Year 10
Final Project Report
2021-2022

Center for Visual & Decision Informatics

Project 10a.005.UL_TAU - Improving Drowsiness and Fatigue Prediction with Multi-modal Sensing and Deep Learning

A National Science Foundation
University Cooperative Research Center

Project # - Project Title

Contents

Personnel

- Principal Investigator: Raju Gottumukkala
- Other team member's first and last name **Moncef Gabbouj – Professor (Co-PI) ,Satya Katragadda – Research Scientist (Co-PI), RaviTeja Bhupatiraju – Research Scientist (Co-PI), Seyed Majid Hosseini – PhD Student**
- Sponsoring IAB member's first and last name Iftikar Ahmad (TietoEvry)

Executive Summary/Abstract

According to the National Highway Traffic Safety Administration, drowsiness is a traffic safety hazard, which results in about 100,000 police-reported crashes each year. These crashes result in more than 1,550 fatalities and 71,000 injuries. Unobtrusive drowsiness detection methods can prevent catastrophic crashes by warning or assisting the drivers [1]. Recently, there have been several efforts at investigating drowsiness based on the facial features of the drivers [2]. Early drowsiness detection is still a challenge. There are no publicly available datasets for early driver drowsiness detection.

We designed and developed a driving simulator for multi-modal driver drowsiness detection and prediction system by fusing video and various other biometric signals (i.e. posture, wrist wearable and pressure sensors). The system includes equipment, hardware and software components that includes (driving simulator, multi-modal sensing modules, software for emotion and driver drowsiness detection). We also developed IRB application for conducting the study, and evaluated the simulator.

Goals and Objectives

The goal of this project is to develop and evaluate multi-modal driver's drowsiness prediction system by fusion of video and different modalities like Biometric signals, posture, pressure signals of drivers. Specific objectives include:

- Design and implementation of a driving simulator
- Design and implementation of multi-modal sensing system for driver drowsiness monitoring and detection
- Develop a study design to simulate drowsiness during driving
- Obtain IRB approval for human subject study for drowsiness
- Preliminary testing and evaluation of various components for the driving drowsiness detection system

Differences from Current State of Art

The proposed simulator collects the data of several sensors in real-time.

Figure 1 shows the driving simulator and the multimodal data collection system

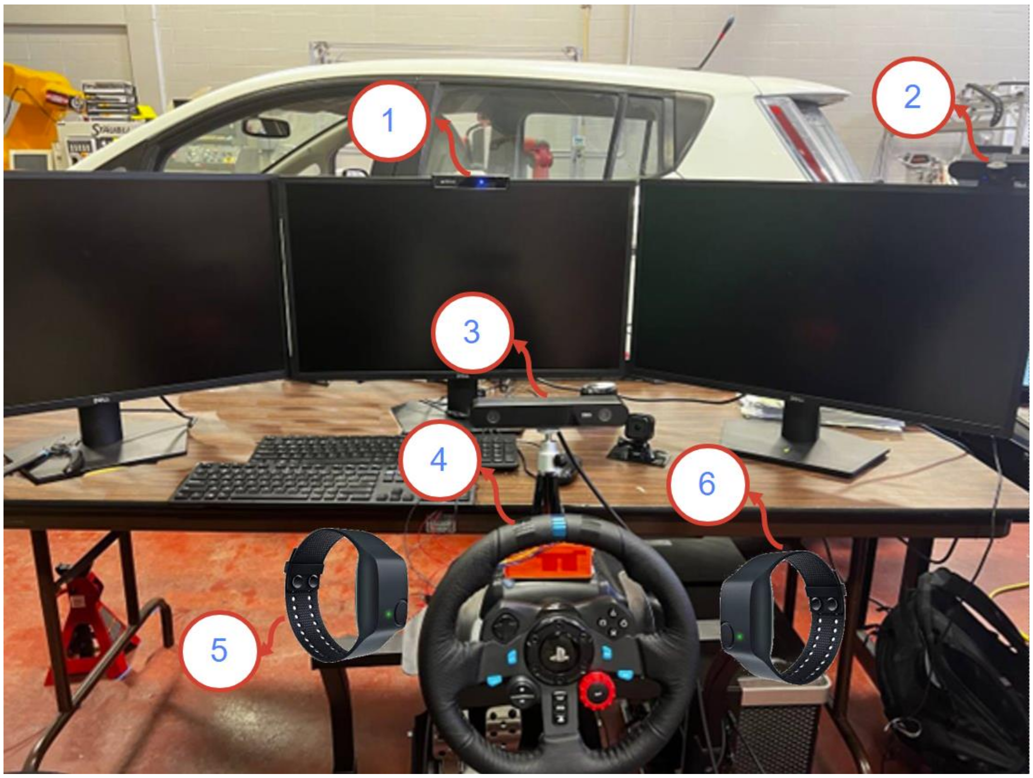

Figure 1: Simulator with different sensors and data collection system

Conventional drowsiness detection models detect the drowsiness using video features (e.g. nodding, blink rate, etc) that are highly influenced by the environmental conditions. The effect of condition and the data collection and sensor problem can disable the model or decrease the accuracy of the system. The proposed models employ the physiological signals along with video and behavioral data (posture and driving behaviors) to improve the performance of drowsiness detection modules. The proposed model is robust in the absence of one or some modalities.

Moreover, the existing drowsiness systems are detecting the drowsiness using facial signs (e.g. nodding, low blink rate). However, the need for early drowsiness detection is felt. In one hand, early drowsiness prediction facilitates the route management for the drivers. On the other hand, the early drowsiness prediction reduces the human error probability that is responsible for 90 percent of accidents investigated [1].

The proposed model fuses video and biometric signals using multi-headed attention that is robust to noise, artifacts, and missing data. The proposed model fuses video and different signals separately that studies the effect of each modality on drowsiness detection without increasing the cost drastically.

Methods and Datasets

Most existing datasets do not provide videos of the participants due to privacy concerns and DDD [2] dataset only provides the videos of the participants in a library. We collected a publicly available dataset that consists of several signals. Table 1 shows the different signals and their frequencies.

No	Signal	Device	Frequency
1	ElectroDermal Activity	Empatica E4	4 Hz
2	Heart Rate	Empatica E4	1 Hz
3	Skin Temperature	Empatica E4	4 Hz
4	3D Accelerometer data	Empatica E4	32 Hz
5	IBI	Empatica E4	---
6	BVP	Empatica E4	32 Hz
7	Facial depth features	Zed2 Camera	30 fps
8	Facial IR video	IR camera	30 fps
9	Posture features	RGB camera	30 fps
10	Grip pressure sensor	PR module	50 Hz
11	Driving behavior	Driving simulator detailed data	---

Figure 1 shows the different approaches that EmpathicDriver dataset covers.

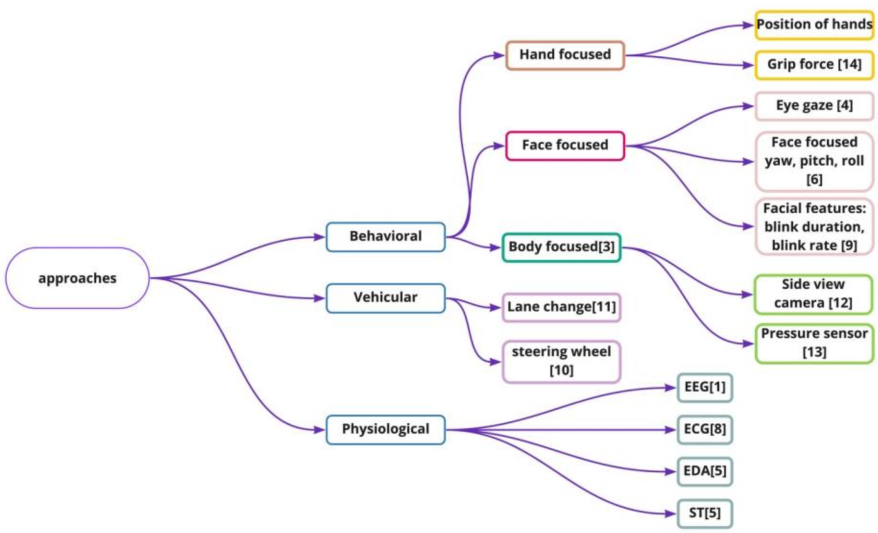

Figure 2: Drowsiness detection approaches covered in the literature

The collected dataset covers all the signals and approaches mentioned in table 1 and figure 1 to give the researchers the opportunity to investigate different combinations of the signals in drowsiness detection.

Our data is labeled and validated by the participants after completing the study and video is played for the users at the speed of 10x to aid the subjects remembering the events. The data labels are shown in the Table 2.

Rating	Description
1	Extremely alert
2	Very alert
3	Alert
4	Rather alert
5	Neither
6	Some signs of sleepiness
7	Sleepy but no effort to stay awake
8	Sleepy but some effort to stay awake
9	great effort to stay awake

Table 2: Drivers drowsiness level

The proposed model fuses video and biometric signals inside a Siamese network that is robust to noise, artifacts, and missing data. The proposed model fuses video and different signals separately that studies the effect of each modality on drowsiness detection without increasing the cost drastically. The model is personalized and detects the drowsiness from the early steps of the training. Another advantage of the models is that the computational cost is lower compared to deep learning models and attention networks.

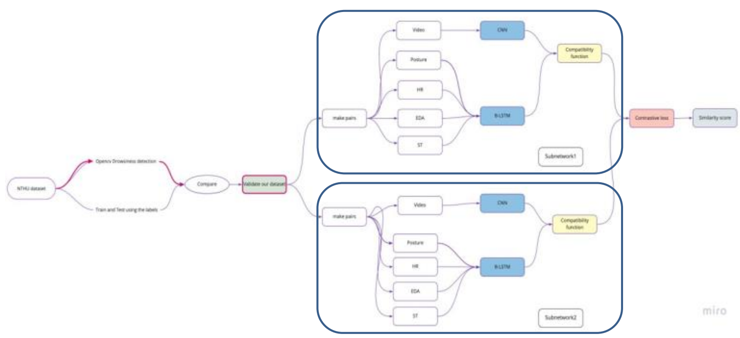

Figure 3: The Siamese network architecture

We also modified Iosifidos et al, [3] architecture to input the biometric signals and video and fuse the representations at intermediate level. Figure 3 shows two transformer modules to fuse the each modality by concatenating the representation of each modality regrading the other modality. In this project we concatenated the biometric signals into a two-dimensional tensor. A set of convolutional blocks had the responsibility of modality-specific feature learning. The representations of each modality concatenated at the penultimate layer.

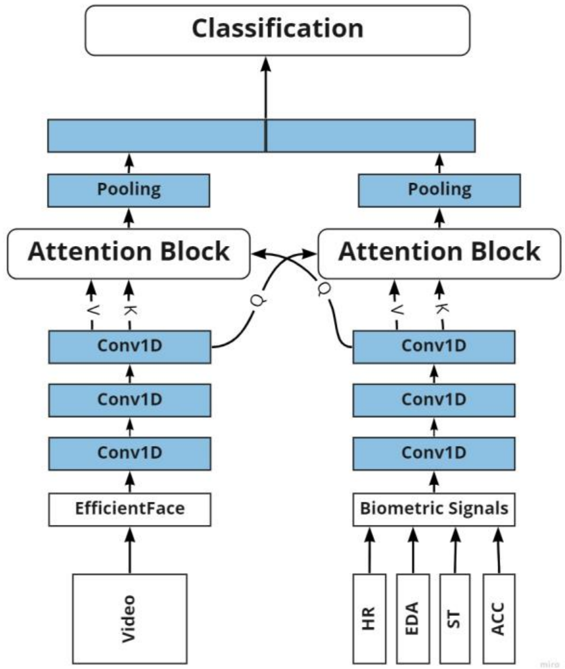

Figure 4: The transformer-based information fusion model

Results

This project is an ongoing 2-year project that we still collect data to reach the number of 40 drivers. The initial results show 80 percent precision on the drivers; however, we believe the model needs more training data and extending the number of the participants while maintaining the gender balance. We split the data labels into three different classes of Alert (extremely alert, very alert, a lert), Semi-alert(Alert, Rather alert, Neither), and drowsy(Sleepy but no effort to stay awake,Sleepy but some effort to stay awake, great effort to stay awake). We used the proposed architecture to detect the drowsiness of the subjects. The model can detect the drowsiness by 76.78 percent accuracy.

Moreover, we used a new transformer-based representation fusion model using the beforementioned labels the model consumes the video and biometric signals and fuses the biometric signals at early fusion steps. The video and biometric representations fuse at intermediate level. Figure shows the proposed architecture that is a variation of AI et al, [3] proposed architecture. The proposed architecture was studied using a different dataset that was a joint CVDI project between UL and Tampere university. Figure 2 shows the architecture

The results showed 85% precent precision. Table 3 compares the results of different architectures.

Model	Recall	Precision	F1_score
Drowsiness			
Siamese	69.35	85.04	75.93
Transformer fusion	62.35	70.81	59.57
Empathic School			
Siamese	73.40	77.74	76.65
Transformer fusion	85.08	89.98	87.87

Functionality of Innovation(s)

Existing work relies on video-based fatigue and drowsiness detection. We believe there is no literature that has integrated physiological data related to affective states for fatigue and drowsiness prediction. Our hypothesis is that we will be able to improve predict fatigue and stress. The project investigates the problem of predicting the drivers affective states (i.e. emotions related to anxiety, stress, drowsiness) using a combination of multiple sensors (i.e. video and physiological signals such as GSR, skin conductance, heart rate, and skin temperature)

We developed the hardware and software system for collecting multi-modal sensor data from driver (i.e. physiological wearables, video camera). This task includes feature engineering to integrate multi-modal sensor and video stream based features, deep learning techniques for improving the performance of drowsiness and fatigue detection.

Conclusions and Recommendations

Recently, there have been several efforts at investigating drowsiness. However, we believe drowsiness detection is in its infancy. Early drowsiness detection is still a challenge and there is not enough literature investigating early drowsiness detection using combination of biometric signals, behavioral patterns, and video. Moreover, there is no publicly available datasets for early driver drowsiness detection.

We designed and developed a driving simulator for multi-modal driver drowsiness detection and prediction system by fusing behavioral, biometric signals (i.e. posture, wrist wearable and pressure sensors), and video. The system includes equipment, hardware and software components that includes (driving simulator, multi-modal sensing modules, software for emotion and driver drowsiness detection). We also developed IRB application for conducting the study and evaluated the simulator.

We investigated the classification and prediction performance of two multimodal architectures using collected data. The results illustrate that the transformer-based model has show better performance while the training data is sufficient. In the further steps of this project we will subjoin the participants to remove the gender bias.

Impact and Uses/Benefits

List of References

[1] Olarte, Olivia. "Human error accounts for 90% of road accidents." AlertDriving Magazine (2011).

[2] Zhao, Lei, et al. "Driver drowsiness detection using facial dynamic fusion information and a DBN." IET Intelligent Transport Systems 12.2 (2018): 127-133.

[3] Chumachenko, Kateryna, Alexandros Iosifidis, and Moncef Gabbouj. "Self-attention fusion for audiovisual emotion recognition with incomplete data." arXiv preprint arXiv:2201.11095 (2022).

Appendix