proved effective in discerning relevant features from noisy data, contributing to improved accuracy levels ranging from 60 to 64 %.

5 Conclusion

In this study, we assessed self-attention mechanisms for ECG and RESP signals collected using a textile wearable device for unobtrusive driver stress monitoring. Conducted with 10 subjects in a controlled environment, our pipeline achieved 64.16% accuracy, demonstrating the effectiveness of self-attention for classification.

References

[1] Selye H. Stress without distress. InPsychopathology of human adaptation 1974 (pp. 137-146). Boston, MA: Springer US.

[2] Sauerzapf VA. Road traffic crash fatalities: An examination of national fatality rates and factors associated with the variation in fatality rates between nations with reference to the World Health Organisation Decade of Action for Road Safety 2011-2020 (Doctoral dissertation, University of East Anglia).

[3] Vivoli, R., Bergomi, M., Rovesti, S., Bussetti, P., Guaitoli, G. M. (2006). Biological and Behavioral Factors Affecting Driving Safety. Journal of Preventive Medicine and Hygiene 2006, 47, (pp. 69–73)

[4] Dan Conway, Ian Dick, Zhidong Li, Yang Wang, and Fang Chen. 2013. The effect of stress on cognitive load measurement. In IFIP Conference on Human-Computer Interaction. Springer, 659–666.

[5] Jennifer A Healey and Rosalind W Picard. 2005. Detecting stress during realworld driving tasks using physiological sensors. IEEE Transactions on intelligent transportation systems 6, 2 (2005), 156–166.

[6] Nermine Munla, Mohamad Khalil, Ahmad Shahin, and Azzam Mourad. 2015. Driver stress level detection using HRV analysis. In Advances in Biomedical Engineering (ICABME), 2015 International Conference on. IEEE, 61–64

[7] Mohammad Naim Rastgoo, Bahareh Nakisa, Andry Rakotonirainy, Vinod Chandran, and Dian Tjondronegoro. 2018. A critical review of proactive detection of driver stress levels based on multimodal measurements. ACM Computing Surveys (CSUR) 51, 5 (2018), 88

[8] Villar R, Beltrame T, Hughson RL. Validation of the Hexoskin wearable vest during lying, sitting, standing, and walking activities. Applied Physiology, Nutrition, and Metabolism. 2015;40(10):1019-24.

[9] Functions—NeuroKit 0.1.7 documentation," Readthedocs.io. [Online]. Available: https://neurokit2.readthedocs.io/en/latest/functions.html

[10] Varma H, Ganapathy N, Deserno TM. Video-based driver emotion recognition using hybrid deep spatio-temporal feature learning. InMedical Imaging 2022: Imaging Informatics for Healthcare, Research, and Applications 2022 Apr 4 (Vol. 12037, pp. 57-63).

[11] Zhang X, Wei X, Zhou Z, Zhao Q, Zhang S, Yang Y, Li R, Hu B. Dynamic alignment and fusion of multimodal physiological patterns for stress recognition. IEEE Transactions on Affective Computing. 2023 Jun 28.

[12] Chung WY, Chong TW, Lee BG. Methods to detect and reduce driver stress: a review. International journal of automotive technology. 2019 Oct;20:1051-63.