Estimating Stress levels of Drivers using Empatica

Thesis to be submitted in partial fulfillment of the requirements for the degree

of

Dual Degree in Computer Science and Engineering

by

Aryaman Jain
18CS30007

Under the guidance of

Prof. Sandip Chakraborty

COMPUTER SCIENCE AND ENGINEERING,
INDIAN INSTITUTE OF TECHNOLOGY KHARAGPUR

Department of Computer science and Engineering,
Indian Institute of Technology,
Kharagpur
India - 721302

CERTIFICATE

This is to certify that we have examined the thesis entitled **Estimating Stress levels of Drivers using Empatica**, submitted by **Aryaman Jain**(*18CS30007*) a dual degree student of **Department of Computer science and Engineering,** in partial fulfillment for the award of degree of Dual Degree in Computer Science and Engineering. We hereby accord our approval of it as a study carried out and presented in a manner required for its acceptance in partial fulfillment for the Dual Degree for which it has been submitted. The thesis has fulfilled all the requirements as per the regulations of the Institute and has reached the standard needed for submission.

Prof. Sandip Chakraborty

Department of Computer science and Engineering,
Indian Institute of Technology,
Kharagpur

Place: Kharagpur
Date: 8th November 2022

ACKNOWLEDGEMENTS

While bringing out this report to its final form, I came across a number of people whose contributions in various ways helped my field of research and they deserve special thanks. It is a pleasure to convey my gratitude to all of them. First and foremost, I would like to express my deep sense of gratitude and indebtedness to my supervisor Prof. Sandip Chakraborty for his invaluable encouragement, suggestions and support from an early stage of this research and providing me extraordinary experiences throughout the work. Above all, his priceless and meticulous supervision at each and every phase of work inspired me in innumerable ways. There involvement with originality has triggered and nourished my intellectual maturity that will help me for a long time to come.

I specially acknowledge my PhD mentor Sugandh Pargal for their advice, supervision, and the vital contribution as and when required during this research. I am highly grateful that I had the opportunity to work such talented mentors and for their support and co-operation that is hard to express.

Aryaman Jain
IIT Kharagpur
Date: 6th November 2021

ABSTRACT

Monitoring driving status has great potential in helping us decline the occurrence probability of traffic accidents and the aim of this research is to develop a novel system for driving stress detection based on multimodal feature analysis and unsupervised classifiers. Physiological signals such as heart rate, electrodermal activity and acceleration were record from two drives executed in a prescribed route in controlled driving environments using Empatica E4 wristband. Features were widely extracted from time, spectral analysis of the features. In order to search for the optimal feature sets, PCA was used to avoid co-related features and top two principal components were chosen. To estimate the driving stress level we use K-Means algorithm with modified initialization of centroids. As the data was unsupervised we generated the ground truth values by manually annotating the video data of the subjects. After classifying we compared the classification with the ground truth obtaining *90.74%* accuracy for first subject and *83.33%* accuracy for second subject.

Contents

List of Figures

Abbreviations

IBI	Inter Beat Interval
BVP	Blood Volume Pulse
ACC	Accelerometer
TEMP	Temperature
HR	Heart Rate
ECG	ElectroCardioGram
ANN	Artificial Neural Networks
GSR	Galvanic Skin Response
SVM	Support Vector Machine
EEG	ElectroEncephaloGram
PCA	Principal Component Analysis

Chapter 1

Introduction

Stress, Fatigue and lack of attention while driving are found to be the leading cause of automobile accidents in the world [4, 8]. In the US 80% of all accidents and 65% of all near-crashes are caused by this reason. Finding the solutions to reduce accidents and improve safety has become an important issue for both government and automobile makers. An automated system that can provide the unfit status of a driver can surely help in reducing accidents.

Systems that measure driver's fitness based on videos, that track head movements, observe eyes and detect facial expressions have been widely explored [4, 7]. But they do not always return acceptable results, this can be due to poor light, driving at night or wearing facial accessories. These methods are also criticized for being very costly and difficult to uniform heuristics.

Methods that include driving behaviour including vehicle speed, accelerator, brake, clutch and gear changes (give references) are also shown to detect stress [14, 11]. These features can be obtained easily but they are mostly dependent on the vehicle type and driver's handling skills.

Features extracted from physiological signals such as heart rate, electrodermal activity (EDA), and electrocardiogram (ECG) had shown relatively high accuracy in getting the driver's state to citeBartra,chen2017detecting. But these methods require specific sensors and cables on the body to measure the features. An increase in the acceptance of wearables avoids the need for specific sensors and cables allowing drivers to drive without obstruction. The physiological signals measured from these wearables can be used to make an automatic system helping in reducing accidents.

In this paper we propose the following methodology and evaluate the results from it:

- Collect EDA, IBI and ACC data from drivers using Empatica E4 wristband.
- Use various time and spatial analysis to generate various features.
- Process the dataframe to remove any invalid or unwanted values.
- Use Dimension Reduction methods to remove co-related features.
- Estimate stress using K-Means unsupervised algorithm with modified initialization
- Generate ground truth by manually annotating video data. Compare ground truth with the estimated values for evaluation

Chapter 2

Related work

Road safety has been a significant concern amongst researchers for more than a decade. Multiple works have been done on monitoring driving behaviour to improve driving habits, such as several works that tried to use physiological sensors to calculate stress [6]. It records the driver's electrocardiogram, electromyogram, skin conductance and respiration. They used different sensors to measure each metric. It detected stress by using supervised data from video encoding of drivers. It used a Fisher projection matrix and a linear discriminant to reduce the dimensions of features. To classify the data between classes scatter was maximized and the within-class scatter was minimized. They achieved 97% on 5-min intervals.

[3] uses same data measure by [6]. It extracted features by using time-domain analysis, frequency-domain analysis and wavelet decomposition. Then applied sparse bayesian learning and principal component analysis (PCA) to further reduce dimensionality. For classification, it uses SVM and an extreme learning machine with 3 different kernels.

[12] proposed an algorithm based on an artificial neural network (ANN) to learn driving overall stress level and correlated it with the changes of statistical, structural and time-frequency features in galvanic skin response (GSR) and photoplethysmography(PPG) signals. Their system achieved a precision of 89.23% based on the data collected from over 19 subjects. [9] constructed a mobile healthcare system for automatic driving sleep onset detection using a support vector machine (SVM) by selecting the most descriptive features obtained from respiration and EEG. Their test results revealed that the integrated application of respiration and EEG resulted in 98.6% identification accuracy.

Chapter 3

Methodology

In this chapter, we will discuss the end-to-end framework. The proposed framework is shown in Fig. 3.1. Our dataset consists of sensor data from IBI, EDA and ACC sensors. Time-domain analysis as well as frequency-domain analysis is applied for feature generation. In our automated system, we detect the stress of the driver every 10s. The dataframe is then processed to filter out unwanted values. We further use dimension reduction methods to combine correlated features. Finally, it predicts the stress of the drivers based on a given dataset. We use the ground truth rating to validate the accuracy.

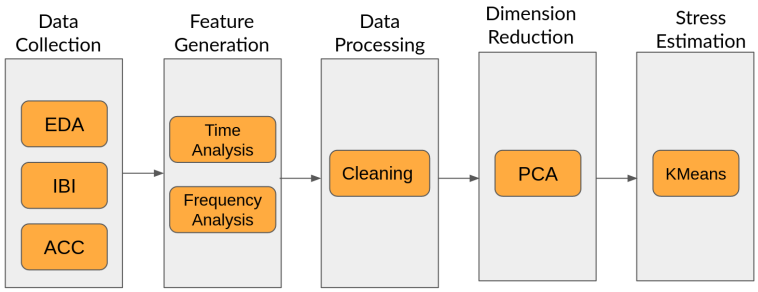

Figure 3.1: Framework for detecting driver's stress

3.1 Data Collection

The data is collected in a controlled driving experiment. There was a total of 2 driver subjects. They were asked to drive on the Indian Institute of Kharagpur roads. Roads were in ideal condition with no potholes. Traffic was minimal, only crowds of students occasionally caused traffic. The drives drove for about 8-9 mins. Drivers were allowed

to use a smartphone and talk to people whenever they wanted to. Both drivers were wearing Empatica E4 bands. Empatica E4 3.3 wristband is equipped with sensors designed to gather high-quality data. E4 is equipped with sensors that measure the following metrics.

Figure 3.2: Empatica E4 wristband

- **EDA** - Electrodermal activity (EDA) is the property of the human body that causes continuous variation in the electrical characteristics of the skin. Historically, EDA has also been known as skin conductance, galvanic skin response (GSR), electrodermal response (EDR), psychogalvanic reflex (PGR), skin conductance response (SCR), sympathetic skin response (SSR) and skin conductance level (SCL). The long history of research into the active and passive electrical properties of the skin by a variety of disciplines has resulted in an excess of names, now standardized to electrodermal activity (EDA). E4 measures EDA at a 4Hz sampling rate.

- **BVP** - BVP measurement is obtained by the use of a photoplethysmography (PPG) sensor. This component measures changes in blood volume in the arteries and capillaries that correspond to changes in the heart rate and blood flow. The PPG sensor detects changes by shining an infrared light, typically via a light-emitting diode (LED), onto the surface of the body. This light is transmitted through the tissues, then backscattered and reflected by the tissue before reaching the photodetector of the PPG sensor. The technology works because red light is selectively absorbed by the haemoglobin of the red blood cells and reflected by other tissues. The amount of light that returns to the

PPG photodetector is proportional to the relative volume of blood present in the tissue. E4 measures BVP at a 64Hz sampling rate.

- **ACC** - It consists of data from a 3-axis accelerometer sensor. The accelerometer is configured to measure acceleration in the range [-2g, 2g]. Therefore the unit in this file is 1/64g. Data from the x, y, and z axis are respectively in the first, second, and third columns.

- **TEMP** - TEMP consists of skin temperature in °C . E4 measures TEMP at 4 Hz.

- **IBI** - It measures the time between individuals heart beats extracted from the BVP signal. This metric does not require any sampling rate.

- **HR** - Measures the average heart rate extracted from BVP signal at 1 Hz sampling rate.

Empatica provides these metrics in the form of a .csv format. The first row is the initial time of the session expressed as a UNIX timestamp in UTC. The second row is the sample rate of different metrics expressed in Hz. The subsequent rows consist of the data of the corresponding metric.

For our framework, we use data from IBI, ACC and EDA metrics. We do not use HR and BVP metrics because they can be obtained using the IBI metric only. So including HR and BVP will include redundant metrics. Also, we are not including the TEMP metric because it takes more than 10s to react to stress [13].

3.2 Feature Generation

We are generating several features from IBI, ACC and EDA data. As we are doing prediction for 10s we divided data into segments and generated multiple features for each segment. Each segment has a window size of 10s and steps size of 10s as these segments are non-overlapping. We leverage the use of FLIRT [5]. FLIRT is a python package it is a feature generation toolkit for wearable data. From FLIRT we get the following IBI features 3.3, EDA features 3.4, and ACC features 3.5.

After feature generation, the data frame consists of 178 different features.

Category	Name	Description	Unit
Statistical	*min/max HR*	Minimum and maximum of the HR	bpm
	mean HR	Mean of the HR_i	bpm
	median HR	Median of the HR_i	bpm
Time	*SDNN*	SD of all NN intervals	ms
domain	*RMSSD*	The square root of the mean of the sum of the squares of differences between adjacent NN intervals	ms
	NN_{50}	Number of pairs of adjacent NN intervals differing by more than 50 ms in the entire recording	ms
	pNN_{50}	NN_{50} count divided by the total number of all NN intervals	%
	NN_{20}	Number of pairs of adjacent NN intervals differing by more than 20 ms in the entire recording	-
	pNN_{20}	NN_{20} count divided by the total number of all NN intervals.	%
	CVNN	Coefficient of variation equal to the ratio of SDNN divided by mean NN interval	-
	CVSD	Coefficient of variation of successive differences equal to the RMSSD divided by mean NN interval	-
	mean	Mean of the IBIs	ms
	std	Standard deviation of the IBIs	ms
	min/max	Minimum and maximum of the IBIs	ms
	pt p	Range (peak to peak) of the IBIs	ms
	sum	Sum of the IBIs	ms
	energy	Energy of the IBIs	ms^2
	skewness	Skewness of the IBIs	-
	kurtosis	Kurtosis of the IBIs	-
	peaks	Number of the IBIs	-
	rms	Root mean square of the IBIs	ms
	line_integral	Integral under the IBIs	ms
	n_above_mean	Number of IBIs above the mean	-
	n_below_mean	Number of IBIs below the mean	-
	n_sign_changes	Number of changes in the IBIs slope	-
	iqr	Interquartile range between the 25th and 75th percentile of the IBIs	ms
	iqr $5-95$	Interquartile range between the 5th and 95th percentile of the IBIs	ms
	pct 5	5th percentile of the IBIs	-
	pct 95	95th percentile of the IBIs	-
	entropy	Entropy of the IBIs	-
	perm entropy	Permutation entropy of the IBIs	-
	svd entropy	Singular value decomposition of the IBIs entropy	-
Frequency	*total power*	The variance of NN intervals over the temporal segment below 0.04 Hz	ms^2
domain	*vlf*	Power in very low frequency range below or equal 0.04 Hz	ms^2
	lf	Power in low frequency range 0.04 Hz and 0.15 Hz	ms^2
	hf	Power in high frequency range 0.15 Hz and 0.4 Hz	ms^2
	lf/hf – ratio	Ratio of LF to HF	-
	lfnu	LF power in normalized units	-
	hfnu	HF power in normalized units	-

High frequency (HF), low frequency (LF), heart rate (HR), inter-beat interval (IBI), and normal-to-normal (NN) interval.

Figure 3.3: HR and HRV features

3.3 Data Processing

Dataframe obtained after feature generation contains *null* and *-inf* values. *null* values were present because of the late initialization of IBI sensors which cause *null* values in all features containing IBI data. We remove the initial 4 rows for each driver because of this reason.

The features 'acc_x_entropy', 'acc_y_entropy', 'acc_z_entropy', 'eda_phasic_entropy', 'eda_tonic_entropy' were removed because of the calculation of entropy where *-inf* is a valid domain. Most of the rows were having *-inf* values so, we decided to drop the features instead of removing the rows.

Processing the data frame, consists of 173 features.

Category	Name	Description	Unit
Time domain	*mean*	Mean of the SCR and SCL	μS
	std	Standard deviation of the SCR and SCL	μS
	min/max	Minimum and maximum of the SCR and SCL	μS
	ptp	Range (peak to peak) of SCR and SCL within a time interval	μS
	sum	Sum of the SCR and SCL values with a time interval	μS
	energy	Energy of the SCR and SCL	μS^2
	skewness	Skewness of the SCR and SCL	-
	kurtosis	Kurtosis of the SCR and SCL	-
	peaks	Number of SCR and SCL peaks with a time interval	-
	rms	Root mean square of the SCR and SCL	μS
	line_integral	Integral under the SCR and SCL curve	$\mu S.s$
	n_above_mean	Number of SCR and SCL data-points above the mean	-
	n_below_mean	Number of SCR and SCL data-points below the mean	-
	n_sign_changes	Number of changes in the SCR and SCL slope	-
	iqr	Interquartile range between the 25th and 75th percentile of the SCR and SCL	μS
	iqr_5-95	Interquartile range between the 5th and 95th percentile of the SCR and SCL	μS
	pct_5	5th percentile of the SCR and SCL	-
	pct_95	95th percentile of the SCR and SCL	-
	entropy	Entropy of the SCR and SCL	-
	perm_entropy	Permutation entropy of the SCR and SCL	-
	svd_entropy	Singular value decomposition of the SCR and SCL entropy	-
Frequency domain	*sma*	Signal magnitude area of the frequency domain SCR and SCL	μS
	energy	Energy of the frequency domain SCR and SCL	μS^2
	kurtosis	Kurtosis of the frequency domain SCR and SCL	-
	iqr	Interquartile range of the frequency domain SCR and SCL	$\mu S/Hz$
	spectral_power	5 spectral power magnitudes in the [0.05-0.55] Hz bands for the power density of the SCR and SCL	$\mu S^2.Hz$
	var_power	Variance of the SCR and SCL spectral power	μS^2
Time-frequency domain	*mean*	Mean of the SCR's and SCL's MFC signal	μS
	std	Mean of the SCR's and SCL's MFC signal	μS
	median	Median of the SCR's and SCL's MFC signal	μS
	iqr	Interquartile range of the SCR's and SCL's MFC signal	μS
	skewness	Skewness of the SCR's and SCL's MFC signal	-
	kurtosis	Kurtosis of the SCR's and SCL's MFC signal	-
SCR time-domain features	*peaks*	Number of SCR peaks	-
	rise_time	Mean of the SCR peaks rise time	s
	max_deriv	Mean value of the maximum derivative of the SCR peaks	$\mu S/s$
	amp	Mean amplitude of the SCR peaks	μS
	decay_time	Mean of the SCR peaks decay time	s
	scr_width	Mean width of the SCR peaks	s
	auc_mean	Mean area-under-curves of the SCR peaks	$\mu S.s$
	auc_sum	Sum of the area-under-curves of the SCR peak	$\mu S.s$

Mel-Frequency Cepstrum (MFC), skin conductance level (SCL), and skin conductance response (SCR).

Figure 3.4: EDA features

3.4 Dimension Reduction

Dimensionality reduction, or dimension reduction, is the transformation of data from a high-dimensional space into a low-dimensional space so that the low-dimensional representation retains some meaningful properties of the original data, ideally close to its intrinsic dimension. Working in high-dimensional spaces can be undesirable for many reasons; raw data are often sparse as a consequence of the curse of dimensionality, and analyzing the data is usually computationally intractable (hard to control or deal with).

Our data frame consists of 173 features. Most of the features are correlated and hence they can be integrated into 1 feature. We use PCA [1] to reduce the dimension of the data frame.

Category	Name	Description	Unit
Time domain	*mean*	Mean of the ACC signal	g
	std	Standard deviation of the ACC signal	g
	min/max	Minimum and maximum of the ACC signal	g
	ptp	Range (peak to peak) of the ACC signal	g
	sum	Sum of the ACC signal	g
	energy	Energy of the ACC signal	g^2
	skewness	Skewness of the ACC signal	-
	kurtosis	Kurtosis of the ACC signal	-
	peaks	Number of the ACC signal	-
	rms	Root mean square of the ACC signal	g
	line_integral	Integral under the ACC signal	g
	n_above_mean	Number of ACC signal above the mean	-
	n_below_mean	Number of ACC signal below the mean	-
	n_sign_changes	Number of changes in the ACC signal slope	-
	iqr	Interquartile range between the 25th and 75th percentile of the ACC signal	g
	iqr_5 – 95	Interquantile range between the 5th and 95th percentile of the ACC signal	g
	pct_5	5th percentile of the ACC signal	-
	pct_95	95th percentile of the ACC signal	-
	entropy	Entropy of the ACC signal	-
	perm_entropy	Permutation entropy of the ACC signal	-
	svd_entropy	Singular value decomposition of the ACC signal entropy	-
Frequency domain	*sma*	Signal magnitude area of the frequency domain ACC signal	g
	energy	Energy of the frequency domain ACC signal	g^2
	kurtosis	Kurtosis of the frequency domain ACC signal	-
	iqr	Interquartile range of the frequency domain ACC signal	g/Hz
	spectral_power	3 spectral power magnitudes in the [1–5.5] Hz bands for the power density of the ACC signal	$g^2.Hz$
	var_power	Variance of the ACC signal spectral power	g^2
Time-frequency domain	*mean*	Mean of the ACC signal's MFC signal	g
	std	Mean of the ACC signal's MFC signal	g
	median	Median of the ACC signal's MFC signal	g
	iqr	Interquartile range of the ACC signal's MFC signal	g
	skewness	Skewness of the ACC signal's MFC signal	-
	kurtosis	Kurtosis of the ACC signal's MFC signal	-

Accelerometer (ACC) and gravitational force equivalent (g).

Figure 3.5: ACC features

3.4.1 Outcome of PCA

The main linear technique for dimensionality reduction, principal component analysis, performs a linear mapping of the data to a lower-dimensional space in such a way that the variance of the data in the low-dimensional representation is maximized. In practice, the covariance (and sometimes the correlation) matrix of the data is constructed and the eigenvectors on this matrix are computed. The eigenvectors that correspond to the largest eigenvalues (the principal components) can now be used to reconstruct a large fraction of the variance of the original data. Moreover, the first few eigenvectors can often be interpreted in terms of the large-scale physical behaviour of the system, because they often contribute the vast majority of the system's energy, especially in low-dimensional systems. Still, this must be proven on a case-by-case basis as not all systems exhibit this behaviour. The original space (with the dimension of the number of points) has been reduced (with data loss, but hopefully retaining

the most important variance) to the space spanned by a few eigenvectors.

After applying PCA on the feature-generated dataset, we observed that the topmost 2 principal component features correspond to about 86.52% variance and the topmost 3 features correspond to about 97.1% for driver1. For driver2 topmost 2 principal component features corresponds to 96.28% and the topmost 3 correspond to 99.8%. We decided to take the topmost 2 features for stress estimation ?? ??.

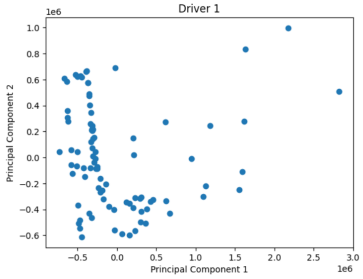

Figure 3.6: Plot of Driver 1 after PCA

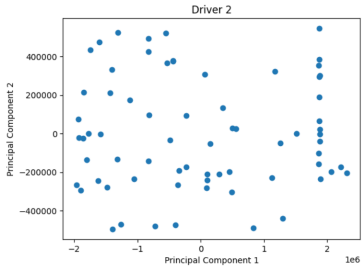

Figure 3.7: Plot of Driver 2 after PCA

3.5 Stress Estimation

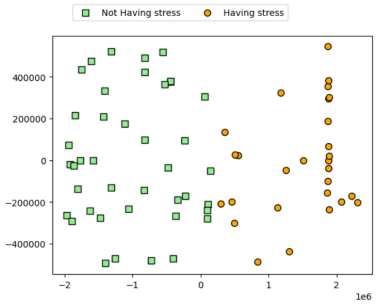

Figure 3.8: Plot of Driver 1

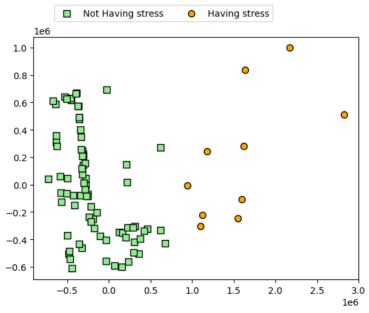

Figure 3.9: Plot of Driver 2

We will use K-means[10] to classify our data into 2 classes. Classes representing the points that belong to *having stress* and *not having stress*. One disadvantage of the K-means algorithm is that it is sensitive to the initialization of the centroids or the mean points. So, if a centroid is initialized to be a "far-off" point, it might just end up with no points associated with it, and at the same time, more than one cluster might end up linked with a single centroid. Similarly, more than one centroid might be initialized into the same cluster resulting in poor clustering. To overcome this

drawback we use K-means++[2]. This algorithm ensures a smarter initialization of the centroids and improves the quality of the clustering. Apart from initialization, the rest of the algorithm is the same as the standard K-means algorithm. That is K-means++ is the standard K-means algorithm coupled with a smarter initialization of the centroids. K-means++ is proved to be O(logk) competitive to the optimal K-means. We ran the K-means algorithm with K-means++ initialization for 300 iterations 3.8 3.9. The class that has a lower x coordinate of its centroid is labelled as *not having stress* and the other class as *having stress*.

Chapter 4

Evaluation

For evaluation, we manually annotated the driver's video data. Annotations were done using intervals of size 10s. We manually divide video data into intervals of 10s into two classes. First represents, the driver not having stress and the other has the driver having stress 4.1 4.2.

Figure 4.1: Driver not in Stress

Figure 4.2: Driver in Stress

Then we compared annotated data with the stress estimation from K-Means to get the evaluation metrics as shown in table 4.1. We use the following metrics to evaluate our estimation.

- Precision - Precision measures the percentage of time instances flagged as *having stress* that were correctly classified
- Recall - Recall measures the percentage of actual *having stress* time intervals that were correctly classified
- Accuracy - Accuracy measures the percentage of correct predictions of both *having stress* and *not having stress* classes

Driver	precision	recall	accuracy
1	66.66	57.14	90.74
2	66.66	83.33	83.33

Table 4.1: Evaluation metrics

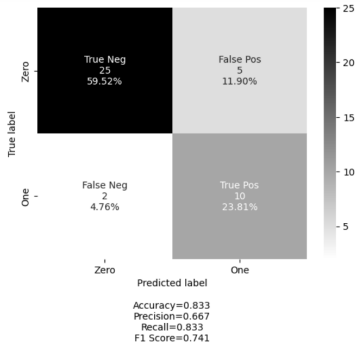

Figure 4.3: Confusion Matrix of Driver 1

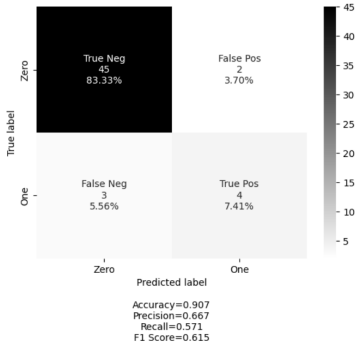

Figure 4.4: Confusion Matrix of Driver 2

Chapter 5

Future Work

- This research only analyzes the physiological measures for the identification of drive-related stress. As multiple features are favourable for detecting the driver's status more reliably and robustly, a hybrid detection system could be developed considering vehicle behaviour, physiological measures and image-based indicators.
- More methods for feature generation and dimension reduction can be explored. Features which are proven to not react to stress or take more time to react to stress can be dropped.
- Testing has to be done extensively. More subjects are needed for testing. As well as our system being tested in a controlled environment, testing should also be done in real-world driving conditions. We need an alternative way to generate ground truth as manual annotations are time taking and introduce bias.

Chapter 6

Conclusion

Monitoring drivers' internal status has great potential in declining the occurrence probability of traffic accidents. The present research developed an automatic system for detecting the driving-related stress level based on multichannel physiological records. A variety of features were extracted using time and spectral analysis. Principal Component Analysis (PCA) was used to search for the optimal and compact feature sets. K-Means with modified K-Means++ initialization was used to classify data into 2 classes. We evaluated the proposed work on the ground truth generated by manually annotating the video data of drivers. We were able to get the accuracy of 90.74% for driver 1 and 83.33% for driver 2.

Bibliography

[1] Hervé Abdi and Lynne J Williams. Principal component analysis. *Wiley interdisciplinary reviews: computational statistics*, 2(4):433–459, 2010.

[2] David Arthur and Sergei Vassilvitskii. k-means++: The advantages of careful seeding. Technical report, Stanford, 2006.

[3] Lan-lan Chen, Yu Zhao, Peng-fei Ye, Jian Zhang, and Jun-zhong Zou. Detecting driving stress in physiological signals based on multimodal feature analysis and kernel classifiers. *Expert Systems with Applications*, 85:279–291, 2017.

[4] Bogusław Cyganek and Sławomir Gruszczyński. Hybrid computer vision system for drivers' eye recognition and fatigue monitoring. *Neurocomputing*, 126:78–94, 2014. Recent trends in Intelligent Data Analysis Online Data Processing.

[5] Simon Föll, Martin Maritsch, Federica Spinola, Varun Mishra, Filipe Barata, Tobias Kowatsch, Elgar Fleisch, and Felix Wortmann. Flirt: A feature generation toolkit for wearable data. *Computer Methods and Programs in Biomedicine*, 212:106461, 2021.

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

[7] Hedyeh A. Kholerdi, Nima TaheriNejad, Reza Ghaderi, and Yaser Baleghi. Driver's drowsiness detection using an enhanced image processing technique inspired by the human visual system. *Connection Science*, 28(1):27–46, 2016.

[8] Boon-Giin Lee and Wan-Young Chung. Driver alertness monitoring using fusion of facial features and bio-signals. *IEEE Sensors journal*, 12(7):2416–2422, 2012.

[9] Boon Giin Lee, Boon-Leng Lee, and Wan-Young Chung. Mobile healthcare for automatic driving sleep-onset detection using wavelet-based eeg and respiration signals. *Sensors (Basel, Switzerland)*, 14:17915–17936, 10 2014.

[10] Aristidis Likas, Nikos Vlassis, and Jakob J Verbeek. The global k-means clustering algorithm. *Pattern recognition*, 36(2):451–461, 2003.

[11] Drew M. Morris, June J. Pilcher, and Fred S. Switzer III. Lane heading difference: An innovative model for drowsy driving detection using retrospective analysis around curves. *Accident Analysis & Prevention*, 80:117–124, 2015.

[12] Rajiv Ranjan Singh, Sailesh Conjeti, and Rahul Banerjee. A comparative evaluation of neural network classifiers for stress level analysis of automotive drivers using physiological signals. *Biomedical Signal Processing and Control*, 8(6):740–754, 2013.

[13] Christiaan H. Vinkers, Renske Penning, Juliane Hellhammer, Joris C. Verster, John H. G. M. Klaessens, Berend Olivier, and Cor J. Kalkman. The effect of stress on core and peripheral body temperature in humans. *Stress*, 16(5):520–530, 2013. PMID: 23790072.

[14] Xuesong Wang and Chuan Xu. Driver drowsiness detection based on non-intrusive metrics considering individual specifics. *Accident Analysis & Prevention*, 95:350–357, 2016.