A Novel Method of Stress Detection using Physiological Measurements of Automobile Drivers

A Dissertation

Submitted in partial fulfillment of the requirements for the award of degree of

Master of Engineering

in

Electronic Instrumentation and Control Engineering

Submitted By:

Abdullah Bin Queyam

(ME-EIC, Reg. No. 801151001)

Under the esteemed guidance of:

Dr. Mandeep Singh

Associate Professor, EIED

Thapar University, Patiala

ELECTRICAL AND INSTRUMENTATION ENGINEERING DEPARTMENT

THAPAR UNIVERSITY

(Established under the section 3 of UGC act, 1956)

PATIALA, 147004, Punjab, India

JULY 2013

CERTIFICATE

I hereby declare that the Dissertation entitled **"A Novel Method of Stress Detection using Physiological Measurements of Automobile Drivers"** is an authentic record of my own work carried out as the requirements for the award of the degree of M.E. (Electronic Instrumentation and Control Engineering) at Thapar University, Patiala, under the esteemed guidance of **Dr. Mandeep Singh**, Associate Professor, EIED, Thapar University, Patiala.

The matter presented in this Dissertation has not been submitted for the award of any other degree of this or any other university.

Abdullah

Abdullah Bin Queyam
(Roll no – 801151001)

Date 10 July 2013

It is certified that the above statement made by the student is correct to the best of my knowledge and belief.

Dr. Mandeep Singh
Associate Professor, EIED
Thapar University, Patiala

Countersigned by:

S. Ghosh

Dr. Smarajit Ghosh
Professor & Head, EIED
Thapar University, Patiala

Dr. S. K. Mohapatra
Dean, Academic Affairs
Thapar University, Patiala

ABSTRACT

Driving a car is a complex cognitive process in which even a small lack of attention can have disastrous consequences. In order to minimize human error while driving, we can monitor stress and fatigue by measuring physiological parameters like ElectroCardioGram (ECG), ElectroMyoGram (EMG), Skin Conductance (SC) also called as Galvanic Skin Response (GSR) and Respiration Rate (RR) continuously over a period of time. Autonomic Nervous System (ANS) primarily depends on emotional responses of the human body to the dynamic surrounding. Further it also controls the smooth muscles, heart muscle and secretion of the glands in human body. As a result of this fact, bio-signal recordings reflecting the operating condition of the physiological systems including the circulatory, respiratory, muscular and endocrine systems can provide useful information representing the dynamics of the internal states in human body. Hence, the dynamic mental stress level of an automobile driver can be derived from those recordings.

In this research we accessed raw physiological signals available at PHYSIONET website and then extracted useful statistical features. Correlation analysis on the selected features showed that Mean HR and Mean Hand GSR are the two statistical features that have a very strong correlation with changing traffic conditions. We presented a method based on a correlation analysis and developed a mathematical function for the estimation of automobile driver stress level. The proposed methodology monitors driver's stress level using features extracted from selected physiological parameters. The results obtained indicate a strong correlation between the stress level of driver and the mathematical function formed. We used threshold approach to perform classification of affective states as "Low Stress", "Moderate Stress" and "High Stress" based on different traffic conditions. The results indicate classification accuracy of more than 80% in most of the driver data sets. Thus, the stress function acts as a direct indicator of stress level of the automobile diver whose physiological parameters are monitored continuously in real-time.

ACKNOWLEDGMENT

I would like to express my sincere gratitude to my supervisor, **Dr. Mandeep Singh,** Associate Professor, EIED, Thapar University, Patiala for all his guidance and invaluable advises throughout the progress. He encouraged and stimulated my interest in stress recognition of automobile drivers and inspired me for doing research on this topic.

I would also like to thank **Dr. Smarajit Ghosh**, Professor and Head, Electrical and Instrumentation Engineering Department, and **Dr. S. K. Mohapatra**, Dean, Academic Affairs, Thapar University, Patiala for giving an opportunity to work in this regard.

Also I would like to thank my parents and my friends who were always there for me. It would not have been possible without their continuous support and encouragement.

(Abdullah Bin Queyam)

CONTENTS

LIST OF ABBREVIATIONS

ANN	Artificial Neural Network
ANOVA	Analysis of Variance
ANS	Autonomic Nervous System
BP	Blood Pressure
BVP	Blood Volume Pressure
C&RT	Classification and Regression Trees
CHAID	Chi Square Automatic Interaction Detector
CPM	Contractions per Minutes
CSV	Comma-Separated Values
CWT	Color Word Test
ECG	Electrocardiogram
EDA	Electro Dermal Activity
EEG	Electroencephalogram
EMG	Electromyogram
FPGA	Field-Programmable Gate Array
FSC	Foot Based Skin Conductance
GMM	Gaussian Mixture Models
GSR	Galvanic Skin Response
HR	Heart Rate
HRV	Heart Rate Variability
HRvB	Heart Rate Variation from Baseline
HS	High Stress
HSC	Hand Based Skin Conductance
IBI	Inter Beat Interval
IHR	Instantaneous Heart Rate
IRR	Instantaneous Respiratory Rate

KNN	K-Nearest Neighbor
KSOM	Kohonen's Self-Organizing Map
LDA	Linear Discriminant Analysis
LS	Low Stress
MANOVA	Multivariate analysis of variance
MFAD	Mean of first absolute differences
ML	Machine Learning
MLP	Multilayer Perceptron
MS	Moderate Stress
NHRV	Normalised HRV
NN	Neural Network
PCA	Principal Component Analysis
PD	Pupil Diameter
PO	Pulse Oximetry
PPG	Photoplethysmogram
RESP	Respiration
ROC	Receiver Operating Characteristics
RR	Respiration Rate
RSA	Respiratory Sinus Arrhythmia
SC	Skin Conductance
SNS	Sympathetic Nervous System
SNS	Sympathetic Nervous System
SOM	Self-Organizing Map
ST	Skin Temperature
SVM	Support Vector Machine
WEKA	Waikato Environment for Knowledge Analysis

LIST OF FIGURES

LIST OF TABLES

Chapter 1

INTRODUCTION

1.1 Overview

Driving is a complex task, requiring full concentration and a vital balance between alertness and a calm attitude. Stress and strong emotions can affect this balance whether they result from the driving task itself, or some unrelated matters. The job of operating public transit vehicles in urban centres may be among the most stressful and unhealthy modern occupations.

Many studies conducted over the last four decades in cities on almost every continent show that bus drivers, when compared to workers in other jobs, are more likely to experience, death from heart and blood vessel disease. Further they suffer from conditions such as chest pain and high blood pressure (BP), digestive disorders, musculoskeletal problems, especially on the back, neck and shoulders. Bus drivers frequently report mental overload fatigue and sleeping disorders. Bus drivers also have more frequent absences from work and of longer duration than workers in other occupations.

Driving is one of the most stress causing activities. And stress is mainly caused by change in environment. Whenever we are behind the wheel we are exposed to unbelievably fast changing circumstances like road conditions, vehicular traffic, speed limits, traffic lights, on road obstacles etc. In addition we are supposed to take into consideration all of them in order to avoid penalties/accidents. In the worst case scenario any negligence may have fatal consequences. Driving also requires several decision-making operations involving complex information processing. These operations often lead to high mental stress. A stressed out driver gets easily irritated and may have what's commonly known as "Road Rage". In fit of this rage, a driver can cause lot of harm, not only to himself or his passengers, but also to the fellow road travellers. It is therefore very important to measure stress level in automobile drivers.

In efforts to improve overall safety and comfort, a number of vehicular technologies have been developed and deployed in the market over the last few decades. A major drawback of available vehicular systems, however, is that they do not include the driver in the loop of decision-making processes. For example, even if the driver is heavily cognitively loaded or distracted, the decision threshold of safety systems as well as the human-computer interface's information exchange protocol remains the same. Little research is done in designing the human-computer interface of a high end that also includes stress monitoring of the automobile driver.

Stress is a physiological condition in which the body becomes excited to face an emergency situation. A number of physiological changes including increase in heart rate, breathing rate, pupil dilation, muscle contraction, sweating etc. occurs during a high stress state. When the brain perceives a stressful situation like high traffic driving, it send messages to a section of nerves called the Autonomic Nervous System (ANS) which then activates the adrenal glands in the kidneys to secrete hormones, such as adrenaline and noradrenaline. This adrenaline hormone secreted by the adrenal glands initiates the process of "Fight or Flight syndrome" that leads to a number of physiological changes in the body. These hormone triggered changes makes the body alert to face the situation. This is mainly due to the activation of Sympathetic Nervous System (SNS). After a short period of time, body returns to normal state by the activation of the Parasympathetic nervous system. Thus, it may be concluded that stress is a natural mechanism to cope up with the changes in the surroundings. These physiological changes in the body can be recorded and analysed for precise detection of stress in an automobile drivers.

A survey by Brake and Green Flag found that 12% of drivers had driven while feeling stressed, angry or annoyed at the behaviour of other road users several times a week. Further 15% reported this feeling once a week, 16% once a month and 4% at least once a day [1]. There are many stressful situations associated with driving such as traffic jams, tailgating and generally dealing with other driver's risk-taking. All drivers are exposed to stressful driving situations from time to time, even if they do not generally suffer from stress in everyday life. Research has shown that angry drivers are more likely to take risks such as speeding, rapidly switching lanes, tailgating and jumping red lights. It may be pertinent to add here that tailgating means; to drive so closely behind another vehicle that one cannot stop or swerve with ease in an emergency.

Though health check-ups are done periodically for bus drivers, there are not many facilities for stress reduction. If a person drives while stressed they run a much greater risk of being involved in a crash that kills or injures them or another road user. At this point it is important to find positive and productive ways to deal with stress. One of the simplest and most effective ways to deal with stress is its continuous monitoring. Stress level of an Automobile driver can be monitored from many different domains as shown in Fig. 1.

Timely detection of the stress in human beings provides a helpful way for people to better understand their stress condition and provides physicians with more reliable data for intervention and stress management.

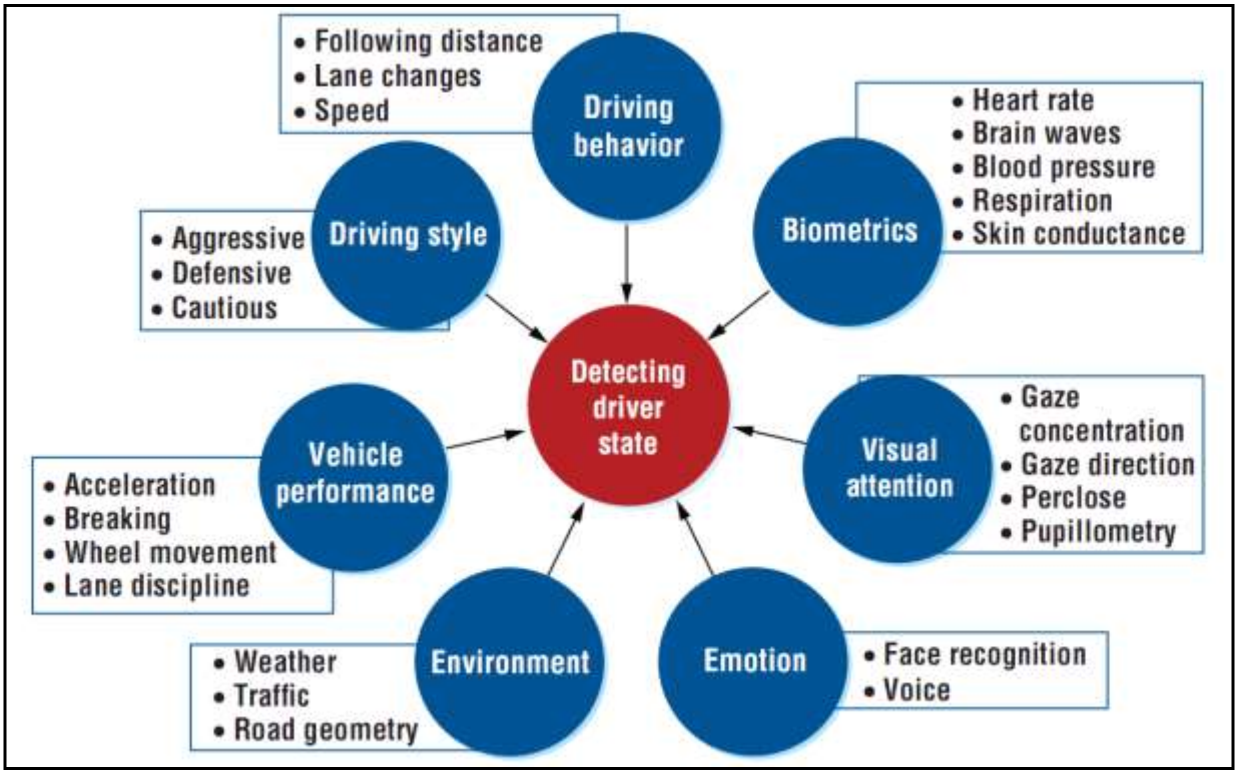

Fig. 1. Domains from which inputs for an integrated driver stress detection system might be drawn.

Existing studies have shown that stress can be recognized by the physiological parameters of the subject. The physiological parameters like ElectroCardioGram (ECG), Galvanic Skin Response (GSR), ElectroMyoGram (EMG), and Respiration Rate (RR) can be acquired for continuous monitoring of stress level. One of the most relevant studies in this field was conducted by Healey and Picard [2], in which they presented methods for collecting and analyzing physiological data during real world driving tasks. They continuously recorded ECG, EMG, Skin Conductance (SC) also known as GSR, and Respiration (RESP) signals of drivers. They showed that physiological measurements can predict mental stress with high accuracy.

Healey and Picard's work, however, lacked the procedure of feature selection, which may result in higher accuracy performance but with the higher computational burden and user resistance especially in real time stress recognition. Also Healey and Picard's approach which was based on judgement of human coders on videotapes of the driving. Their approach might be biased toward visually discernible effects, and may not well reflect the true stress state of the drivers.

The information's gathered from Autonomic Nervous Systems (ANS) are considered useful in detecting, analyzing, and modelling psycho-physiological states of the human body. ANS primarily involve emotional responses and controls the smooth muscles, heart muscle and secretion of the glands in the human body. Bio-signal recordings reflect the operating condition of the physiological systems of human-being like the circulatory, respiratory, muscular and endocrine systems can provide useful metrics representing the dynamics of the internal states in the human body.

The database created by Healey and Picard mentions the drive time through highway and through city conditions, along with the period during which the driver is resting [2]. One simple approach in classification of stress level is to consider the rest period as low stress condition, highway driving as moderate stress condition and city drive as high stress condition. Neglecting all other factor towards the contribution of stress, an analysis of correlation of various physiological parameters with the three mentioned states of mental stress has been made in this paper.

From a driving process, task related data can be derived from physiological recordings that can be collected continuously without interfering with the driver's performance. Information obtained in this way can provide a continuous measure to determine how different road and traffic conditions affect the drivers.

1.2 Biometric Domain

In our work we have identified a number of physiological markers of stress like an electrocardiogram (ECG), galvanic skin response (GSR), electromyogram (EMG), and respiration (RESP) that can be acquired for continuous monitoring of stress level of an automobile driver. The physiological signals that we used in our research will each be described briefly:

1.2.1 Galvanic skin response

GSR is a measure of the conductivity of the skin. There are specific sweat glands (eccrine glands) that cause this conductivity to change and result in the GSR. Located in the palms of the hands and soles of the feet, these sweat glands respond to psychological stimulation rather than simply to temperature changes in the body [3]. Some of the features that can be extracted from a GSR signal segment corresponding to busy driving scenario are shown in Fig. 2. Galvanic skin response is a linear correlate to arousal and reflects both emotional responses as well as cognitive activity. The GSR signal comprises of two components: tonic (Frequency Range: 0.0Hz to 0.16 Hz) and phasic (Frequency Range: 0.16Hz and above).

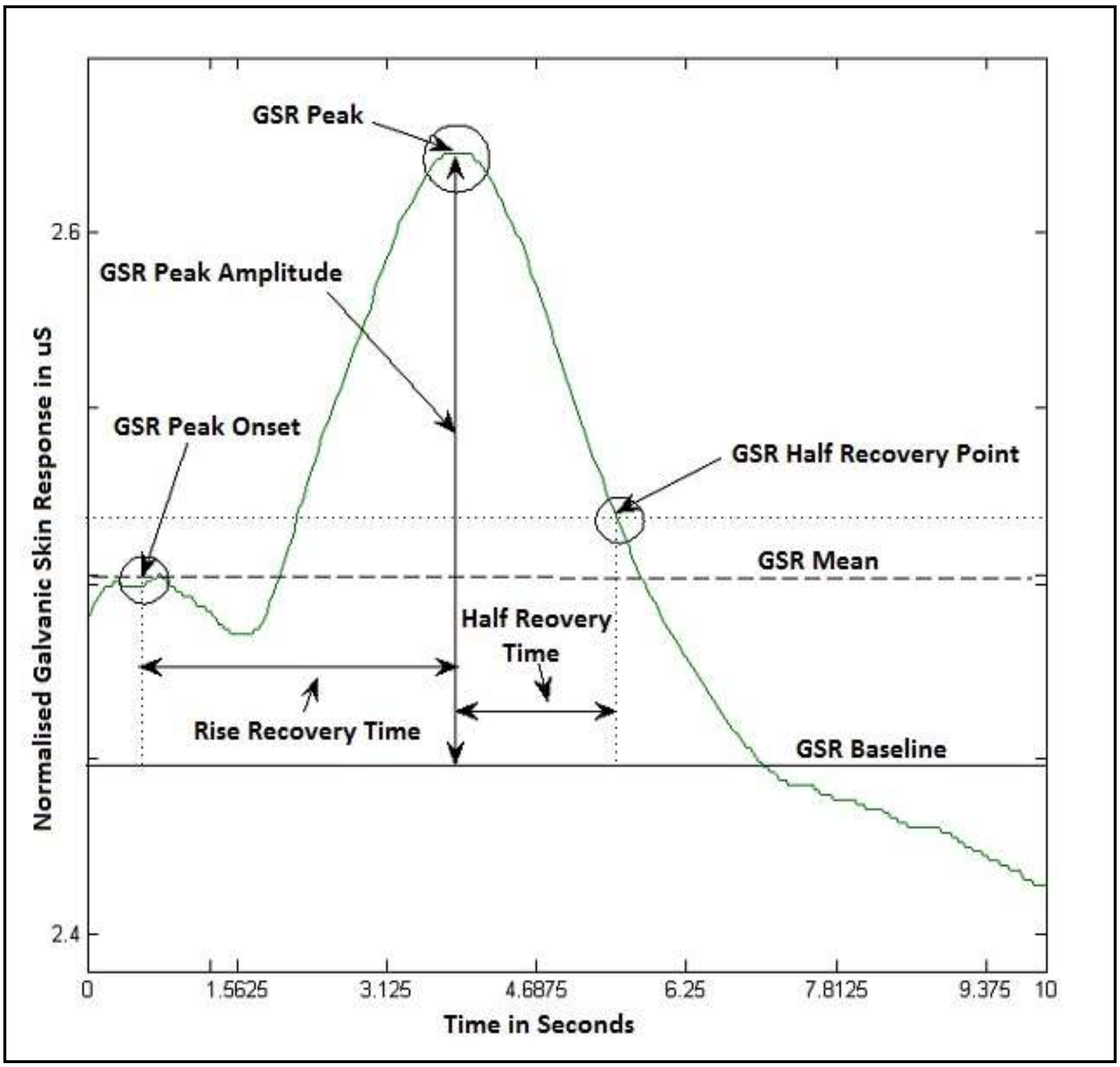

Fig. 2. Schematic GSR to a hypothetical stimulus.

For example, many people have cold clammy hands when they are nervous. In fact, subjects do not have to even be sweating to see differences in skin conductance in the palms of the hands or soles of the feet because the eccrine sweat glands act as variable resistors on the surface. As sweat rises in a particular gland, the resistance of that gland decreases even though the sweat may not reach the surface of the skin [4].

1.2.2 Cardiovascular measures

The cardiovascular system includes the organs that regulate blood flow through the body. Measures of cardiovascular activity include Heart Rate (HR), inter beat interval (IBI), heart rate variability (HRV), blood pressure (BP), and BVP. ECG measures electrical activity of the heart. HR, HRV, and respiratory sinus arrhythmia (RSA) can all be gathered from ECG. HR reflects emotional and stress related activity. It has been used to differentiate between positive and negative emotions. HRV refers to the oscillation of the interval between consecutive heartbeats. When subjects are under stress, HRV is suppressed and when they are relaxed, HRV emerges as shown in Fig. 3. Similarly, HRV decreases with mental effort, but if the mental effort needed for a task increases beyond the capacity of working memory, HRV will increase [5].

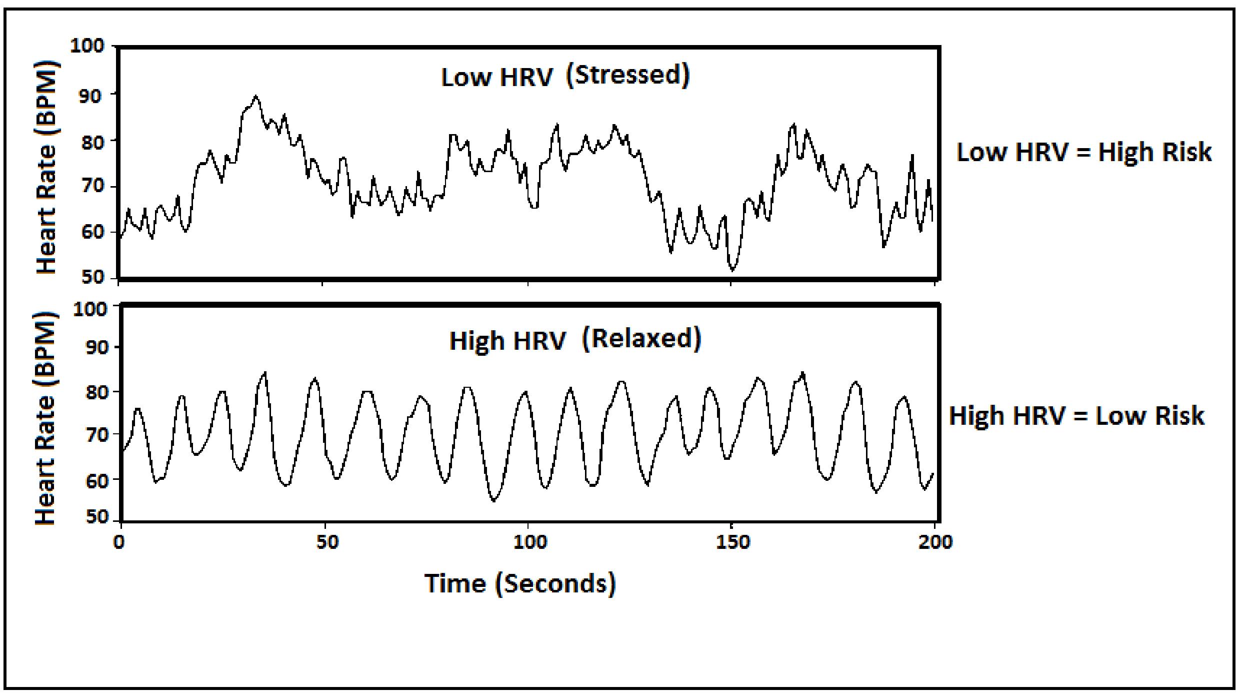

Fig. 3. Shows an original HR signal notice the different arousal of this signal, depending on the stressing stimulus.

1.2.3 Respiratory measures

Respiration can be measured as the rate or volume at which an individual exchanges air in their lungs. Rate of respiration (RespRate) and depth of breath (RespAmp) are the most common measures of respiration. Emotional arousal increases respiration rate while rest and relaxation decreases respiration rate [4]. Although respiration rate generally decreases with relaxation, startle events and tense situations may result in momentary respiration cessation. Negative emotions cause irregularity in the respiration pattern. Because respiration is closely linked to cardiac function, a deep breath can affect cardiac measures. Respiration is most accurately measured by gas exchange in the lungs, but the sensor technology inhibits talking and moving. Instead, chest cavity expansion can be used to capture breathing activity using either a Hall Effect sensor, strain gauge, or a stretch sensor [4].

1.2.4 Electromyography

Electromyography (EMG) measures muscle activity by detecting surface voltages that occur when a muscle is contracted [4]. In isometric conditions (no movement) EMG is closely correlated with muscle tension, however, this is not true of isotonic movements (when the muscle is moving). When used on the jaw, EMG provides a very good indicator of tension in an individual due to jaw clenching [6]. On the face, EMG has been used to distinguish between positive and negative emotions. EMG activity over the brow (frown muscle) region is lower and EMG activity over the cheek (smile muscle) is higher when emotions are mildly positive, as opposed to mildly negative [6].

1.3 General Stress Detection System

Measuring physiological conditions offers a practical method of determining the mental stress level of an automobile driver. Driver's stress level can be measured using recording acquired from physiological sensors. Such recordings can be used by stress detection system installed within the automobiles. These in-vehicle electronic systems may further improve the decision making capability of the driver, and makes an intelligent transportation system.

Galvanic Skin Response (GSR) and Heart Rate (HR) are the two most important physiological parameters for detecting driver's drowsiness, fatigue or stress level under variable traffic conditions. Several other physiological parameters like Respiration Rate (RR), ElectroMyoGram (EMG) also affect the driver stress level by a small factor. Their effect can also be combined to form a reliable and more accurate stress detection system/model. The work described in this research aims to design a stress detection system that can monitor the stress level of a driver using physiological parameters such as GSR, EMG, RR and HR as shown in Fig. 4.

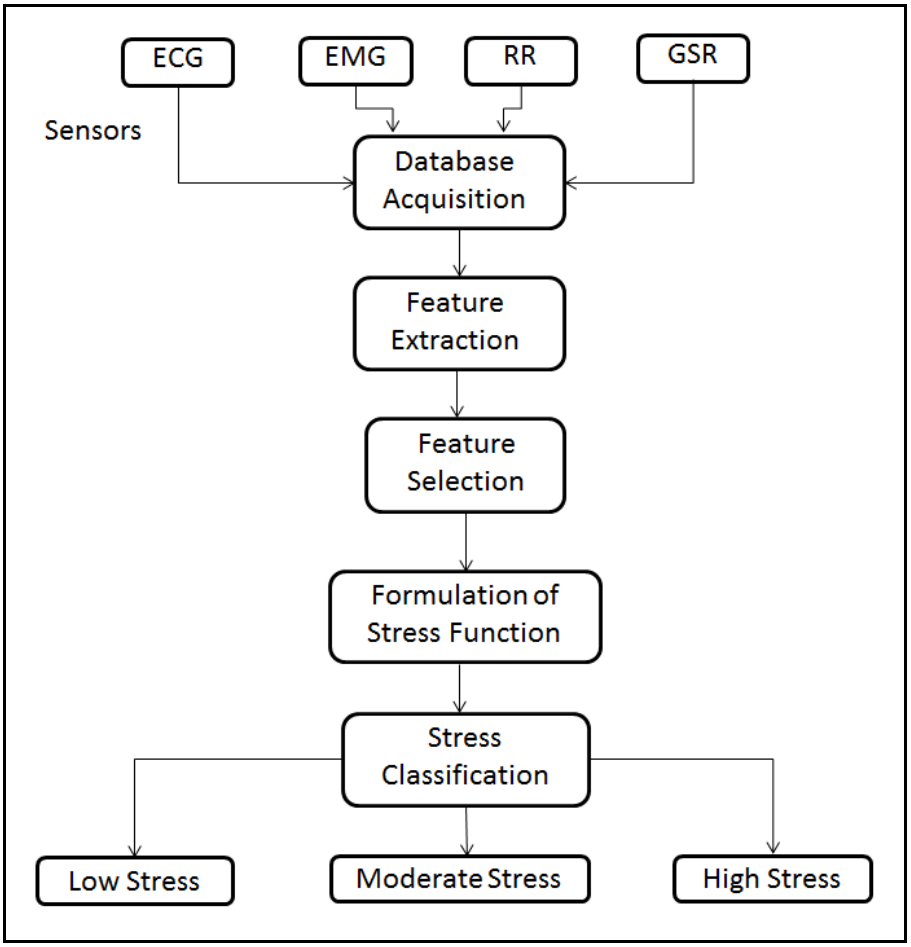

Fig. 4. General Layout of the proposed Stress Detection System

Healey and Picard presented methods to evaluate driver's relative stress level using physiological signals. The study indicated that Galvanic Skin Response (GSR) and heart rate (HR) signals were significantly correlated to driver stress [2]. Stress is known to activate the Sympathetic Nervous System (SNS). Much research has already been done on the detection of stress from physiological parameters that are influenced by the SNS. Examples are muscle activity, heart rate, heart rate variability, skin conductance and pupil diameter. Other studies have shown that a combination of these physiological parameters facilitates differentiation between stressful situations and situations without stress.

Chapter 2
LITERATURE REVIEW

Assessing human stress in real-time is more difficult and challenging today. The present review deals about the measurement of stress in laboratory environment using different stress inducement stimuli by the help of physiological signals. Previous researchers have been used different stress inducement stimuli such as stroop colour word test (CWT), mental arithmetic test, public speaking task, cold pressure test, computer games and works used to induce the stress. Most of the researchers have been analysed stress using questionnaire based approach and physiological signals. The several physiological signals like ElectroCardioGram (ECG), ElectroMyoGram (EMG), Galvanic Skin Response (GSR), Blood Pressure (BP), Skin Temperature (ST), Blood Volume Pulse (BVP), Respiration Rate (RR) and ElectroEncephaloGram (EEG) were briefly investigated to identify the stress. Different statistical methods like Analysis of variance (ANOVA), two-way ANOVA, Multivariate analysis of variance (MANOVA), t-test, paired t-tests and student t-tests have used to describe the correlation between stress inducement stimuli, subjective parameters (age, gender and etc.,) and physiological signals.

The purpose of this chapter is to present a review of the various studies conducted in the past on the subject of stress detection of an automobile driver using physiological parameters and subjects related to features extraction and selection from the acquired physiological parameters. The brief details about these studies are listed underneath:

G. Rigas *et al.* proposed a method based on a Dynamic Bayesian Network for the estimation of car driver's stress produced due to specific driving events. To evaluate how strongly the extracted features, Heart Rate Variation from Baseline (HRvB) and Electro Dermal Activity (EDA) mean of first absolute differences (MFAD) are related to the underlying stress level, they utilized the correlation of ECG and EDA features and the correlation of those features with the stress metric described for each tour. They noticed that correlations of HRvB and MFAD are non-random with high probability. Moreover, both features are significantly non-random correlated with the stress metric. They also monitors driver's stress using selected bio-signals and provides a probabilistic reasoning-based framework and infer results that indicate a strong correlation between the level of the stress as reported by the driver and the outcome of their model [7].

S. Li *et al.* proposed a system that can actively monitor the driver's fatigue level in real time for the prevention of accidents. For that they used Support vector machine (SVM) technique to identify driver's fatigue based on psychological features, such as EEG and ECG. Driver's fatigue is expressed as alert, mild fatigue, deep fatigue and drowsiness, and they are used as output variables of SVM model. Field experiments were carried out to collect the required data to validate the SVM model. Driver's fatigue levels obtained are used as output variables of SVM model and that resulted in a model that can recognize driver's fatigue levels effectively and recognition precisions of all states are larger than 87.5% [8].

I. C. Jeong *et al.* intends to develop devices which detects the electrocardiogram (ECG) signal of driver in real time and provide autonomic nervous system (ANS) information and degree of stress through analysis of Heart Rate Variability (HRV). ECG signal is converted into HRV for analysis of time area and frequency area and the response of autonomic nervous system and stress level is calculated through algorithm. ECG reflects the health state of subject as well as checks the degree of operation of autonomic nervous system through analysis of Heart Rate Variability (HRV). Any alteration of autonomic nervous system predicts the stress level of drivers during operation and provides the possibility of warning by continuous detection, it may lead to the recognition and safe driving of driver by detecting the stress during driving altered by diverse factors such as changing mood, bio rhythm, condition, fatigue, boredom or disease and preventing the driver from reaching the inappropriate state for driving [9].

R. R. Singh *et al.* presents a comprehensive analysis of extraction and signal processing techniques of features from physiological signals. They adopted Self-Organizing Map to cluster data into topographically distinct clusters of low, medium and high stress states and developed cumulative sum-based stress metric, capable of detecting over-stress conditions, using Page's Technique. They worked on feature extraction and processing techniques and used on GSR and PPG signals. KSOM based cluster analysis of principal components resulted in three topographically distinct clusters of stress state with an overall predictive rate of 81.60%. Further, the drivers were categorized into three broad groups based on the Stress Susceptibility Index during profile analysis. Finally, combining the knowledge of the driver's stress susceptibility with the observed stress states, a CUSUM based stress metric was developed using incremental weights obtained from reciprocal analysis of the mean driving times [10].

J. A. Healey and R. W. Picard collected and analysed physiological signals like electrocardiogram, electromyogram, skin conductance, and respiration continuously during real-world driving tasks to determine a driver's relative stress level. Their experiment was designed to monitor driver's physiological reactions during real-world driving situations. The data were analysed in two ways. Analysis I used features from 5-min intervals of data during the rest, highway, and city driving conditions to distinguish three levels of driver stress with an accuracy of over 97% across multiple drivers and driving days. Analysis II compared continuous features, calculated at 1-s intervals throughout the entire drive, with a metric of observable stressors created by independent coders from videotapes. Results showed that for most drivers studied, skin conductivity and heart rate metrics are most closely correlated with driver stress level. Such a metric could be used to help manage noncritical in-vehicle information systems and could also provide a continuous measure of how different road and traffic conditions affect drivers. Their findings indicate that physiological signals can provide a metric of driver stress in future cars capable of physiological monitoring [2].

Y. Deng *et al.* proposed a feature selection method based on Principal Component Analysis (PCA) and evaluated their effectiveness in terms of correct rate and computational time using five classification algorithms, Linear Discriminant Function, C4.5 induction tree, Support Vector Machine (SVM), Naïve Bayes and K Nearest Neighbor (KNN). Their study demonstrated the importance of feature selection and the effectiveness of the methods used in accurately classifying stress levels [11].

Y. Shi *et al.* build a personalized stress detection model based on Support Vector Machines, and evaluated it on the collected. Each subject in their study was exposed to a protocol containing four stressors and six rest periods. The four stressors were: one public speaking stressor, two mental arithmetic stressors, and one cold pressure stressor. These stressors represent the social, mental, or physical challenges that might lead to either mental or physical stress. Experimental results show that their models can detect stress at high precision and recall values, especially when personalized information is used [12].

A. Barreto and J. Zhai developed a stress detection system based on the physiological signals monitored by non-invasive and non-intrusive sensors. The development of this emotion recognition system involved three stages: experiment setup for physiological sensing, signal pre-processing for the extraction of affective features and affective recognition using a learning system. Four signals: Galvanic Skin Response (GSR), Blood Volume Pulse (BVP), Pupil Diameter (PD) and Skin Temperature (ST) are monitored and analyzed to differentiate affective states in a computer user. A Support Vector Machine is used to perform the supervised classification of affective states between "stress" and "relaxed". Their results indicated that the physiological signals monitored have a strong correlation with the changes in emotional state of their experimental subjects when stress stimuli are applied to the interaction environment. They also found that the pupil diameter was the most significant affective state indicator, compared to the other three physiological signals monitored [13].

M. Kumar *et al.* presented a method of HRV analysis for mental stress assessment using fuzzy clustering and robust identification techniques. Their approach consists of 1) online monitoring of heart rate signals, 2) signal processing (e.g., using the continuous wavelet transform to extract the local features of HRV in time-frequency domain), 3) exploiting fuzzy clustering and fuzzy identification techniques to render robustness in HRV analysis against uncertainties due to individual variations, and 4) monitoring the functioning of autonomic nervous system under different stress conditions [14].

Y. Shi *et al.* used a physiological measure, namely Galvanic Skin Response (GSR), to evaluate stress and arousal levels while using unimodal and multimodal versions of the same interface. Preliminary results showed that user's GSR readings significantly increases when task cognitive load level increases. Moreover, users' GSR readings were found to be lower when using a multimodal interface, instead of a unimodal interface. Cross-examination of GSR data with multimodal data annotation showed promising results in explaining the peaks in the GSR data, which are found to correlate with sub-task user events [15].

J. Wijsman *et al.* they designed two new stress tests, which aimed at creating circumstances that, are similar to work stress. An experiment was described in which EMG signals of the upper trapezius muscle were measured during three different stressful situations. Stress tests included a calculation task, a logical puzzle task and a memory task, of which the last two were newly designed. The results showed significant higher amplitudes of the EMG signals during stress compared to rest and fewer gaps (periods of relaxation) during stress. Also, mean and median frequencies were significantly lower during stress than during rest. The differences in EMG features between rest and stress conditions indicate that EMG is a useful parameter to detect stress. These results show opportunities for the inclusion of EMG sensors in a wireless system for ambulatory monitoring of stress levels [16].

J. W. Shin *et al.* had done a knowledge based synthetic analysis using 5 types of bio-signals to reason and to evaluate the mutual physiological phenomena with fuzzy theory. With knowledge based synthetic analysis using human intelligence algorithm not only the evaluation on stress stages but also on the five senses and various emotions can be done. Due to fast calculation of evaluation achievement this technique can be applied to real-time sensibility monitor [17].

A. de S. Sierra *et al.* described a stress detection system based on fuzzy logic and two physiological signals: Galvanic Skin Response and Heart Rate. Instead of providing a global stress classification, their approach created an individual stress template, gathering the behavior of individuals under situations with different degrees of stress. The proposed method is able to detect stress properly with a rate of 99.5%, being evaluated with a database of 80 individuals. Their results were improved from former approaches in the literature and well-known machine learning techniques like SVM, k-NN, GMM and Linear Discriminant Analysis. Their proposed method was highly suitable for real-time applications [18].

M. M. Bundele. and R. Banerjee aimed to design a simple wearable computing system using noninvasive type of physiological parameter-based sensors for the detection of fatigue / stress level of a driver and alerts the driver in time so as to avoid any accident. The skin conductance and the oximetry pulse signals of the vehicular drivers have been recorded for various states. The features extracted were subsequently used to design multilayer perceptron neural network (MLP NN) to fetch an optimal set of performance measures. They have done analysis of one hidden layer and two hidden layer MLP NN using Receiver Operating Characteristics (ROC). They designed a two-state classifier using MLP NN and the classifier performance has been analyzed using the ROC method and the independent validation method. The work establishes the correlation of fatigue with Skin Conductance (SC) and Oximetry Pulse (PO) and consequently presents a design for MLP NNs for the detection of fatigue level of a vehicular driver [19].

A. Akbas identified the usefulness of metrics that indicate the real-time stress level of drivers. He completed the evaluations by using the available segment based arrays of instantaneous heart rate (IHR), hand based skin conductance (HSC), foot based skin conductance (FSC) and electromyography (EMG) signals. He derived the segment based data arrays including instantaneous respiratory rate (IRR) and average number of contractions/minutes (CPM) from the respiratory and EMG signals, respectively by using a peak detection algorithm. Statistical comparisons using the overall mean and mean values of the IHR, HSC, FSC, EMG, IRR and CPM data arrays showed that with an optional exception of IRR all of these metrics can be identified as useful parameters in the future car technology to determine the dynamic stress level of drivers [20].

A. Drachen *et al.* had done a case study on HR and EDA correlations with subjective gameplay experience, testing the feasibility of these measures in commercial game development contexts. Their results indicate a significant correlation ($p < 0.01$) between psychophysiological arousal (i.e., HR, EDA) and self-reported gameplay experience. However, the covariance between psychophysiological measures and self-reports varies between the two measures. The results are consistent across three different contemporary major commercial first-person shooter (FPS) games [21].

Chapter 3
DATABASE ACQUISITION

3.1 Data Collection

A lot of work has already been done, but only a few physiological sensor data sets which can be used for detection of stress have been made available for further exploration. Fortunately, there is one available from the PHYSIONET website [22] which contains sensors signal data directly obtained from different drivers. These driver data sets are originally collected by Healy & Picard [2] from MIT Media Lab. It provides sufficient data for us to do some further work about the feature selection and stress identification.

PHYSIONET is an on-line forum for the dissemination and exchange of recorded biomedical signals and open-source software for analyzing them. It provides facilities for the cooperative analysis of data and the evaluation of proposed new algorithms. It also provides free electronic access to PhysioBank data and PhysioToolkit software via the World Wide Web (http://www.physionet.org).

We have obtained all our raw physiological signals from the database of PHYSIONET website located at http://www.physionet.org/physiobank/database/drivedb/. This database, contributed to PHYSIONET by its creator, Jennifer Healey, contains a collection of multi-parameter recordings from healthy volunteers, taken while they were driving on a prescribed route including city streets and highways in and around Boston, Massachusetts. However these data sets actually are not as complete as those in the experiment of Healey and Picard [2]. In their original work, a total of seventeen drivers participated in the experiment and for each driver eight types (Time Stamp, ECG, EMG, Foot GSR, Hand GSR, IHR, Marker, and Respiration) of raw data are acquired from the sensors that the drivers wear. On carefully inspecting the data it was found that amongst the seventeen data sets, only ten drives' data sets (drives 05, 06, 07, 08, 09, 10, 11, 12, 15 and 16) are complete, which include all the sensor information as well as have clear mark identification. The remaining seven drives' data sets (drives 01, 02, 03, 04, 13, 14, 17a and 17b) do not contain all the sensor information and the mark of different driving period is not clear. The work conducted by Yong Deng et al. [11] also helped us to clearly understand the selection of different drivers' data sets.

We have downloaded our entire database from PHYSIONET Website in the form of .csv files which is in the form of comma-separated values and can be directly opened in MS Excel Software [23]. But one major problem that we encountered while opening our downloaded .csv files is that they are very long columns of data and not within the permissible range of MS Excel which is 65,536 rows by 256 columns. To overcome this problem, two software's viz; CSV Splitter v1.1 by Sopheap Ly and Merge Excel Sheets v 4.0 © 2010 by Abdel Yezza were used for opening these .csv files. We divided single .csv files into multiple file of one minute duration each for the whole one hour long drive using CSV Splitter and then merged them column-wise using Merge Excel Sheets. By this method we were able to download and save all ten different drivers data set in MS Excel.

Each signal acquired from PHYSIONET was sampled at a rate appropriate for capturing the information contained in the signal constrained by the sampling rates available on the FlexComp system. The ECG was sampled at 496 Hz, the skin conductivity (SC) and RESP signals were sampled at 31 Hz, and the EMG was sampled at 15.5 Hz after first passing through a 0.5 s averaging filter. The signals were collected by embedded computer in the testing car [2].

The data acquired from **Healey and Picard's** [2] experiment lacks the information regarding the duration of each Rest, City and Highway driving task, but the same durations were mentioned in **Ahmet Akbas** [20] research and we found it to be very useful in our research for stress detection in automobile drivers. List of the time intervals for different driving segment for each of the 10 selected Drive's i.e. Drive 05, 06, 07, 08, 09, 10, 11, 12, 15 and 16 are shown in Table 1.

Table 1

Time intervals of the 7 driving segments of available bio-signal datasets.

Drive No.	Driving period (min)							Total rec. time (min)
	Initial Rest	City 1	Highway 1	City 2	Highway 2	City 3	Final Rest	
Drive05	15.13	16	7.74	6.06	7.56	14.96	15.78	83.23
Drive06	15.05	14.49	7.32	6.53	7.64	12.29	15.05	78.38
Drive07	15.04	16.23	10.96	9.83	7.64	10.15	15.03	84.87
Drive08	15	12.31	7.23	9.51	7.64	13.43	15.07	80.19
Drive09	15.66	19.21	8.47	5.2	7.06	13.21	NA	68.82
Drive10	15.04	15.3	8.66	5.27	7.04	12.06	14.79	78.15
Drive11	15.02	15.81	7.43	7.15	6.96	11.72	14.99	79.08
Drive12	15.01	13.41	7.56	6.5	8.06	11.68	15.01	77.23
Drive15	15	12.54	7.24	5.99	6.82	12.12	15	74.7
Drive16	15.01	16.12	7.14	5.12	6.81	13.91	NA	64.1

NA: Not available

3.2 Bio-signals Acquired

In our research, physiological data were gathered from PHYSIONET website. And based on previous literature survey, we chose to collect Galvanic Skin Response (GSR), ElectroCardioGraphy (ECG), ElectroMyoGram (EMG) of the trapezius muscle, and Respiration Rate (RR). Heart rate (HR) was also available, while Respiration Rate (RR) was computed from the raw respiration data. As the subjects were operating an automobile, it wasn't possible to constrain their movements so no data were collected by Healey and Picard [2], which would include artifact's. Different signals acquired for 'drive05' in Healey and Picard's experiment [2] are shown in Fig. 5.

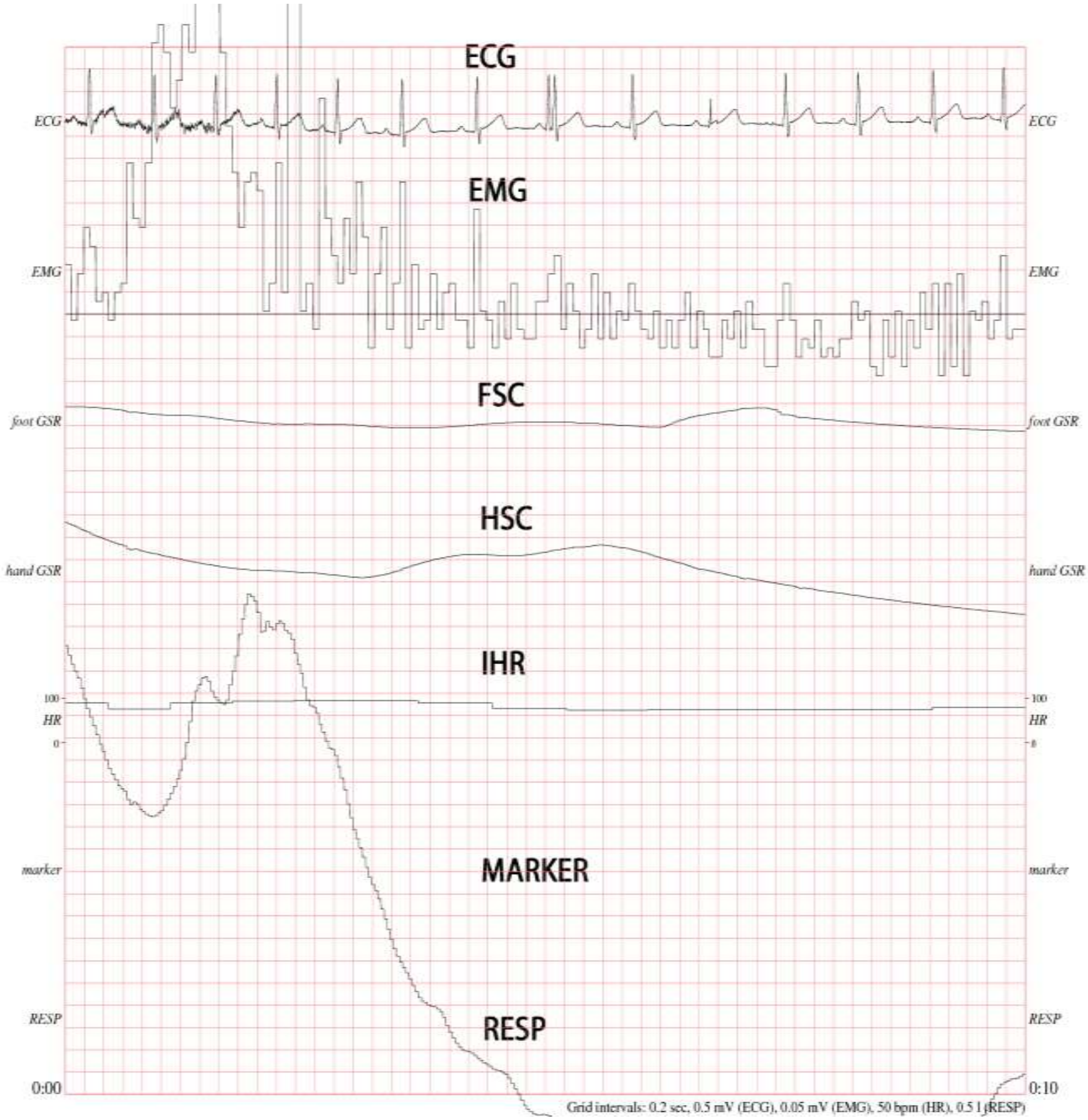

Fig. 5. A 10-s close up of original 'drive05' bio-signal dataset.

Chapter 4
MATERIALS and METHODS

4.1 Pre-processing of acquired database

The data acquired from PHYSIONET website contains recording of 17 different drives. Each drive dataset consists of ECG, EMG, HR, GSR and RESP. signal of approximately 1 hour duration or more. We divided the whole 1 hour long data set into many 1 minute dataset. After that we extracted features from that dataset for each minute and similarly for every drives.

4.2. Feature Extraction

We have done a literature survey on the topic of stress detection and selected four physiological parameters like ECG, EMG, GSR and RR that will help in detecting stress of automobile driver's easily. These four signals were selected based on their properties regarding non-invasivity when being acquired and because their variation is strongly related to stress stimuli [2]. Overall 8 statistical features were extracted from the 4 available physiological parameters.

Heart rate variability (HRV) analysis is commonly used as a quantitative marker depicting the activity of autonomic nervous system (ANS) that may be related to mental stress [24]. Although HRV features can be extracted by detecting QRS complexes from electrocardiogram (ECG) signals, the signals acquired from PHYSIONET website already contained HR signal along with ECG signal. So, we have used available HR signal directly to extract features like Mean HR, HRV and normalised HRV (NHRV).

GSR is a bio-electric physiological signal controlled by the sympathetic nervous system. The GSR signal comprises of two components: tonic (Frequency Range: 0.0Hz to 0.16Hz) and phasic (Frequency Range: 0.16Hz and above) [10]. Features extracted from GSR can give significant emotion information about an individual like surprise, fear, disgust, grief, and happy [25]. In our work we have extracted Mean hand GSR and Mean foot GSR features. All mean values are taken over a period of 60 seconds. Thus minutes to minute mean values are used for analysis.

The stress changes amplitude and temporal features of the EMG of the upper trapezius muscle. A statistically significant relation was found between stress level and features extracted from EMG signals of the upper trapezius muscle. Amplitude of the EMG signal is reported to be significantly higher during stress conditions than during rest [16][26]. We used Mean EMG and rms EMG features, again on minute to minute basis.

Respiration is primarily regulated for metabolic and homeostatic purposes in the brainstem. However, breathing can also change in response to changes in emotions, such as sadness, happiness, anxiety or fear [27]. We used Respiration Rate per minute (RR) as a feature for stress detection. Table 2 gathers a list of the physiological signals and the corresponding statistical features extracted from them. These extracted features can be categorized in the following groups:

Table 2

List of the physiological signals and the corresponding features extracted from them.

Physiological Signals	Features Selected
GSR (Galvanic Skin Response)	Mean hand GSR, Mean foot GSR
EMG (Electromyogram)	Mean EMG, rms EMG
Respiration Rate (RR)	Respiration Rate per minute (RR)
ECG (Electrocardiogram)	Mean Heart Rate (HR), Heart Rate Variability (HRV), Normalised HRV (NHRV)

We have calculated 8 statistical features for every minute of the driving period for each of the 10 selected driver's data set i.e. drive 5, 6, 7, 8, 9, 10, 11, 12, 15 and 16 as shown in Table 3. All these features are then accumulated to form a feature matrix and correlation is evaluated to check the strength of relationship between these features.

Table 3

List of 8 statistical features extracted and their description.

S. No.	Features	Description
1	Mean hand GSR	The mean of the hand GSR data samples for each minute
2	Mean foot GSR	The mean of the foot GSR data samples for each minute
3	Mean EMG	The mean of the EMG data samples for each minute
4	rms EMG	The root-mean-square of mean EMG for each minute
5	RR	The mean of respiration rate for each minute
6	HR	The mean of the heart rate data samples for each minute
7	HRV	The standard deviation of heart rate data for each minute
8	NHRV	The normalised standard deviation of heart rate data for each minute

4.3 Making Feature Matrix

From the duration of different driving segment obtained from Table 1 we have defined a Class for our database so that the correlation analysis of the database can be done properly. These Category and Class are shown in Table 4. Our approach in defining the Stress Class of automobile driver is to consider the rest period as low stress condition, highway driving as moderate stress condition and city drive as high stress condition. Neglecting all other factor towards the contribution of stress, an analysis of correlation of various physiological parameters with the three mentioned states of mental stress has been made in this research. Further we have chosen numerical value 1 for low stress, 3 for moderate stress and 5 for high.

After selecting 8 statistical features, we arranged all the data in a matrix form that shows the mean of data acquired at every minute of the total driving duration. We have calculated Feature Matrix for all drivers' data sets (drivers 5, 6, 7, 8, 9, 10, 11, 12, 15 and 16) and only Drive07 data set is shown below in Table 5.

Table 4
Driving Segments divided into different Category and Numerical Stress Class.

Driving Segments Used	Category	Stress Class
Initial Rest + Final Rest	Relaxed or Low Stress (LS)	**1**
Highway 1 + Highway 2	Moderate Stress (MS)	**5**
City 1 + City 2 + City 3	High Stress (HS)	**10**

Table 5
Feature Matrix for Drive07.

Sample no. (per minute)	Mean Hand GSR	Mean Foot GSR	Mean EMG	rms EMG	RR	Mean HR	HRV	NHRV	Class
1	5.405	8.519	0.057	0.061	12	81.833	7.191	0.088	**1**
2	4.662	7.123	0.074	0.100	16	76.780	6.472	0.084	**1**
3	5.553	8.983	0.057	0.091	16	78.316	7.228	0.092	**1**
4	5.613	9.066	0.053	0.055	14	69.015	5.039	0.073	**1**
5	4.835	7.752	0.054	0.063	15	70.960	4.774	0.067	**1**
6	4.531	6.891	0.058	0.065	14	73.792	5.074	0.069	**1**
7	4.867	6.763	0.081	0.170	14	78.218	6.261	0.080	**1**
8	4.475	6.065	0.080	0.179	14	74.114	3.865	0.052	**1**
9	4.209	5.820	0.051	0.053	13	72.760	5.899	0.081	**1**
10	4.212	5.834	0.051	0.053	13	71.631	7.864	0.110	**1**
11	3.919	5.450	0.054	0.059	12	73.971	6.642	0.090	**1**
12	3.658	5.086	0.056	0.058	13	69.633	4.826	0.069	**1**
13	3.479	4.898	0.082	0.164	13	70.898	6.125	0.086	**1**
14	3.345	4.736	0.052	0.055	12	67.438	5.765	0.085	**1**
15	3.244	4.584	0.059	0.077	12	68.628	7.225	0.105	**1**
16	3.166	4.482	0.055	0.061	13	67.990	4.893	0.072	**10**
17	3.657	4.905	0.327	0.788	15	67.862	9.874	0.146	**10**
18	10.577	12.362	0.468	0.854	23	89.503	6.190	0.069	**10**
19	7.761	10.920	0.159	0.240	15	79.481	11.516	0.145	**10**

20	8.994	11.988	0.165	0.277	19	78.983	6.139	0.078	**10**
21	7.653	11.383	0.101	0.188	18	78.132	5.179	0.066	**10**
22	7.915	11.491	0.144	0.322	17	79.863	6.314	0.079	**10**
23	6.815	10.252	0.127	0.205	17	75.717	5.210	0.069	**10**
24	6.553	9.942	0.111	0.207	17	77.171	7.358	0.095	**10**
25	7.175	10.471	0.135	0.245	17	75.858	5.547	0.073	**10**
26	7.420	11.141	0.143	0.234	18	78.634	6.141	0.078	**10**
27	7.776	11.662	0.185	0.515	17	78.337	6.189	0.079	**10**
28	7.865	11.338	0.279	0.407	19	81.855	4.583	0.056	**10**
29	8.568	12.614	0.410	0.624	21	86.956	5.981	0.069	**10**
30	8.449	12.275	0.172	0.459	18	78.926	6.140	0.078	**10**
31	9.029	12.970	0.446	0.846	21	79.302	5.007	0.063	**10**
32	7.769	12.594	0.234	0.779	21	78.998	5.902	0.075	**5**
33	8.615	13.676	0.383	0.513	23	83.080	5.903	0.071	**5**
34	7.157	12.885	0.141	0.190	19	71.349	3.454	0.048	**5**
35	6.669	12.076	0.125	0.548	19	72.735	4.808	0.066	**5**
36	9.586	14.884	0.248	0.336	21	82.368	10.693	0.130	**5**
37	8.288	14.166	0.118	0.149	24	73.699	5.507	0.075	**5**
38	6.741	12.934	0.157	1.177	17	71.101	2.862	0.040	**5**
39	6.609	12.442	0.353	0.870	19	73.775	5.174	0.070	**5**
40	6.496	12.264	0.061	0.066	16	75.896	5.680	0.075	**5**
41	5.961	11.536	0.106	0.261	16	72.530	3.843	0.053	**5**
42	6.725	12.501	0.142	0.182	17	76.559	5.279	0.069	**5**
43	6.505	12.059	0.069	0.076	17	73.087	3.445	0.047	**5**
44	6.714	12.179	0.167	0.236	19	75.833	4.049	0.053	**10**
45	8.446	14.538	0.364	0.673	23	85.689	5.806	0.068	**10**
46	6.824	12.525	0.170	0.834	16	72.566	4.450	0.061	**10**
47	6.541	11.877	0.131	0.163	18	74.768	5.082	0.068	**10**
48	7.268	13.050	0.219	0.482	18	76.799	5.594	0.073	**10**

49	6.543	12.536	0.195	0.224	18	74.455	5.645	0.076	**10**
50	7.264	12.873	0.138	0.207	17	73.595	4.119	0.056	**10**
51	8.868	14.495	0.357	0.861	18	78.555	6.515	0.083	**10**
52	7.478	13.735	0.257	0.904	19	76.695	4.644	0.061	**10**
53	7.405	13.434	0.295	0.377	18	77.609	4.326	0.056	**10**
54	7.379	14.040	0.302	0.474	18	80.022	7.822	0.098	**10**
55	6.893	13.564	0.171	0.223	17	71.534	3.654	0.051	**5**
56	5.974	12.529	0.135	0.509	17	71.580	3.928	0.055	**5**
57	5.719	12.186	0.093	0.139	16	71.733	2.803	0.039	**5**
58	5.695	11.942	0.082	0.103	16	70.982	4.226	0.060	**5**
59	5.484	11.707	0.070	0.075	16	71.431	3.271	0.046	**5**
60	6.284	12.729	0.160	0.256	18	71.757	3.595	0.050	**5**
61	7.699	14.605	0.207	0.243	17	73.589	3.935	0.053	**10**
62	8.651	14.622	0.496	0.990	18	79.270	6.472	0.082	**10**
63	7.584	13.597	0.184	0.569	19	74.971	7.127	0.095	**10**
64	7.184	12.904	0.158	0.461	16	68.663	3.562	0.052	**10**
65	8.442	14.651	0.242	0.525	18	79.867	6.134	0.077	**10**
66	7.209	13.820	0.135	0.372	18	72.704	5.422	0.075	**10**
67	7.388	14.338	0.267	0.578	17	72.715	4.595	0.063	**10**
68	7.685	14.518	0.277	0.448	15	74.467	6.384	0.086	**10**
69	8.007	14.703	0.160	0.379	18	72.686	6.715	0.092	**10**
70	7.593	14.588	0.242	0.875	19	80.497	9.320	0.116	**1**
71	7.505	14.866	0.402	0.547	21	85.337	3.866	0.045	**1**
72	7.175	14.903	0.385	0.909	16	81.974	10.208	0.125	**1**
73	5.953	13.523	0.056	0.060	12	68.127	6.520	0.096	**1**
74	5.372	12.746	0.054	0.056	14	66.483	5.532	0.083	**1**
75	5.065	11.778	0.052	0.057	13	63.897	4.816	0.075	**1**
76	4.933	11.034	0.052	0.053	14	64.958	4.390	0.068	**1**
77	5.634	12.315	0.085	0.121	10	76.452	11.123	0.145	**1**

78	5.433	11.484	0.053	0.054	12	67.866	4.858	0.072	**1**
79	6.274	12.063	0.080	0.235	13	73.711	11.590	0.157	**1**
80	7.718	13.244	0.059	0.067	13	71.318	6.757	0.095	**1**
Mean	**6.597**	**11.400**	**0.168**	**0.341**	**16.650**	**75.036**	**5.799**	**0.077**	
Std.Dev.	**1.599**	**2.958**	**0.115**	**0.286**	**2.924**	**5.064**	**1.897**	**0.024**	

4.4 Standardization of Feature Matrix

As we can see from the above table that the range of all the data is different so standardisation of these data is required. Standardization of data is necessary because, if the input variables are combined linearly, then it is strictly necessary to standardize the inputs, at least in theory. The reason is that any rescaling of an input vector can be effectively undone by changing the corresponding weights and biases, leaving the user with the exact same outputs as the user had before. It will make all the data to come into one particular range so that the data interpretation and analysis can be done easily. Table 6 shows standardized matrix for 'Drive 07'.

Hence, to standardize the data, you will want the data to reflect how many standard deviations from the average that that data lies, with the following normal distribution curve representing the probability of each standard deviation for a normal distribution

The conclusion is that data should be normalized or standardized to remove their scale from your model, but both techniques produce identical results to this desired outcome. However, standardizing is the preferred method because it produces meaningful information about each data point, and where it falls within its normal distribution, plus provides a crude indicator of outliers (i.e., anything above or below a Z-Score of ±4).

$$\boxed{Z = \frac{X_i - \mu}{\sigma}}$$

Where,

X_i = each data point i

μ = the average of all the sample data points

σ = the sample standard deviation of all sample data points

Z = the data point i standardized to 1σ, also known as **Z-Score**

Table 6

The Standardized Features of Drive07.

Sample no. (per minute)	Mean Hand GSR	Mean Foot GSR	Mean EMG	rms EMG	RR	Mean HR	HRV	NHRV	Class
1	-0.745	-0.974	-0.959	-0.976	-1.590	1.342	0.734	0.444	**1**
2	-1.210	-1.446	-0.818	-0.840	-0.222	0.344	0.355	0.296	**1**
3	-0.653	-0.817	-0.960	-0.871	-0.222	0.648	0.753	0.627	**1**
4	-0.615	-0.789	-0.997	-0.999	-0.906	-1.189	-0.401	-0.170	**1**
5	-1.103	-1.234	-0.984	-0.970	-0.564	-0.805	-0.541	-0.407	**1**
6	-1.292	-1.525	-0.949	-0.962	-0.906	-0.246	-0.382	-0.346	**1**
7	-1.082	-1.568	-0.754	-0.596	-0.906	0.628	0.244	0.121	**1**
8	-1.327	-1.804	-0.757	-0.565	-0.906	-0.182	-1.019	-1.032	**1**
9	-1.494	-1.887	-1.014	-1.006	-1.248	-0.449	0.053	0.163	**1**
10	-1.492	-1.882	-1.014	-1.006	-1.248	-0.672	1.089	1.350	**1**
11	-1.675	-2.012	-0.986	-0.983	-1.590	-0.210	0.444	0.523	**1**
12	-1.839	-2.135	-0.973	-0.989	-1.248	-1.067	-0.513	-0.324	**1**
13	-1.951	-2.199	-0.746	-0.618	-1.248	-0.817	0.172	0.383	**1**
14	-2.034	-2.253	-1.000	-0.999	-1.590	-1.501	-0.018	0.346	**1**
15	-2.097	-2.305	-0.947	-0.920	-1.590	-1.265	0.752	1.164	**1**
16	-2.146	-2.339	-0.975	-0.978	-1.248	-1.391	-0.477	-0.213	**10**
17	-1.839	-2.196	1.376	1.561	-0.564	-1.417	2.149	2.827	**10**
18	2.489	0.325	2.596	1.791	2.171	2.857	0.206	-0.330	**10**
19	0.728	-0.162	-0.079	-0.353	-0.564	0.878	3.014	2.801	**10**
20	1.499	0.199	-0.022	-0.223	0.804	0.779	0.179	0.025	**10**
21	0.660	-0.006	-0.582	-0.531	0.462	0.611	-0.327	-0.448	**10**
22	0.824	0.031	-0.209	-0.065	0.120	0.953	0.272	0.080	**10**
23	0.136	-0.388	-0.354	-0.475	0.120	0.135	-0.311	-0.344	**10**
24	-0.027	-0.493	-0.493	-0.467	0.120	0.422	0.822	0.753	**10**
25	0.361	-0.314	-0.282	-0.332	0.120	0.162	-0.133	-0.166	**10**

26	0.514	-0.088	-0.220	-0.372	0.462	0.711	0.180	0.040	**10**
27	0.737	0.088	0.147	0.609	0.120	0.652	0.206	0.077	**10**
28	0.793	-0.021	0.965	0.230	0.804	1.347	-0.641	-0.874	**10**
29	1.233	0.410	2.101	0.990	1.487	2.354	0.096	-0.345	**10**
30	1.158	0.296	0.036	0.413	0.462	0.768	0.180	0.028	**10**
31	1.521	0.531	2.408	1.763	1.487	0.842	-0.418	-0.579	**10**
32	0.733	0.404	0.569	1.532	1.487	0.782	0.054	-0.100	**5**
33	1.262	0.769	1.859	0.600	2.171	1.588	0.055	-0.251	**5**
34	0.350	0.502	-0.235	-0.526	0.804	-0.728	-1.236	-1.187	**5**
35	0.045	0.228	-0.368	0.725	0.804	-0.454	-0.523	-0.456	**5**
36	1.869	1.178	0.696	-0.018	1.487	1.448	2.580	2.178	**5**
37	1.057	0.935	-0.435	-0.668	2.513	-0.264	-0.154	-0.099	**5**
38	0.090	0.519	-0.095	2.919	0.120	-0.777	-1.549	-1.525	**5**
39	0.008	0.352	1.607	1.847	0.804	-0.249	-0.329	-0.289	**5**
40	-0.063	0.292	-0.924	-0.960	-0.222	0.170	-0.062	-0.094	**5**
41	-0.398	0.046	-0.537	-0.278	-0.222	-0.495	-1.032	-0.998	**5**
42	0.080	0.372	-0.222	-0.552	0.120	0.301	-0.274	-0.338	**5**
43	-0.058	0.223	-0.858	-0.924	0.120	-0.385	-1.241	-1.240	**5**
44	0.073	0.263	-0.008	-0.365	0.804	0.157	-0.923	-0.981	**10**
45	1.156	1.061	1.698	1.159	2.171	2.104	0.004	-0.388	**10**
46	0.142	0.380	0.015	1.722	-0.222	-0.488	-0.711	-0.653	**10**
47	-0.035	0.161	-0.316	-0.622	0.462	-0.053	-0.378	-0.379	**10**
48	0.420	0.558	0.446	0.495	0.462	0.348	-0.108	-0.178	**10**
49	-0.034	0.384	0.232	-0.409	0.462	-0.115	-0.081	-0.054	**10**
50	0.417	0.498	-0.260	-0.468	0.120	-0.285	-0.886	-0.875	**10**
51	1.420	1.046	1.642	1.818	0.462	0.695	0.378	0.240	**10**
52	0.551	0.789	0.770	1.966	0.804	0.328	-0.609	-0.685	**10**
53	0.505	0.688	1.103	0.125	0.462	0.508	-0.776	-0.884	**10**
54	0.489	0.893	1.163	0.464	0.462	0.984	1.067	0.852	**10**

55	0.185	0.731	0.023	-0.410	0.120	-0.691	-1.131	-1.077	**5**
56	-0.390	0.382	-0.286	0.589	0.120	-0.683	-0.986	-0.920	**5**
57	-0.550	0.266	-0.652	-0.706	-0.222	-0.652	-1.579	-1.573	**5**
58	-0.565	0.183	-0.741	-0.831	-0.222	-0.801	-0.829	-0.727	**5**
59	-0.697	0.104	-0.848	-0.926	-0.222	-0.712	-1.333	-1.295	**5**
60	-0.196	0.449	-0.068	-0.294	0.462	-0.648	-1.162	-1.118	**5**
61	0.689	1.083	0.334	-0.340	0.120	-0.286	-0.983	-0.978	**10**
62	1.284	1.089	2.845	2.267	0.462	0.836	0.355	0.187	**10**
63	0.617	0.743	0.136	0.798	0.804	-0.013	0.700	0.741	**10**
64	0.367	0.509	-0.089	0.419	-0.222	-1.258	-1.179	-1.044	**10**
65	1.154	1.099	0.642	0.642	0.462	0.954	0.176	-0.014	**10**
66	0.383	0.818	-0.283	0.109	0.462	-0.460	-0.199	-0.106	**10**
67	0.495	0.993	0.861	0.828	0.120	-0.458	-0.635	-0.576	**10**
68	0.680	1.054	0.942	0.373	-0.564	-0.112	0.308	0.355	**10**
69	0.882	1.117	-0.073	0.135	0.462	-0.464	0.483	0.630	**10**
70	0.623	1.078	0.642	1.867	0.804	1.078	1.857	1.598	**1**
71	0.568	1.172	2.026	0.721	1.487	2.034	-1.019	-1.316	**1**
72	0.361	1.184	1.878	1.983	-0.222	1.370	2.324	1.959	**1**
73	-0.403	0.718	-0.972	-0.982	-1.590	-1.364	0.380	0.768	**1**
74	-0.767	0.455	-0.985	-0.994	-0.906	-1.689	-0.141	0.251	**1**
75	-0.958	0.128	-1.000	-0.990	-1.248	-2.200	-0.518	-0.072	**1**
76	-1.041	-0.124	-1.007	-1.004	-0.906	-1.990	-0.743	-0.395	**1**
77	-0.603	0.309	-0.719	-0.768	-2.274	0.280	2.807	2.826	**1**
78	-0.728	0.028	-0.998	-1.000	-1.590	-1.416	-0.496	-0.229	**1**
79	-0.202	0.224	-0.760	-0.370	-1.248	-0.262	3.053	3.312	**1**
80	0.701	0.623	-0.945	-0.957	-1.248	-0.734	0.505	0.728	**1**

4.5 Evaluating Correlation Matrix

The cross correlation function is a measure of the similarity between two data sets. Corresponding values of the two sets are multiplied together, and the products are summed to give the value of cross correlation without any phase lag. It is a standard method of estimating the degree to which two different series or signals are correlated. If the signals are identical, then the correlation coefficient is 1; if they are totally different, the correlation coefficient is 0, and if they are identical except that the phase is shifted by exactly 180^{o} (i.e. mirrored), then the correlation coefficient is -1. Strength of correlation between the features is evaluated based on the range shown in Table 7.

Table 7

Interpretation of the Strength of correlation results.

Correlation Coefficient Range	Strength of Correlation
|0.00 to 0.30|	Weak
|0.31 to 0.50|	Moderate
|0.51 to 0.80|	Strong
|0.81 to 1.00|	Very Strong

Correlation coefficient is also known as Pearson product-moment correlation coefficient, or "Pearson's correlation". If we have a series of n measurements of X and Y written as x_i and y_i where i = 1,2,...,n, then the sample correlation coefficient can be used to estimate the population Pearson correlation r_{xy} between X and Y. The sample correlation coefficient r_{xy} is written as:

$$\boxed{r_{xy} = \frac{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)\left(y_i - \bar{y}\right)}{(n-1)s_x s_y} = \frac{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)\left(y_i - \bar{y}\right)}{\sqrt{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2 \sum_{i=1}^{n}\left(y_i - \bar{y}\right)^2}}}$$

Where $\bar{x}$ and $\bar{y}$ are the sample means of X and Y, and s_x and s_y are the sample standard deviations of X and Y.

From the duration of different driving segment obtained from Table 1 we have defined a stress class for our database so that the correlation analysis of the database can be done properly. Here we have considered the rest period as low stress condition, highway driving as moderate stress condition and city drive as high stress condition. Neglecting all other factor towards the contribution of stress, an analysis of correlation of various physiological parameters with the three mentioned states of mental stress has been made in this paper.

E.g. the states changes from moderate to high stress where the first three minutes are in moderate stress and the subsequent three minutes in high stress condition. Instead of making this change over six minutes as 3,3,3,5,5,5 the transition is recorded as 3,3,3,4,5,5.

We have calculated the correlation between all the 8 selected statistical features with respect to different stress class values as stated in Case 1, Case 2 and Case 3. These category and stress class are shown in Table 8. Case 1 and Case 2 shows abrupt change in stress level while Case 3 shows a gradual change in stress level of driver with respect to change driving segment.

Table 8

Driving Segments divided into different Category and Gradual Stress Class.

Driving Segments Used	Category	Stress Class		
		Case 1	Case 2	Case 3
Initial Rest + Final Rest	Relaxed or Low Stress (LS)	**1**	**1**	**1**
Transition State	-	**-**	**-**	**3**
City 1 + City 2 + City 3	High Stress (HS)	**10**	**5**	**5**
Transition State	-	**-**	**-**	**4**
Highway 1 + Highway 2	Medium Stress (MS)	**5**	**3**	**3**

We evaluated Pearson's coefficient r for each drive's dataset. These r values show the strength of relationship between the different features and the stress class value of automobile driver defined in this research. Correlation matrix obtained for all Drives for three different cases are shown in Table 9 to Table 18 along with its corresponding r values.

Table 9

Correlation matrix obtained for Drive05 and its corresponding *r* values including all three cases Case 1, Case 2 and Case 3.

Drive05		Mean Hand GSR	Mean Foot GSR	Mean EMG	rms EMG	RR	Mean HR	HRV	NHRV
Mean Hand GSR		1							
Mean Foot GSR		0.966	1						
Mean EMG		0.624	0.697	1					
rms EMG		0.612	0.673	0.975	1				
RR		0.509	0.570	0.493	0.465	1			
Mean HR		0.836	0.860	0.799	0.805	0.519	1		
HRV		0.342	0.279	0.166	0.211	0.017	0.299	1	
NHRV		0.253	0.201	0.134	0.170	-0.001	0.179	0.888	1
Stress Class	Case 1	0.762	0.789	0.724	0.706	0.567	0.812	0.170	0.107
Stress Class	Case 2	0.770	0.800	0.723	0.704	0.579	0.815	0.171	0.110
Stress Class	Case 3	0.802	0.826	0.731	0.713	0.568	0.831	0.178	0.112

Table 10

Correlation matrix obtained for Drive06 and its corresponding *r* values including all three cases Case 1, Case 2 and Case 3.

Drive06		Mean Hand GSR	Mean Foot GSR	Mean EMG	rms EMG	RR	Mean HR	HRV	NHRV
Mean Hand GSR		1							
Mean Foot GSR		0.472	1						
Mean EMG		0.419	0.285	1					
rms EMG		0.379	0.312	0.971	1				
RR		-0.039	0.287	0.353	0.392	1			
Mean HR		0.413	0.684	0.650	0.661	0.602	1		
HRV		0.189	-0.026	0.394	0.355	-0.096	0.086	1	
NHRV		0.047	-0.265	0.119	0.084	-0.328	-0.287	0.919	1
Stress Class	Case 1	0.291	0.469	0.491	0.514	0.637	0.709	0.078	-0.180
Stress Class	Case 2	0.263	0.460	0.490	0.515	0.651	0.711	0.079	-0.181
Stress Class	Case 3	0.257	0.459	0.522	0.543	0.662	0.714	0.087	-0.177

Table 11

Correlation matrix obtained for Drive07 and its corresponding r values including all three cases Case 1, Case 2 and Case 3.

Drive07		Mean Hand GSR	Mean Foot GSR	Mean EMG	rms EMG	RR	Mean HR	HRV	NHRV
Mean Hand GSR		1							
Mean Foot GSR		0.821	1						
Mean EMG		0.677	0.523	1					
rms EMG		0.560	0.478	0.810	1				
RR		0.783	0.603	0.685	0.561	1			
Mean HR		0.673	0.353	0.656	0.449	0.623	1		
HRV		0.066	-0.055	0.114	0.081	-0.166	0.296	1	
NHRV		-0.085	-0.149	-0.012	-0.010	-0.306	0.093	0.977	1
Stress Class	Case 1	0.614	0.417	0.500	0.433	0.552	0.346	-0.108	-0.182
Stress Class	Case 2	0.622	0.432	0.501	0.437	0.571	0.343	-0.125	-0.201
Stress Class	Case 3	0.667	0.482	0.536	0.509	0.617	0.387	-0.078	-0.163

Table 12

Correlation matrix obtained for Drive08 and its corresponding r values including all three cases Case 1, Case 2 and Case 3.

Drive08		Mean Hand GSR	Mean Foot GSR	Mean EMG	rms EMG	RR	Mean HR	HRV	NHRV
Mean Hand GSR		1							
Mean Foot GSR		0.921	1						
Mean EMG		0.385	0.390	1					
rms EMG		0.331	0.311	0.972	1				
RR		0.450	0.591	0.218	0.114	1			
Mean HR		0.334	0.366	0.613	0.613	0.160	1		
HRV		0.158	0.210	0.445	0.442	0.083	0.876	1	
NHRV		0.137	0.196	0.337	0.326	0.101	0.698	0.942	1
Stress Class	Case 1	0.745	0.833	0.443	0.412	0.595	0.459	0.323	0.316
Stress Class	Case 2	0.753	0.846	0.449	0.416	0.608	0.456	0.320	0.314
Stress Class	Case 3	0.772	0.861	0.427	0.386	0.596	0.466	0.328	0.323

Table 13

Correlation matrix obtained for Drive09 and its corresponding *r* values including all three cases Case 1, Case 2 and Case 3.

Drive09		Mean Hand GSR	Mean Foot GSR	Mean EMG	rms EMG	RR	Mean HR	HRV	NHRV
Mean Hand GSR		1							
Mean Foot GSR		0.959	1						
Mean EMG		0.350	0.407	1					
rms EMG		0.337	0.378	0.908	1				
RR		0.451	0.474	0.005	-0.039	1			
Mean HR		0.536	0.531	0.457	0.490	0.192	1		
HRV		0.212	0.218	0.473	0.506	0.066	0.631	1	
NHRV		0.217	0.226	0.463	0.491	0.064	0.605	0.997	1
Stress Class	Case 1	0.848	0.812	0.237	0.228	0.462	0.511	0.138	0.130
Stress Class	Case 2	0.857	0.829	0.264	0.248	0.472	0.524	0.150	0.143
Stress Class	Case 3	0.861	0.842	0.252	0.234	0.483	0.505	0.158	0.154

Table 14

Correlation matrix obtained for Drive10 and its corresponding *r* values including all three cases Case 1, Case 2 and Case 3.

Drive10		Mean Hand GSR	Mean Foot GSR	Mean EMG	rms EMG	RR	Mean HR	HRV	NHRV
Mean Hand GSR		1							
Mean Foot GSR		0.841	1						
Mean EMG		0.552	0.679	1					
rms EMG		0.535	0.675	0.955	1				
RR		-0.021	0.344	0.403	0.406	1			
Mean HR		0.741	0.747	0.706	0.670	0.305	1		
HRV		0.496	0.364	0.348	0.316	0.109	0.588	1	
NHRV		0.399	0.247	0.231	0.214	0.032	0.425	0.975	1
Stress Class	Case 1	0.329	0.646	0.547	0.491	0.462	0.487	0.184	0.104
Stress Class	Case 2	0.323	0.650	0.548	0.495	0.486	0.486	0.177	0.096
Stress Class	Case 3	0.357	0.696	0.545	0.492	0.496	0.477	0.156	0.072

Table 15

Correlation matrix obtained for Drive11 and its corresponding *r* values including all three cases Case 1, Case 2 and Case 3.

Drive11		Mean Hand GSR	Mean Foot GSR	Mean EMG	rms EMG	RR	Mean HR	HRV	NHRV
Mean Hand GSR		1							
Mean Foot GSR		0.878	1						
Mean EMG		0.225	0.266	1					
rms EMG		0.174	0.196	0.964	1				
RR		0.143	0.024	-0.227	-0.194	1			
Mean HR		0.723	0.672	0.326	0.310	0.012	1		
HRV		0.372	0.365	0.173	0.184	-0.021	0.853	1	
NHRV		0.334	0.334	0.159	0.176	-0.012	0.816	0.991	1
Stress Class	Case 1	0.860	0.734	0.232	0.188	0.184	0.645	0.312	0.282
Stress Class	Case 2	0.869	0.742	0.227	0.184	0.202	0.656	0.324	0.293
Stress Class	Case 3	0.881	0.737	0.178	0.139	0.176	0.665	0.338	0.300

Table 16

Correlation matrix obtained for Drive12 and its corresponding *r* values including all three cases Case 1, Case 2 and Case 3.

Drive12		Mean Hand GSR	Mean Foot GSR	Mean EMG	rms EMG	RR	Mean HR	HRV	NHRV
Mean Hand GSR		1							
Mean Foot GSR		0.240	1						
Mean EMG		0.445	-0.129	1					
rms EMG		0.428	-0.097	0.991	1				
RR		0.352	0.019	0.263	0.206	1			
Mean HR		0.838	0.216	0.456	0.437	0.265	1		
HRV		0.334	0.097	0.030	0.045	-0.290	0.362	1	
NHRV		0.187	0.060	-0.054	-0.035	-0.355	0.179	0.980	1
Stress Class	Case 1	0.570	0.328	0.234	0.225	0.149	0.498	0.192	0.098
Stress Class	Case 2	0.586	0.331	0.249	0.239	0.162	0.515	0.185	0.088
Stress Class	Case 3	0.610	0.311	0.259	0.250	0.164	0.519	0.200	0.103

Table 17

Correlation matrix obtained for Drive15 and its corresponding *r* values including all three cases Case 1, Case 2 and Case 3.

Drive15		Mean Hand GSR	Mean Foot GSR	Mean EMG	rms EMG	RR	Mean HR	HRV	NHRV
Mean Hand GSR		1							
Mean Foot GSR		0.797	1						
Mean EMG		0.533	0.498	1					
rms EMG		0.594	0.556	0.938	1				
RR		0.436	0.370	0.336	0.328	1			
Mean HR		0.317	0.312	0.166	0.214	-0.161	1		
HRV		-0.227	-0.067	-0.151	-0.119	-0.287	0.736	1	
NHRV		-0.244	-0.087	-0.168	-0.136	-0.279	0.706	0.994	1
Stress Class	Case 1	0.729	0.739	0.461	0.539	0.553	0.160	-0.180	-0.180
Stress Class	Case 2	0.737	0.748	0.479	0.553	0.568	0.156	-0.189	-0.188
Stress Class	Case 3	0.738	0.751	0.499	0.552	0.584	0.129	-0.224	-0.226

Table 18

Correlation matrix obtained for Drive16 and its corresponding *r* values including all three cases Case 1, Case 2 and Case 3.

Drive16		Mean Hand GSR	Mean Foot GSR	Mean EMG	rms EMG	RR	Mean HR	HRV	NHRV
Mean Hand GSR		1							
Mean Foot GSR		0.789	1						
Mean EMG		0.608	0.359	1					
rms EMG		0.620	0.329	0.971	1				
RR		0.607	0.614	0.214	0.183	1			
Mean HR		0.718	0.701	0.499	0.494	0.756	1		
HRV		0.544	0.453	0.348	0.356	0.251	0.482	1	
NHRV		0.435	0.335	0.251	0.263	0.110	0.302	0.979	1
Stress Class	Case 1	0.793	0.729	0.418	0.446	0.702	0.751	0.489	0.377
Stress Class	Case 2	0.803	0.737	0.414	0.442	0.713	0.758	0.490	0.377
Stress Class	Case 3	0.804	0.777	0.385	0.410	0.709	0.768	0.514	0.399

We have evaluated Correlation Matrix for all ten driver's data set and found that in most of the cases the Stress Class is strongly correlated with Mean Hand GSR. The correlation results of Stress Class with Mean EMG, rms EMG, RR, Mean HR also shows moderate to strong correlation but not as strong as Mean Hand GSR. On inspecting last three rows of Table 9 to Table 18, it is clear that Case 3 has stronger correlation of almost all physiological parameters with stress class. This makes Case 3 for stress class a probable candidate to be verified for the remaining drives as well. Table 19 summarises all the correlation coefficient values obtained between Stress Class and Mean Hand GSR in Case 1, Case 2, and Case 3 for all the 10 drives.

Table 19

Comparison of correlation coefficient obtained between Stress Class and Mean Hand GSR in Case 1, Case 2, and Case 3 for all 10 drives.

S No.	Drive No.	Correlation Coefficient between Stress Class and Mean Hand GSR		
		Case 1 (1-5-10)	Case 2 (1-3-5)	Case 3 (1~3~5)
1	Drive05	0.762	0.770	0.802
2	Drive06	0.291	0.263	0.257
3	Drive07	0.614	0.622	0.667
4	Drive08	0.745	0.753	0.771
5	Drive09	0.848	0.857	0.860
6	Drive10	0.329	0.323	0.357
7	Drive11	0.860	0.869	0.881
8	Drive12	0.570	0.586	0.609
9	Drive15	0.729	0.737	0.738
10	Drive16	0.793	0.803	0.804

The reason behind this very strong correlation between Stress and Mean Hand GSR is that the Autonomic Nervous System (ANS) of the human body get activated when people experience any stressful situation like driving in high traffic condition. In case of physical arousal sweat is produced in the eccrine glands, which measurably changes the conductivity of the skin. The sweat glands used for measurement are typically those in the palms of the hand or the soles of the feet. Thus, GSR is the most important indicator of stress. Also, GSR is one of the most straight-forward and low-cost psycho-physiological measures. Further, choice of Case 3 is most suitable way to classify the level of stress and hence will be used in the subsequent part of our study.

4.5.1 Conclusion of Correlation Analysis

We have divided the stress class in 3 different ways: Case 1, Case 2 and Case 3 based on the change in traffic condition. Case 1 and Case 2 includes step change in traffic condition while Case 3 includes gradual change in traffic condition. 9 drive's out of total 10 drive's, shows that the correlation results obtained in Case 3 was always greater than Case 1 and Case 2 as shown in Fig. 6. This is in contrast that when a driver goes from high traffic condition to low traffic condition or vice versa, there is a gradual change in stress level along with the physiological measures of the automobile drivers.

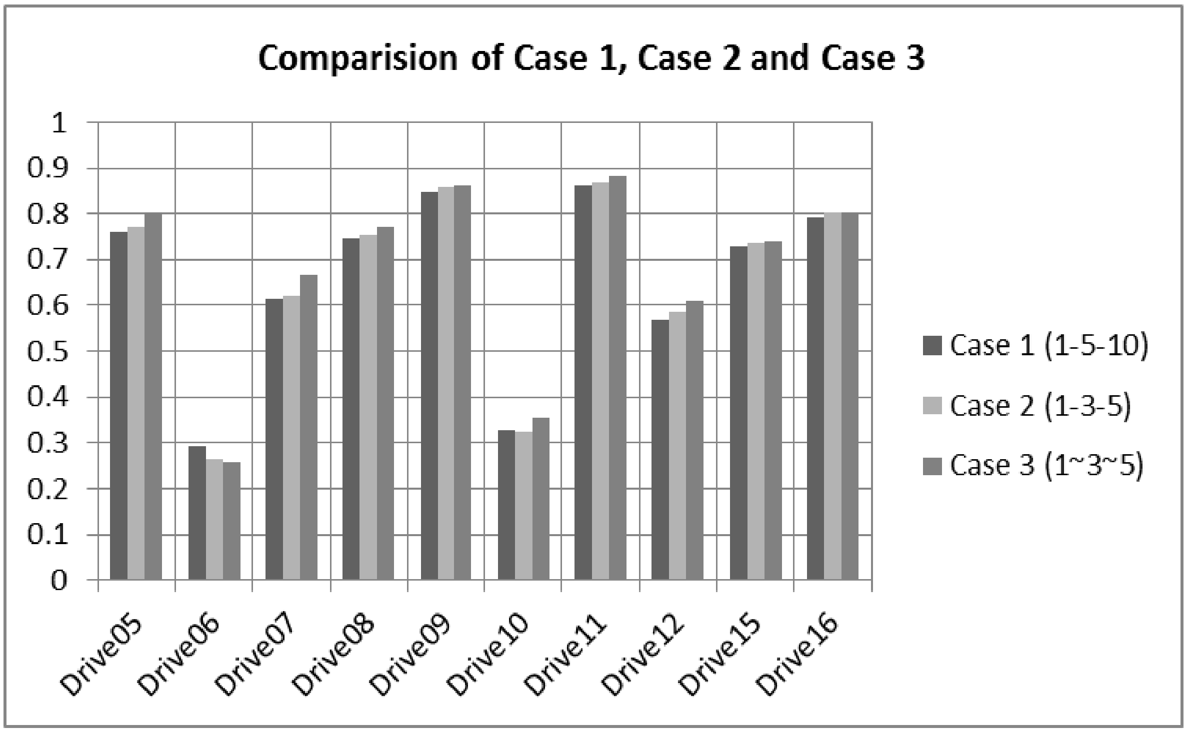

Fig. 6. Comparision of Case 1, Case 2 and Case 3 for choice of stress classification.

Assuming that the psychological stress in driver is only on account of the traffic conditions which again depend on the terrain, correlation analysis shows very strong linear relationship between Stress and Mean Hand GSR. Stress class as indicated by Case 3 shows the best correlation result, thus we select Case 3 as our final stress class for further analysis.

4.6 Feature Selection

As many pattern recognition techniques were originally not designed to cope with large amounts of irrelevant features, combining them with Feature Selection techniques has become a necessity in many applications. The objectives of feature selection are manifold, the most important ones being:

a) to avoid overfitting and improve model performance, i.e. prediction performance in the case of supervised classification and better cluster detection in the case of clustering,
b) to provide faster and more cost-effective models, and
c) to gain a deeper insight into the underlying processes that generated the data.

For selecting the best features out of 8 statistical features we used software called 'WEKA' [28]. WEKA (Waikato Environment for Knowledge Analysis) is a data mining system developed by the University of Waikato in New Zealand that implements data mining algorithms using the JAVA language. WEKA is a state of-the-art facility for developing machine learning (ML) techniques and their application to real-world data mining problems. It is a collection of machine learning algorithms for data mining tasks. The algorithms are applied directly to a dataset. WEKA is open source software issued under General Public License. WEKA contains tools for data pre-processing, classification, regression, clustering, association rules, attribute selection and visualization. It is also well-suited for developing new machine learning schemes.

Features are specified attribute numbers before applying attribute selection analysis on our database as shown in Table 20.

Table 20

Attributes used in 'WEKA' shown by corresponding attribute no.

Attribute No.	Feature/Attribute Name
1	Mean hand GSR
2	Mean foot GSR
3	Mean EMG
4	rms EMG
5	RR
6	Mean HR
7	HRV
8	NHRV

Five different feature selection algorithms are applied on the attribute no. 1, 2, 3, 4, 5, 6, 7 and 8 as shown in Table 21 to Table 25.

Evaluator: ClassifierSubsetEval -B weka.classifiers.trees.J48 -T -H

Search: BestFirst -D 1 -N 5

Table 21

Using 'ClassifierSubsetEval' Attribute Evaluator.

S No.	Drive No.	Selected Attributes	Overall Selected
1	Drive 05	1,4,5,6	1,2,3,4,5,6
2	Drive 06	2,3,4,6	
3	Drive 07	1,3,5,6	
4	Drive 08	2	
5	Drive 09	1,2,8	
6	Drive 10	1,2,3,4,5	
7	Drive 11	1,6	
8	Drive 12	2,3,5,6	
9	Drive 15	2,3,4,5,6	
10	Drive 16	1,2,3,4,5	

Evaluator: ChiSquaredAttributeEval

Search: Ranker -T -1.7976931348623157E308 -N -1

Table 22

Using 'ChiSquaredAttributeEval' Attribute Evaluator.

S No.	Drive No.	Attributes arranged by	Overall Selected
1	Drive 05	6,4,3,2,1,5,7,8	1,2,3,4,6
2	Drive 06	1,5,6,4,3,2,7,8	
3	Drive 07	1,5,3,6,4,2,7,8	
4	Drive 08	2,1,4,3,6,5,7,8	
5	Drive 09	1,2,3,4,6,7,8,5	
6	Drive 10	5,1,3,2,4,7,6,8	
7	Drive 11	1,2,6,8,7,5,3,4	
8	Drive 12	2,1,6,4,3,5,7,8	
9	Drive 15	1,2,3,4,5,7,8,6	
10	Drive 16	1, 4,3,6,5,2,7,8	

Evaluator: CfsSubsetEval

Search: BestFirst -D 1 -N 5

Table 23

Using ‘CfsSubsetEval’ Attribute Evaluator.

S No.	Drive No.	Attributes arranged by	Overall Selected
1	Drive 05	3,6	1,2,3,4,5,6
2	Drive 06	1,5,6	
3	Drive 07	1,3,5,6	
4	Drive 08	1,2,4,6	
5	Drive 09	1,2,4,7	
6	Drive 10	1,3,5	
7	Drive 11	1,2,5,6	
8	Drive 12	1,2	
9	Drive 15	1	
10	Drive 16	1,2,3,6	

Evaluator: SVMAttributeEval -X 1 -Y 0 -Z 0 -P 1.0E-25 -T 1.0E-10 -C 1.0 -N 0

Search: Ranker -T -1.7976931348623157E308 -N -1

Table 24

Using ‘SVMAttributeEval’ Attribute Evaluator.

S No.	Drive No.	Attributes arranged by	Overall Selected
1	Drive 05	2,6,3,8,1,4,5,7	1,2,3,5,6
2	Drive 06	5,6,1,4,8,2,3,7	
3	Drive 07	5,1,4,3,2,8,7,6	
4	Drive 08	2,7,5,6,8,4,1,3	
5	Drive 09	2,1,3,6,5,4,8,7	
6	Drive 10	5,3,2,6,1,4,8,7	
7	Drive 11	1,3,5,2,4,8,6,7	
8	Drive 12	1,6,2,8,5,3,7,4	
9	Drive 15	2,1,6,3,5,4,7,8	
10	Drive 16	6,4,5,2,3,7,1,8	

Evaluator: WrapperSubsetEval -B weka.classifiers.bayes.NaiveBayes -F 5 -T 0.01 -R 1

Search: GreedyStepwise -T -1.7976931348623157E308 -N -1

Table 25

Using 'WrapperSubsetEval' Attribute Evaluator.

S No.	Drive No.	Attributes arranged by	Overall Selected
1	Drive 05	1,2,3,4,6	1,2,3,5,6
2	Drive 06	1,5	
3	Drive 07	1,8	
4	Drive 08	3	
5	Drive 09	1,2,3	
6	Drive 10	1,2,3,5	
7	Drive 11	1,5	
8	Drive 12	2,6	
9	Drive 15	1,2,6	
10	Drive 16	5,6	

After applying feature selection method the results obtained shows that attribute no. 1, 2, 3, 4, 5, 6 are the attributes selected by different algorithms as shown in Table 26.

Table 26

Results obtained after using different feature selection algorithms.

S No.	Drive No.	Attribute Selector Used	Selected Attributes	Overall Selected Attributes
1	All Drive	ClassifierSubsetEval	1,2,3,4,5,6	1,2,3,4,5,6
2	All Drive	ChiSquaredAttributeEval	1,2,3,4,6	
3	All Drive	CfsSubsetEval	1,2,3,4,5,6	
4	All Drive	SVMAttributeEval	1,2,3,5,6	
5	All Drive	WrapperSubsetEval	1,2,3,5,6	

Out of 8 statistical features we selected only 6 features as shown in Table 27 for our further research.

Table 27

Final attributes selected using different algorithms in 'WEKA' Software.

Attribute No.	Feature/Attribute Name
1	Mean Hand GSR
2	Mean Foot GSR
3	Mean EMG
4	rms EMG
5	RR
6	Mean HR

On applying correlation analysis on the 6 selected statistical features, we observed that in most of the cases the correlation between Mean Hand GSR vs Mean Foot GSR and Mean EMG vs rms EMG are very strong as shown in Table 28. Thus, using Mean EMG instead of rms EMG will not make much difference in our aim to form a function that will act as a stress level indicator.

Table 28

Correlation coefficients obtained between Mean Hand GSR vs Mean Foot GSR and Mean EMG vs rms EMG for all ten Drives.

S No.	Driver's data set No.	Corresponding r values between	
		Mean Hand GSR vs Mean Foot GSR	Mean EMG vs rms EMG
1	Drive 05	0.97	0.97
2	Drive 06	0.47	0.97
3	Drive 07	0.82	0.81
4	Drive 08	0.92	0.97
5	Drive 09	0.96	0.91
6	Drive 10	0.84	0.96
7	Drive 11	0.88	0.96
8	Drive 12	0.24	0.99
9	Drive 15	0.80	0.94
10	Drive 16	0.79	0.97

Also shown in Table 28 is correlation between Mean Hand GSR and Mean Foot GSR. Theoretically, there should be high correlation between these two physiological parameters. In case of some drives like Drive06 and Drive12, the low correlation may be due to some data acquisition problems. These drives thus may need special techniques for stress classification. This is elaborated in the later sections of this paper.

4.7 Formation of Stress Detection Function

The general purpose of multiple regression (the term was first used by Pearson, 1908) is to learn more about the relationship between several independent or predictor variables and a dependent or criterion variable. For example, a real estate agent might record for each listing the size of the house (in square feet), the number of bedrooms, the average income in the respective neighborhood according to census data, and a subjective rating of appeal of the house. Once this information has been compiled for various houses it would be interesting to see whether and how these measures relate to the price for which a house is sold. For example, you might learn that the number of bedrooms is a better predictor of the price for which a house sells in a particular neighborhood than how "pretty" the house is (subjective rating). You may also detect "outliers," that is, houses that should really sell for more, given their location and characteristics.

E.g. personnel professionals customarily use multiple regression procedures to determine equitable compensation. You can determine a number of factors or dimensions such as "amount of responsibility" (Resp) or "number of people to supervise" (No_Super) that you believe to contribute to the value of a job. The personnel analyst then usually conducts a salary survey among comparable companies in the market, recording the salaries and respective characteristics (i.e., values on dimensions) for different positions. This information can be used in a multiple regression analysis to build a regression equation of the form:

Salary = 0.5*Resp + 0.8*No_Super

Once this so-called regression line has been determined, the analyst can now easily construct a graph of the expected (predicted) salaries and the actual salaries of job incumbents in his or her company. Thus, the analyst is able to determine which position is underpaid (below the regression line) or overpaid (above the regression line), or paid equitably.

In the social and natural sciences multiple regression procedures are very widely used in research. In general, multiple regression allows the researcher to ask (and hopefully answer) the general question “What is the best predictor of ......?”. For example, educational researchers might want to learn “What are the best predictors of success in high-school ?”. Psychologists may want to determine which personality variable best predicts social adjustment. Sociologists may want to find out which of the multiple social indicators best predict whether or not a new immigrant group will adapt and be absorbed into society.

In statistics, regression analysis is a statistical technique for estimating the relationships among variables. It includes many techniques for modelling and analyzing several variables, when the focus is on the relationship between a dependent variable and one or more independent variables. More specifically, regression analysis helps one understand how the typical value of the dependent variable changes when any one of the independent variables is varied, while the other independent variables are held fixed. In all cases, the estimation target is a function of the independent variables called the regression function. In regression analysis, it is also of interest to characterize the variation of the dependent variable around the regression function, which can be described by a probability distribution.

Regression analysis is widely used for prediction and forecasting, where its use has substantial overlap with the field of machine learning. Regression analysis is also used to understand which among the independent variables are related to the dependent variable, and to explore the forms of these relationships.

We applied Regression Analysis by taking Stress Class of automobile driver as Output variable and Mean Hand GSR, Mean EMG, RR, Mean HR and Mean Foot GSR as Input variables as shown in Table 29.

Table 29
Variables used in Regression analysis.

Inputs (X)	**Output (Y)**
1. Mean Hand GSR ($\mathbf{X_1}$) 2. Mean EMG ($\mathbf{X_2}$) 3. RR ($\mathbf{X_3}$) 4. Mean HR ($\mathbf{X_4}$) 5. Mean Foot GSR ($\mathbf{X_5}$)	Stress Class (**LS, MS and HS**)

We have divided our analysis into 4 Cases:

1. **Case 1** includes Stress Class as Output variable and Mean Hand GSR, Mean EMG, RR and Mean HR as Input variables.
2. **Case 2** includes Stress Class as Output variable and Mean Hand GSR, RR and Mean HR as Input variables.
3. **Case 3** includes Stress Class as Output variable and Mean Foot GSR, RR and Mean HR as Input variables.
4. **Case 4** includes Stress Class as Output variable and Mean Hand GSR, Mean Foot GSR, RR and Mean HR as Input variables.

The results were computed on MS Excel and the coefficients of function obtained from Case 1 are shown in Table 30. Similarly, Case 2, Case 3 and Case 4 have been analyzed and corresponding functions are formed as shown in Table 31.

Table 30

Results obtained from Regression Analysis of Case 1.

S No.	**Driver's data set No.**	**Coefficients**				**Intercept**
		Mean Hand GSR (X_1)	**Mean EMG (X_2)**	**RR (X_3)**	**Mean HR (X_4)**	
1	Drive 05	0.6707	0.3482	0.2471	0.4868	3.3000
2	Drive 06	0.1276	0.1737	0.7271	0.6647	3.1500
3	Drive 07	0.9273	0.2820	0.3728	-0.3789	3.2500
4	Drive 08	1.0022	-0.0488	0.5814	0.4096	3.2000
5	Drive 09	1.2903	-0.1174	0.1910	0.1554	3.6286
6	Drive 10	0.3559	0.4445	0.7108	0.0614	3.0750
7	Drive 11	1.4506	-0.1217	0.0900	0.1633	3.1000
8	Drive 12	1.0239	-0.0340	-0.0845	0.1104	3.0750
9	Drive 15	0.9763	0.1877	0.5500	-0.0535	3.0260
10	Drive 16	0.9580	-0.3615	0.2602	0.5659	3.5938
MEAN		**0.8783**	**0.0753**	**0.3646**	**0.2185**	**3.2398**

The results obtained in Table 6, shows that Mean EMG has very less impact on change in value of stress function ($Ƒ_1$). Thus, Mean EMG has been neglected in Case 2 to find the function ($Ƒ_2$). We observed that there is a negligible change in value of correlation between Stress Class and function ($Ƒ_2$) with respect to value of correlation between Stress Class and function ($Ƒ_1$). Functions ($Ƒ_3$ & $Ƒ_4$) obtained in Case 3 and Case 4 also show very strong correlation with Stress Class. Table 32 shows a comparison of the correlations obtained between all the four functions formed with Stress Class.

Table 31

Function obtained from Regression Analysis of Case 1, Case 2, Case 3, and Case 4.

Case No.	Function Formed
Case 1	**Function ($Ƒ_1$) = 0.8783(X_1) + 0.0753(X_2) + 0.3646(X_3) + 0.2185(X_4) + 3.2398**
Case 2	**Function ($Ƒ_2$) = 0.8783(X_1) + 0.3646(X_3) + 0.2185(X_4) + 3.2398**
Case 3	**Function ($Ƒ_3$) = 0.7907(X_5) + 0.4184(X_3) + 0.3587(X_4) + 3.2398**
Case 4	**Function ($Ƒ_4$) = 0.4916(X_1) + 0.5257(X_5) +0.2851(X_3) + 0.2007(X_4) + 3.2398**

Table 32

Comparison of Correlation between Stress Class and Different Functions for All Drives.

S No.	Driver's data set No.	Correlation Coefficients				
		Stress Class vs Mean Hand GSR	Case 1	Case 2	Case 3	Case 4
			Stress Class vs Function ($Ƒ_1$)	Stress Class vs Function ($Ƒ_2$)	Stress Class vs Function ($Ƒ_3$)	Stress Class vs Function ($Ƒ_4$)
1	Drive 5	0.813	0.860	0.853	0.862	**0.862**
2	Drive 6	0.256	0.593	0.582	**0.704**	0.636
3	Drive 7	**0.680**	0.677	0.675	0.602	0.642
4	Drive 8	0.783	0.857	0.858	**0.887**	0.885
5	Drive 9	0.856	0.850	0.853	0.830	**0.860**
6	Drive 10	0.359	0.550	0.540	**0.731**	0.643
7	Drive 11	**0.877**	0.852	0.854	0.776	0.840
8	Drive 12	**0.609**	0.572	0.573	0.471	0.584
9	Drive 15	0.736	0.774	0.770	0.772	**0.802**
10	Drive 16	0.782	0.838	0.846	0.850	**0.864**
Mean Correlation.		**0.6751**	**0.7423**	**0.7404**	**0.7485**	**0.7618**
Score		**3**	**0**	**0**	**3**	**4**

Stress is most strongly related to the Mean Hand GSR feature of the automobile driver. There are several other physiological features like Mean EMG, RR, Mean HR, Mean Foot GSR that are also affected by the dynamic stress level of the automobile drivers [29][30]. Thus we have formulated a stress function (Ƒ) which combines important features drawn from different sensors in specific proportion. The function (Ƒ) obtained have higher degree of strength of correlation with Stress Class as compared to just Mean Hand GSR feature. Changing the features used for the formulation of function (Ƒ), changes value of the correlation coefficient. For each drive the best correlation function has been highlighted in Table 32. The score evaluated in the last row of this table is the count of the best correlations.

Function ($Ƒ_4$) has the best correlation results with the stress class as indicated by the overall Mean Correlation and Score value. Thus, instead of Mean Hand GSR, the function ($Ƒ_4$) can directly be used for the detection of dynamic stress level of an automobile driver continuously over a period of time. In our further research of stress classification we have used only function ($Ƒ_4$) as a standard function for determination of stress level.

Function ($Ƒ_4$) = $0.4916(X_1) + 0.5257(X_5) + 0.2851(X_3) + 0.2007(X_4) + 3.2398$

Stress Function ($Ƒ_4$) can directly be used for the detection of dynamic stress level of an automobile driver continuously over a period of time.

4.8 Comparison of Features after each Analysis

Our aim was to select the best possible features extracted from physiological parameters. Hence, we have achieved our target of feature selection and function formation. Table 33 illustrates the comparison of features available after each Analysis.

Table 33

Comparison of Features after every Analysis.

Initial Features	Features after Attribute Selection using 'WEKA'	Features after Correlation Analysis	Features after Function Formation
Mean Hand GSR Mean Foot GSR Mean EMG rms EMG RR Mean HR HRV NHRV	Mean Hand GSR Mean Foot GSR Mean EMG rms EMG RR Mean HR	Mean Hand GSR Mean EMG RR Mean HR	Function (Ƒ$_4$)

Chapter 5
STRESS DETECTION

5.1 Category and Class Formation

In order to make evaluation for 3 types of driver stress corresponding to 3 types of driving conditions, all the segment based statistical parameters related to the Hand GSR, Foot GSR, EMG, RR and HR arrays have been collected for the complete rest, city driving and highway driving periods of each driving. For this aim, the data arrays related to the 2 segments of the rest period have been added together to form the data arrays related to the complete rest (R) segment of each drive corresponding to the low level stress conditions. Similarly, the data arrays related to the 2 segments of highway driving periods have been added together to form the data arrays related to the complete highway (H) segment of each drive corresponding to the medium level stress conditions, and the data arrays related the 3 segments of the city driving periods have been added together to form the data arrays related to the complete city (C) segment of each drive corresponding to the high level stress conditions [20].

The data acquired from **Healey and Picard's** [2] experiment lacks the information regarding the duration of each Rest, City and Highway driving task, but the same durations were found in **Ahmet Akbas** [20] research and we found it to be very useful in our research for stress detection in automobile drivers. List of the time intervals for different driving segment for each of the 10 selected Drive's i.e. Drive 05, 06, 07, 08, 09, 10, 11, 12, 15 and 16 are shown in Table 34.

Table 34

Time intervals of the 7 driving segments of available bio-signal datasets.

Drive No.	Driving period (min)							Total rec. time (min)
	Initial Rest	City 1	Highway 1	City 2	Highway 2	City 3	Final Rest	
Drive05	15.13	16	7.74	6.06	7.56	14.96	15.78	83.23
Drive06	15.05	14.49	7.32	6.53	7.64	12.29	15.05	78.38
Drive07	15.04	16.23	10.96	9.83	7.64	10.15	15.03	84.87
Drive08	15	12.31	7.23	9.51	7.64	13.43	15.07	80.19
Drive09	15.66	19.21	8.47	5.2	7.06	13.21	NA	68.82
Drive10	15.04	15.3	8.66	5.27	7.04	12.06	14.79	78.15
Drive11	15.02	15.81	7.43	7.15	6.96	11.72	14.99	79.08
Drive12	15.01	13.41	7.56	6.5	8.06	11.68	15.01	77.23
Drive15	15	12.54	7.24	5.99	6.82	12.12	15	74.7
Drive16	15.01	16.12	7.14	5.12	6.81	13.91	NA	64.1

NA: Not available

5.2 Stress Classification

Classification is a task of training a model that maps each attribute set X to one of the predefined class labels Y. In our case class labels Y is the Stress Class of automobile driver defined by different traffic conditions (Resting, Highway and City) and attribute set X is the function ($Ƒ_4$) which is a linear combination of four statistical features namely Mean Hand GSR, Mean Foot GSR, Respiration Rate and Mean HR extracted from the physiological parameters. Function ($Ƒ_4$) is the only attribute used for stress classification because this function is alone sufficient for detection of stress level of the automobile driver. Resting phase of the driving segment is considered as Low Stress (LS) category. Highway phase is considered as Medium Stress (MS) category, while driving in City is considered as High Stress (HS) category as shown in Table 35. It is assumed that the change in stress level of the driver is solely due to the variable traffic conditions and all other factor towards the contribution of stress is neglected.

Table 35
Driving Segment defined by Category and Stress Class.

Driving Segments Used	Category	Stress Class
Initial Rest + Final Rest	Relaxed or Low Stress	**LS (1)**
City 1 + City 2 + City 3	High Stress	**HS (5)**
Highway 1 + Highway 2	Moderate Stress	**MS (3)**

In our research the classification of three stress class is done by the use of two threshold values. One threshold value differentiates Low Stress category and Moderate Stress category while other differentiates Moderate Stress category and High Stress category.

E.g. Stress Class LS has min value of 1 and max value of 2, HS has min value of 2 and max value of 3, while HS has min value of 3 and max value of 5. Clearly the two threshold values which classify the stress class into three categories (LS, MS and HS) would be 2 and 3. Similarly, the classifications of all the ten Drives are carried out by using a common threshold value and also by changing the threshold value with respect to individual Drive. Thresholds based classification of all ten drives as shown from Table 36 to Table 45 and summary of all threshold classification is given in Table 46.

Table 36
Threshold based classification of Stress Class using Function for Drive05.

S No.	Function ($Ƒ_4$)	Actual Stress Class	Predicted Stress Class	Correctly Classified in its class (TRUE/FALSE)
1	3.482559	1	3	FALSE
2	3.064901	1	1	TRUE
3	2.21291	1	1	TRUE
4	1.558975	1	1	TRUE
5	1.473714	1	1	TRUE
6	1.547599	1	1	TRUE
7	1.464676	1	1	TRUE
8	1.567954	1	1	TRUE
9	1.734413	1	1	TRUE
10	1.719835	1	1	TRUE
11	1.636098	1	1	TRUE
12	1.622905	1	1	TRUE
13	1.507161	1	1	TRUE
14	1.618369	1	1	TRUE
15	1.652224	1	1	TRUE
16	1.624481	1	1	TRUE
17	1.693111	1	1	TRUE
18	1.308608	1	1	TRUE
19	1.255644	1	1	TRUE
20	1.019267	1	1	TRUE
21	1.044265	1	1	TRUE
22	1.298794	1	1	TRUE
23	1.610025	1	1	TRUE
24	1.619114	1	1	TRUE
25	1.459516	1	1	TRUE
26	1.252513	1	1	TRUE
27	1.638183	1	1	TRUE
28	3.910174	3	5	FALSE
29	4.126007	3	5	FALSE
30	4.301584	3	5	FALSE
31	3.880595	3	5	FALSE
32	3.95461	3	5	FALSE
33	3.706982	3	5	FALSE
34	3.477979	3	3	TRUE
35	3.471411	3	3	TRUE
36	3.21851	3	3	TRUE
37	3.407138	3	3	TRUE
38	3.279052	3	3	TRUE
39	3.139869	3	3	TRUE
40	3.28969	3	3	TRUE

41	3.24883	3	3	TRUE
42	1.596138	5	1	FALSE
43	3.1906	5	3	FALSE
44	3.231671	5	3	FALSE
45	3.314984	5	3	FALSE
46	3.375733	5	3	FALSE
47	3.344031	5	3	FALSE
48	3.03236	5	1	FALSE
49	3.423655	5	3	FALSE
50	3.043019	5	1	FALSE
51	5.470108	5	5	TRUE
52	5.836192	5	5	TRUE
53	6.548755	5	5	TRUE
54	5.117903	5	5	TRUE
55	4.873737	5	5	TRUE
56	4.485559	5	5	TRUE
57	4.197468	5	5	TRUE
58	3.769316	5	5	TRUE
59	4.325054	5	5	TRUE
60	3.824683	5	5	TRUE
61	4.608513	5	5	TRUE
62	4.611346	5	5	TRUE
63	3.807737	5	5	TRUE
64	4.241073	5	5	TRUE
65	5.239767	5	5	TRUE
66	4.465891	5	5	TRUE
67	3.968649	5	5	TRUE
68	4.451004	5	5	TRUE
69	4.2354	5	5	TRUE
70	4.243242	5	5	TRUE
71	4.527368	5	5	TRUE
72	4.267737	5	5	TRUE
73	4.241573	5	5	TRUE
74	4.019427	5	5	TRUE
75	3.886469	5	5	TRUE
76	3.685692	5	5	TRUE
77	5.885593	5	5	TRUE
78	4.780615	5	5	TRUE
79	4.699497	5	5	TRUE
80	4.216192	5	5	TRUE
Correctly Classified Instances in %age				**(80-16)/80=80%**

Table 37
Threshold based classification of Stress Class using Function for Drive06.

S No.	Function (F_4)	Actual Stress Class	Predicted Stress Class	Correctly Classified in its class (TRUE/FALSE)
1	5.267335766	1	5	FALSE
2	3.485930053	1	5	FALSE
3	2.954228003	1	1	TRUE
4	2.887734543	1	1	TRUE
5	2.982020495	1	1	TRUE
6	3.06794781	1	1	TRUE
7	3.043197039	1	1	TRUE
8	3.148082931	1	1	TRUE
9	2.842920018	1	1	TRUE
10	2.925647221	1	1	TRUE
11	3.08660177	1	1	TRUE
12	2.6247487	1	1	TRUE
13	2.300937841	1	1	TRUE
14	2.117679502	1	1	TRUE
15	2.453251976	1	1	TRUE
16	2.234706156	1	1	TRUE
17	2.118383072	1	1	TRUE
18	2.370986455	1	1	TRUE
19	1.886602601	1	1	TRUE
20	1.903423539	1	1	TRUE
21	1.649220683	1	1	TRUE
22	2.225044656	1	1	TRUE
23	2.074248454	1	1	TRUE
24	1.926650369	1	1	TRUE
25	2.46487297	1	1	TRUE
26	2.14178403	1	1	TRUE
27	2.162280223	1	1	TRUE
28	1.839086913	1	1	TRUE
29	2.08057251	1	1	TRUE
30	3.818570473	3	5	FALSE
31	3.617733236	3	5	FALSE
32	2.427695483	3	1	FALSE
33	2.050839036	3	1	FALSE
34	1.830246944	3	1	FALSE
35	1.742091478	3	1	FALSE
36	2.276334203	3	1	FALSE
37	4.221102487	3	5	FALSE
38	2.128259099	3	1	FALSE
39	1.920772796	3	1	FALSE
40	2.019762541	3	1	FALSE

41	1.937833493	3	1	FALSE
42	1.980042745	3	1	FALSE
43	3.874713767	3	5	FALSE
44	2.513164991	3	1	FALSE
45	2.974308514	3	1	FALSE
46	2.814272185	5	1	FALSE
47	3.28235383	5	3	FALSE
48	2.628903236	5	1	FALSE
49	3.184231232	5	1	FALSE
50	3.282576548	5	3	FALSE
51	2.885846826	5	1	FALSE
52	3.038217692	5	1	FALSE
53	2.716555213	5	1	FALSE
54	3.469909573	5	5	TRUE
55	3.528483595	5	5	TRUE
56	6.614792864	5	5	TRUE
57	6.446958832	5	5	TRUE
58	5.256163071	5	5	TRUE
59	5.065872256	5	5	TRUE
60	5.216108016	5	5	TRUE
61	4.980181582	5	5	TRUE
62	4.809768788	5	5	TRUE
63	3.66099105	5	5	TRUE
64	4.583607922	5	5	TRUE
65	4.608331754	5	5	TRUE
66	4.195525545	5	5	TRUE
67	4.312280443	5	5	TRUE
68	4.553241609	5	5	TRUE
69	3.694414868	5	5	TRUE
70	4.211125799	5	5	TRUE
71	3.77886073	5	5	TRUE
72	4.153250374	5	5	TRUE
73	4.267968244	5	5	TRUE
74	4.279009656	5	5	TRUE
75	4.27907487	5	5	TRUE
76	3.928042338	5	5	TRUE
77	3.978551193	5	5	TRUE
78	3.89914739	5	5	TRUE
79	3.879799699	5	5	TRUE
80	4.099983588	5	5	TRUE
Correctly Classified Instances in %age				**(80-26)/80=67.5%**

Table 38
Threshold based classification of Stress Class using Function for Drive07.

S No.	Function (F_4)	Actual Stress Class	Predicted Stress Class	Correctly Classified in its class (TRUE/FALSE)
1	4.557925	1	5	FALSE
2	4.967325	1	5	FALSE
3	4.251615	1	5	FALSE
4	3.408978	1	3	FALSE
5	2.177331	1	1	TRUE
6	1.890362	1	1	TRUE
7	2.555703	1	1	TRUE
8	2.025427	1	1	TRUE
9	1.726873	1	1	TRUE
10	1.495299	1	1	TRUE
11	1.751163	1	1	TRUE
12	1.344064	1	1	TRUE
13	1.067461	1	1	TRUE
14	1.026155	1	1	TRUE
15	0.862922	1	1	TRUE
16	0.643666	1	1	TRUE
17	0.605236	1	1	TRUE
18	0.300755	1	1	TRUE
19	0.289887	1	1	TRUE
20	2.691954	1	1	TRUE
21	2.50478	1	1	TRUE
22	2.038419	1	1	TRUE
23	2.005315	1	1	TRUE
24	2.513878	1	1	TRUE
25	2.159393	1	1	TRUE
26	2.849795	1	1	TRUE
27	4.393437	3	5	FALSE
28	5.202628	3	5	FALSE
29	3.758778	3	5	FALSE
30	3.519877	3	5	FALSE
31	5.492688	3	5	FALSE
32	4.914667	3	5	FALSE
33	3.60786	3	5	FALSE
34	3.569162	3	5	FALSE
35	3.610444	3	5	FALSE
36	2.834521	3	1	FALSE
37	2.745635	3	1	FALSE
38	2.905562	3	1	FALSE
39	2.915013	3	1	FALSE
40	3.434775	3	3	TRUE

41	3.332959	3	3	TRUE
42	3.285434	3	3	TRUE
43	3.145839	3	3	TRUE
44	3.381204	3	3	TRUE
45	0.319963	5	1	FALSE
46	0.735979	5	1	FALSE
47	3.16377	5	3	FALSE
48	3.085863	5	3	FALSE
49	3.318909	5	3	FALSE
50	3.348116	5	3	FALSE
51	3.428244	5	3	FALSE
52	3.371592	5	3	FALSE
53	5.82696	5	5	TRUE
54	3.527486	5	5	TRUE
55	4.466809	5	5	TRUE
56	3.815657	5	5	TRUE
57	3.886418	5	5	TRUE
58	3.720717	5	5	TRUE
59	3.813559	5	5	TRUE
60	4.117807	5	5	TRUE
61	4.958012	5	5	TRUE
62	4.250507	5	5	TRUE
63	4.859714	5	5	TRUE
64	3.67483	5	5	TRUE
65	5.407224	5	5	TRUE
66	3.940693	5	5	TRUE
67	3.533405	5	5	TRUE
68	3.683481	5	5	TRUE
69	4.759245	5	5	TRUE
70	4.220458	5	5	TRUE
71	4.083197	5	5	TRUE
72	4.278613	5	5	TRUE
73	4.12486	5	5	TRUE
74	4.743188	5	5	TRUE
75	4.160215	5	5	TRUE
76	4.707812	5	5	TRUE
77	3.897325	5	5	TRUE
78	3.947342	5	5	TRUE
79	3.945104	5	5	TRUE
80	4.29876	5	5	TRUE
Correctly Classified Instances in %age				**(80-25)/80=68.7%**

Table 39
Threshold based classification of Stress Class using Function for Drive08.

S No.	Function (F_4)	Actual Stress Class	Predicted Stress Class	Correctly Classified in its class (TRUE/FALSE)
1	3.318008	1	3	FALSE
2	2.461616	1	1	TRUE
3	2.143356	1	1	TRUE
4	2.061727	1	1	TRUE
5	1.819801	1	1	TRUE
6	1.897946	1	1	TRUE
7	1.550949	1	1	TRUE
8	1.503412	1	1	TRUE
9	1.691239	1	1	TRUE
10	1.54106	1	1	TRUE
11	1.235978	1	1	TRUE
12	1.358849	1	1	TRUE
13	1.377647	1	1	TRUE
14	1.61738	1	1	TRUE
15	1.436204	1	1	TRUE
16	2.818296	1	1	TRUE
17	2.606372	1	1	TRUE
18	2.187832	1	1	TRUE
19	1.883521	1	1	TRUE
20	1.964001	1	1	TRUE
21	1.560829	1	1	TRUE
22	1.601508	1	1	TRUE
23	1.478996	1	1	TRUE
24	1.164686	1	1	TRUE
25	1.440288	1	1	TRUE
26	1.5395	1	1	TRUE
27	1.637875	1	1	TRUE
28	1.474864	1	1	TRUE
29	1.392807	1	1	TRUE
30	4.044443	3	5	FALSE
31	3.723587	3	3	TRUE
32	4.01571	3	5	FALSE
33	4.336484	3	5	FALSE
34	3.522835	3	3	TRUE
35	3.534886	3	3	TRUE
36	3.915035	3	5	FALSE
37	3.747374	3	3	TRUE
38	3.720873	3	3	TRUE
39	3.640704	3	3	TRUE
40	3.38248	3	3	TRUE

41	3.42358	3	3	TRUE
42	3.470033	3	3	TRUE
43	3.29618	3	3	TRUE
44	1.647228	5	1	FALSE
45	3.15513	5	3	FALSE
46	2.856788	5	1	FALSE
47	3.297067	5	3	FALSE
48	5.728108	5	5	TRUE
49	5.264902	5	5	TRUE
50	4.840299	5	5	TRUE
51	4.546413	5	5	TRUE
52	4.627613	5	5	TRUE
53	4.330221	5	5	TRUE
54	4.027881	5	5	TRUE
55	4.678588	5	5	TRUE
56	4.45749	5	5	TRUE
57	4.224961	5	5	TRUE
58	5.195035	5	5	TRUE
59	4.808199	5	5	TRUE
60	4.440198	5	5	TRUE
61	3.835877	5	3	FALSE
62	4.26711	5	5	TRUE
63	4.248087	5	5	TRUE
64	4.042109	5	5	TRUE
65	4.625104	5	5	TRUE
66	4.65569	5	5	TRUE
67	4.260383	5	5	TRUE
68	4.141656	5	5	TRUE
69	4.311772	5	5	TRUE
70	4.244553	5	5	TRUE
71	3.571393	5	3	FALSE
72	4.121942	5	5	TRUE
73	3.730537	5	3	FALSE
74	3.954809	5	5	TRUE
75	4.264776	5	5	TRUE
76	3.543526	5	3	FALSE
77	3.736714	5	3	FALSE
78	4.069051	5	5	TRUE
79	4.617779	5	5	TRUE
80	5.274259	5	5	TRUE
Correctly Classified Instances in %age				**(80-13)/80=83.7%**

Table 40
Threshold based classification of Stress Class using Function for Drive09.

S No.	Function (F_4)	Actual Stress Class	Predicted Stress Class	Correctly Classified in its class (TRUE/FALSE)
1	1.801967	1	1	TRUE
2	1.393152	1	1	TRUE
3	1.51097	1	1	TRUE
4	1.696488	1	1	TRUE
5	1.1484	1	1	TRUE
6	0.788411	1	1	TRUE
7	0.9505	1	1	TRUE
8	1.268865	1	1	TRUE
9	1.328504	1	1	TRUE
10	1.246816	1	1	TRUE
11	1.215064	1	1	TRUE
12	1.109627	1	1	TRUE
13	1.660973	1	1	TRUE
14	1.155549	1	1	TRUE
15	0.980252	1	1	TRUE
16	0.971344	1	1	TRUE
17	1.520192	1	1	TRUE
18	3.581129	3	5	FALSE
19	3.805758	3	5	FALSE
20	3.629043	3	5	FALSE
21	4.340732	3	5	FALSE
22	3.876998	3	5	FALSE
23	4.293952	3	5	FALSE
24	3.453524	3	5	FALSE
25	3.348963	3	5	FALSE
26	3.446747	3	5	FALSE
27	3.264403	3	5	FALSE
28	2.794406	3	3	TRUE
29	2.70459	3	3	TRUE
30	2.884196	3	3	TRUE
31	2.937719	5	3	FALSE
32	2.708674	5	3	FALSE
33	2.753239	5	3	FALSE
34	3.33069	5	5	TRUE
35	3.477856	5	5	TRUE
36	3.126301	5	5	TRUE
37	3.074378	5	5	TRUE
38	3.464987	5	5	TRUE
39	3.282396	5	5	TRUE
40	3.081587	5	5	TRUE

41	3.480705	5	5	TRUE
42	3.479429	5	5	TRUE
43	4.463871	5	5	TRUE
44	3.843177	5	5	TRUE
45	4.972622	5	5	TRUE
46	4.686967	5	5	TRUE
47	4.426376	5	5	TRUE
48	3.808239	5	5	TRUE
49	4.073642	5	5	TRUE
50	5.13842	5	5	TRUE
51	4.111518	5	5	TRUE
52	5.119994	5	5	TRUE
53	4.166468	5	5	TRUE
54	4.434069	5	5	TRUE
55	4.532803	5	5	TRUE
56	4.577119	5	5	TRUE
57	4.533271	5	5	TRUE
58	4.1746	5	5	TRUE
59	4.137289	5	5	TRUE
60	5.344503	5	5	TRUE
61	4.087958	5	5	TRUE
62	4.49624	5	5	TRUE
63	3.656509	5	5	TRUE
64	4.516952	5	5	TRUE
65	3.53193	5	5	TRUE
66	3.852884	5	5	TRUE
67	3.821323	5	5	TRUE
68	4.29004	5	5	TRUE
69	4.290498	5	5	TRUE
70	4.327239	5	5	TRUE
Correctly Classified Instances in %age				**(70-13)/70=81.4%**

Table 41
Threshold based classification of Stress Class using Function for Drive10.

S No.	Function (F_4)	Actual Stress Class	Predicted Stress Class	Correctly Classified in its class (TRUE/FALSE)
1	1.583434	1	1	TRUE
2	1.431123	1	1	TRUE
3	2.035463	1	1	TRUE
4	1.928792	1	1	TRUE
5	1.507846	1	1	TRUE
6	1.275416	1	1	TRUE
7	1.333964	1	1	TRUE
8	1.272451	1	1	TRUE
9	1.259638	1	1	TRUE
10	1.228792	1	1	TRUE
11	1.166337	1	1	TRUE
12	1.158397	1	1	TRUE
13	1.152302	1	1	TRUE
14	1.211575	1	1	TRUE
15	1.202804	1	1	TRUE
16	1.20832	1	1	TRUE
17	1.230303	1	1	TRUE
18	3.332533	1	5	FALSE
19	2.795129	1	3	FALSE
20	3.385365	1	5	FALSE
21	3.725339	1	5	FALSE
22	4.335343	1	5	FALSE
23	3.581389	1	5	FALSE
24	3.53981	1	5	FALSE
25	4.055735	1	5	FALSE
26	3.637363	1	5	FALSE
27	3.612237	1	5	FALSE
28	3.174754	1	5	FALSE
29	4.380632	1	5	FALSE
30	3.131962	1	5	FALSE
31	2.470673	1	1	TRUE
32	5.402645	3	5	FALSE
33	4.729493	3	5	FALSE
34	3.646135	3	5	FALSE
35	2.915597	3	3	TRUE
36	3.544633	3	5	FALSE
37	4.201473	3	5	FALSE
38	3.489529	3	5	FALSE
39	2.763633	3	3	TRUE
40	2.639463	3	3	TRUE

41	3.466707	3	5	FALSE
42	2.946937	3	3	TRUE
43	2.685105	3	3	TRUE
44	2.608187	3	3	TRUE
45	3.240541	3	5	FALSE
46	3.403867	3	5	FALSE
47	1.91808	5	1	FALSE
48	4.423891	5	5	TRUE
49	4.783643	5	5	TRUE
50	4.254879	5	5	TRUE
51	3.68685	5	5	TRUE
52	3.044846	5	5	TRUE
53	3.147019	5	5	TRUE
54	2.769501	5	3	FALSE
55	2.836109	5	3	FALSE
56	3.515077	5	5	TRUE
57	4.091062	5	5	TRUE
58	3.477361	5	5	TRUE
59	2.827764	5	3	FALSE
60	4.272514	5	5	TRUE
61	2.865443	5	3	FALSE
62	4.969774	5	5	TRUE
63	5.003853	5	5	TRUE
64	4.134195	5	5	TRUE
65	4.604792	5	5	TRUE
66	4.018684	5	5	TRUE
67	5.001	5	5	TRUE
68	3.837756	5	5	TRUE
69	5.874575	5	5	TRUE
70	4.241688	5	5	TRUE
71	4.094644	5	5	TRUE
72	3.706211	5	5	TRUE
73	3.612149	5	5	TRUE
74	3.636263	5	5	TRUE
75	3.984343	5	5	TRUE
76	3.366726	5	5	TRUE
77	3.560152	5	5	TRUE
78	4.109651	5	5	TRUE
79	5.885942	5	5	TRUE
80	5.598394	5	5	TRUE
Correctly Classified Instances in %age				**(80-27)/80=66.2%**

Table 42
Threshold based classification of Stress Class using Function for Drive11.

S No.	Function (F_4)	Actual Stress Class	Predicted Stress Class	Correctly Classified in its class (TRUE/FALSE)
1	3.970864	1	5	FALSE
2	3.384377	1	3	FALSE
3	2.253638	1	1	TRUE
4	1.656621	1	1	TRUE
5	2.138081	1	1	TRUE
6	1.215104	1	1	TRUE
7	1.393875	1	1	TRUE
8	1.328875	1	1	TRUE
9	1.80774	1	1	TRUE
10	1.703806	1	1	TRUE
11	1.898412	1	1	TRUE
12	1.712177	1	1	TRUE
13	1.4953	1	1	TRUE
14	1.353617	1	1	TRUE
15	1.412249	1	1	TRUE
16	1.498821	1	1	TRUE
17	1.517055	1	1	TRUE
18	2.800865	1	1	TRUE
19	2.633052	1	1	TRUE
20	2.404264	1	1	TRUE
21	2.190696	1	1	TRUE
22	1.932795	1	1	TRUE
23	2.302984	1	1	TRUE
24	2.062211	1	1	TRUE
25	1.881453	1	1	TRUE
26	1.53609	1	1	TRUE
27	1.507547	1	1	TRUE
28	1.632539	1	1	TRUE
29	1.513593	1	1	TRUE
30	1.587882	1	1	TRUE
31	1.426835	1	1	TRUE
32	4.078406	3	5	FALSE
33	4.601557	3	5	FALSE
34	3.927501	3	5	FALSE
35	3.772192	3	5	FALSE
36	3.412997	3	3	TRUE
37	3.680264	3	5	FALSE
38	4.651412	3	5	FALSE
39	4.234318	3	5	FALSE
40	3.993609	3	5	FALSE

41	3.708475	3	5	FALSE
42	3.789214	3	5	FALSE
43	3.439714	3	3	TRUE
44	3.432099	3	3	TRUE
45	4.147688	3	5	FALSE
46	4.461322	5	5	TRUE
47	6.013616	5	5	TRUE
48	4.0828	5	5	TRUE
49	3.87422	5	5	TRUE
50	3.675212	5	5	TRUE
51	3.550162	5	5	TRUE
52	3.654672	5	5	TRUE
53	4.821835	5	5	TRUE
54	4.247035	5	5	TRUE
55	4.249141	5	5	TRUE
56	3.938766	5	5	TRUE
57	3.939277	5	5	TRUE
58	3.406761	5	3	FALSE
59	4.039877	5	5	TRUE
60	4.60591	5	5	TRUE
61	3.719624	5	5	TRUE
62	4.294983	5	5	TRUE
63	4.890436	5	5	TRUE
64	4.899538	5	5	TRUE
65	4.570902	5	5	TRUE
66	4.10441	5	5	TRUE
67	4.561128	5	5	TRUE
68	4.178268	5	5	TRUE
69	4.02118	5	5	TRUE
70	4.417398	5	5	TRUE
71	3.902308	5	5	TRUE
72	4.071671	5	5	TRUE
73	4.199193	5	5	TRUE
74	3.695525	5	5	TRUE
75	3.435482	5	3	FALSE
76	3.097276	5	3	FALSE
77	3.552475	5	5	TRUE
78	4.026964	5	5	TRUE
79	4.788078	5	5	TRUE
80	4.173693	5	5	TRUE
Correctly Classified Instances in %age				**(80-16)/80=80%**

Table 43

Threshold based classification of Stress Class using Function for Drive12.

S No.	Function (F_4)	Actual Stress Class	Predicted Stress Class	Correctly Classified in its class (TRUE/FALSE)
1	3.305234	1	3	FALSE
2	3.410604	1	3	FALSE
3	3.810859	1	5	FALSE
4	4.222709	1	5	FALSE
5	4.41572	1	5	FALSE
6	3.422057	1	3	FALSE
7	2.898779	1	3	FALSE
8	3.535003	1	3	FALSE
9	3.073612	1	3	FALSE
10	2.714278	1	1	TRUE
11	2.153743	1	1	TRUE
12	2.089198	1	1	TRUE
13	1.754873	1	1	TRUE
14	1.760377	1	1	TRUE
15	1.5197	1	1	TRUE
16	1.960488	1	1	TRUE
17	1.917288	1	1	TRUE
18	2.306618	1	1	TRUE
19	2.693896	1	1	TRUE
20	2.072951	1	1	TRUE
21	1.96526	1	1	TRUE
22	1.608651	1	1	TRUE
23	1.476774	1	1	TRUE
24	1.459388	1	1	TRUE
25	1.501156	1	1	TRUE
26	1.652869	1	1	TRUE
27	1.756848	1	1	TRUE
28	1.829827	1	1	TRUE
29	1.77282	1	1	TRUE
30	1.481456	1	1	TRUE
31	1.893812	1	1	TRUE
32	4.425289	3	5	FALSE
33	3.837428	3	5	FALSE
34	4.663134	3	5	FALSE
35	4.002492	3	5	FALSE
36	4.668538	3	5	FALSE
37	4.493892	3	5	FALSE
38	4.412196	3	5	FALSE
39	3.712421	3	5	FALSE
40	3.543563	3	3	TRUE

41	3.6018	3	3	TRUE
42	3.334315	3	3	TRUE
43	3.383391	3	3	TRUE
44	3.362938	3	3	TRUE
45	3.683915	3	3	TRUE
46	3.384174	3	3	TRUE
47	3.265377	5	3	FALSE
48	2.62112	5	1	FALSE
49	2.173395	5	1	FALSE
50	2.266411	5	1	FALSE
51	3.413423	5	3	FALSE
52	3.599266	5	3	FALSE
53	3.03606	5	3	FALSE
54	2.633976	5	1	FALSE
55	2.707398	5	1	FALSE
56	2.60112	5	1	FALSE
57	2.52027	5	1	FALSE
58	2.481013	5	1	FALSE
59	3.747193	5	5	TRUE
60	5.348762	5	5	TRUE
61	4.809711	5	5	TRUE
62	4.818108	5	5	TRUE
63	3.764646	5	5	TRUE
64	4.316446	5	5	TRUE
65	4.118533	5	5	TRUE
66	4.560331	5	5	TRUE
67	5.145817	5	5	TRUE
68	4.317492	5	5	TRUE
69	4.279286	5	5	TRUE
70	4.282316	5	5	TRUE
71	4.255036	5	5	TRUE
72	3.858192	5	5	TRUE
73	4.280223	5	5	TRUE
74	3.972759	5	5	TRUE
75	3.968376	5	5	TRUE
76	4.287532	5	5	TRUE
77	4.377212	5	5	TRUE
78	3.746228	5	5	TRUE
79	3.875167	5	5	TRUE
80	3.789469	5	5	TRUE
Correctly Classified Instances in %age				**(80-29)/80=63.7%**

Table 44
Threshold based classification of Stress Class using Function for Drive15.

S No.	Function (F_4)	Actual Stress Class	Predicted Stress Class	Correctly Classified in its class (TRUE/FALSE)
1	3.225255	1	3	FALSE
2	2.572232	1	1	TRUE
3	1.968844	1	1	TRUE
4	2.729812	1	1	TRUE
5	1.771066	1	1	TRUE
6	2.093651	1	1	TRUE
7	2.39562	1	1	TRUE
8	2.45672	1	1	TRUE
9	2.17282	1	1	TRUE
10	2.198379	1	1	TRUE
11	2.476527	1	1	TRUE
12	1.863022	1	1	TRUE
13	2.188579	1	1	TRUE
14	1.908138	1	1	TRUE
15	1.743692	1	1	TRUE
16	3.089995	1	3	FALSE
17	2.4431	1	1	TRUE
18	2.001892	1	1	TRUE
19	1.891879	1	1	TRUE
20	1.597455	1	1	TRUE
21	1.533605	1	1	TRUE
22	1.36053	1	1	TRUE
23	1.293202	1	1	TRUE
24	1.271499	1	1	TRUE
25	1.788879	1	1	TRUE
26	1.485192	1	1	TRUE
27	1.299446	1	1	TRUE
28	1.597641	1	1	TRUE
29	1.220075	1	1	TRUE
30	4.083276	1	5	FALSE
31	2.953937	1	1	TRUE
32	3.270248	3	3	TRUE
33	3.240663	3	3	TRUE
34	3.761193	3	5	FALSE
35	3.271494	3	3	TRUE
36	2.767319	3	1	FALSE
37	3.531259	3	5	FALSE
38	4.322697	3	5	FALSE
39	3.747554	3	5	FALSE
40	3.436072	3	3	TRUE

41	3.633375	3	5	FALSE
42	3.428656	3	3	TRUE
43	3.773124	3	5	FALSE
44	4.583643	3	5	FALSE
45	4.746207	3	5	FALSE
46	4.208291	5	5	TRUE
47	6.904758	5	5	TRUE
48	5.283928	5	5	TRUE
49	4.331924	5	5	TRUE
50	4.082544	5	5	TRUE
51	3.153272	5	3	FALSE
52	3.613007	5	5	TRUE
53	3.629202	5	5	TRUE
54	3.601488	5	5	TRUE
55	3.666237	5	5	TRUE
56	3.508803	5	5	TRUE
57	4.341002	5	5	TRUE
58	4.307352	5	5	TRUE
59	3.928084	5	5	TRUE
60	4.544185	5	5	TRUE
61	4.224493	5	5	TRUE
62	4.295059	5	5	TRUE
63	3.772609	5	5	TRUE
64	4.263113	5	5	TRUE
65	4.273636	5	5	TRUE
66	4.082141	5	5	TRUE
67	3.706984	5	5	TRUE
68	4.53726	5	5	TRUE
69	3.929862	5	5	TRUE
70	4.063006	5	5	TRUE
71	3.742935	5	5	TRUE
72	3.597267	5	5	TRUE
73	3.669392	5	5	TRUE
74	3.833136	5	5	TRUE
75	4.78118	5	5	TRUE
76	5.736074	5	5	TRUE
77	3.662908	5	5	TRUE
Correctly Classified Instances in %age				**(77-13)/77=83.1%**

Table 45
Threshold based classification of Stress Class using Function for Drive16.

S No.	Function (F_4)	Actual Stress Class	Predicted Stress Class	Correctly Classified in its class (TRUE/FALSE)
1	2.512633	1	3	FALSE
2	1.825258	1	1	TRUE
3	1.886143	1	1	TRUE
4	1.735168	1	1	TRUE
5	1.802404	1	1	TRUE
6	1.699059	1	1	TRUE
7	1.571548	1	1	TRUE
8	1.223254	1	1	TRUE
9	1.144929	1	1	TRUE
10	1.014772	1	1	TRUE
11	0.80597	1	1	TRUE
12	0.746543	1	1	TRUE
13	0.705663	1	1	TRUE
14	0.641523	1	1	TRUE
15	0.482834	1	1	TRUE
16	0.471228	1	1	TRUE
17	4.966778	3	5	FALSE
18	3.915204	3	5	FALSE
19	3.369314	3	5	FALSE
20	3.157144	3	3	TRUE
21	3.265911	3	5	FALSE
22	3.249131	3	5	FALSE
23	3.579077	3	5	FALSE
24	3.202252	3	5	FALSE
25	3.112695	3	3	TRUE
26	2.886329	3	3	TRUE
27	2.719679	3	3	TRUE
28	2.986349	3	3	TRUE
29	4.052324	5	5	TRUE
30	5.657707	5	5	TRUE
31	5.222302	5	5	TRUE
32	4.846339	5	5	TRUE
33	4.036354	5	5	TRUE
34	3.024366	5	3	FALSE
35	3.977657	5	5	TRUE
36	3.649438	5	5	TRUE
37	4.912907	5	5	TRUE
38	4.431224	5	5	TRUE
39	5.016858	5	5	TRUE
40	4.628236	5	5	TRUE

41	3.937386	5	5	TRUE
42	3.083007	5	3	FALSE
43	5.198012	5	5	TRUE
44	3.66632	5	5	TRUE
45	4.019046	5	5	TRUE
46	4.715076	5	5	TRUE
47	4.181242	5	5	TRUE
48	4.412239	5	5	TRUE
49	3.836053	5	5	TRUE
50	4.190005	5	5	TRUE
51	3.697933	5	5	TRUE
52	4.461799	5	5	TRUE
53	3.242788	5	5	TRUE
54	3.046099	5	3	FALSE
55	3.349231	5	5	TRUE
56	3.49574	5	5	TRUE
57	3.226322	5	5	TRUE
58	3.283888	5	5	TRUE
59	3.206424	5	5	TRUE
60	4.776734	5	5	TRUE
61	4.492155	5	5	TRUE
62	4.17084	5	5	TRUE
63	3.479303	5	5	TRUE
64	4.045055	5	5	TRUE
Correctly Classified Instances in %age				**(64-11)/64=82.8%**

Table 46
Percentage of Correctly Classified Instances after Classification.

Drive No.	**Classification Accuracy using Common Thresholds (3, 3.5)**	**Classification Accuracy for Individual thresholds**	
Drive05	78.7 %	**80 %**	(3.1, 3.6)
Drive06	61.2 %	67.5 %	(3.2, 3.4)
Drive07	68.7 %	68.7 %	(3, 3.5)
Drive08	81.2 %	**83.7 %**	(3, 3.9)
Drive09	70 %	**81.4 %**	(1.9, 3)
Drive10	60 %	66.2 %	(2.6, 3)
Drive11	80 %	**80 %**	(3, 3.5)
Drive12	62.5 %	63.7 %	(2.8, 3.7)
Drive15	83.1 %	**83.1 %**	(3, 3.5)
Drive16	75 %	**82.8 %**	(2.5, 3.2)

It is depicted from Table 46, that the stress function ($\digamma_4$), alone is able to classify 6 out of 10 drives with accuracy of 80% or more than 80% into their respective Stress Class of automobile driver (LS, MS and HS). These accuracies have been highlighted in the table. However, the classification accuracy in case of Drive06, Drive07, Drive10 and Drive 12 is very low. To further improve the accuracy of classification of drives Drive06, Drive07, Drive10 and Drive 12, we have used Artificial Neural Network (ANN) approach. As ANN requires multiple attributes, the attributes used in function ($\digamma_4$), namely Mean Hand GSR, Mean Foot GSR, Respiration Rate and Mean HR are chosen as four inputs to ANN used for stress classification. Different Algorithmic approach with supervised classification is also used to verify the classification accuracy.

5.3 Classification Algorithms

Classification is a task of training a classification model that maps each attribute set X to one of the predefined class labels Y. We have again used 'WEKA' [28] for classification of stress by implementing different machine learning. Machine learning covers such a broad range of processes that it is difficult to define precisely. A dictionary definition includes phrases such as to gain knowledge or understanding of or skill by studying the instruction or experience and modification of a behavioral tendency by experienced zoologists and psychologists study learning in animals and humans [31]. The extraction of important information from a large pile of data and its correlations is often the advantage of using machine learning. New knowledge about tasks is constantly being discovered by humans and vocabulary changes. There is a constant stream of new events in the world and continuing redesign of Artificial Intelligent systems to conform to new knowledge is impractical but machine learning methods might be able to track much of it [31]. Some of the important features of different algorithms used in our research are discussed below:

5.3.1 Naïve Bayes

A naive Bayes classifier is a simple probabilistic classifier based on applying Bayes' theorem with strong (naive) independence assumptions. In simple terms, a naive Bayes classifier assumes that the presence or absence of a particular feature is unrelated to the presence or absence of any other feature, given the class variable. For example, a fruit may be considered to be an apple if it is red, round, and about 3" in diameter. A naive Bayes classifier considers each of these features to contribute independently to the probability that this fruit is an apple, regardless of the presence or absence of the other features.

For some types of probability models, naive Bayes classifiers can be trained very efficiently in a supervised learning setting. In many practical applications, parameter estimation for naive Bayes models uses the method of maximum likelihood; in other words, one can work with the naive Bayes model without accepting Bayesian probability or using any Bayesian methods. An advantage of naive Bayes is that it only requires a small amount of training data to estimate the parameters (means and variances of the variables) necessary for classification. Because independent variables are assumed, only the variances of the variables for each class need to be determined and not the entire covariance matrix.

Naive Bayes Classifier Introductory Overview

The Naive Bayes Classifier technique is based on the so-called Bayesian theorem and is particularly suited when the dimensionality of the inputs is high. Despite its simplicity, Naive Bayes can often outperform more sophisticated classification methods.

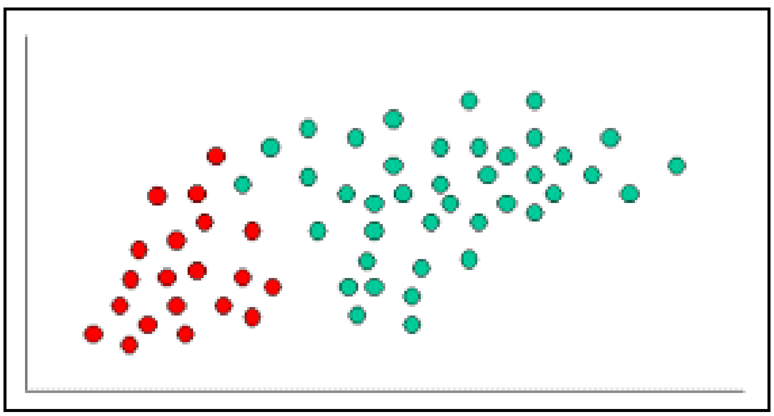

Fig. 7. An example of naïve Bayes showing red and green labels.

To demonstrate the concept of Naïve Bayes Classification, consider the example displayed in the illustration above in Fig. 7. As indicated, the objects can be classified as either GREEN or RED. Our task is to classify new cases as they arrive, i.e., decide to which class label they belong, based on the currently exiting objects.

Since there are twice as many GREEN objects as RED, it is reasonable to believe that a new case (which hasn't been observed yet) is twice as likely to have membership GREEN rather than RED. In the Bayesian analysis, this belief is known as the prior probability. Prior probabilities are based on previous experience, in this case the percentage of GREEN and RED objects, and often used to predict outcomes before they actually happen.

Thus, we can write:

$$\text{Prior probability for GREEN} \propto \frac{\text{Number of GREEN objects}}{\text{Total number of objects}}$$

$$\text{Prior probability for RED} \propto \frac{\text{Number of RED objects}}{\text{Total number of objects}}$$

Since there is a total of 60 objects, 40 of which are GREEN and 20 RED, our prior probabilities for class membership are:

$$\text{Prior probability for GREEN} \propto \frac{40}{60}$$

$$\text{Prior probability for RED} \propto \frac{20}{60}$$

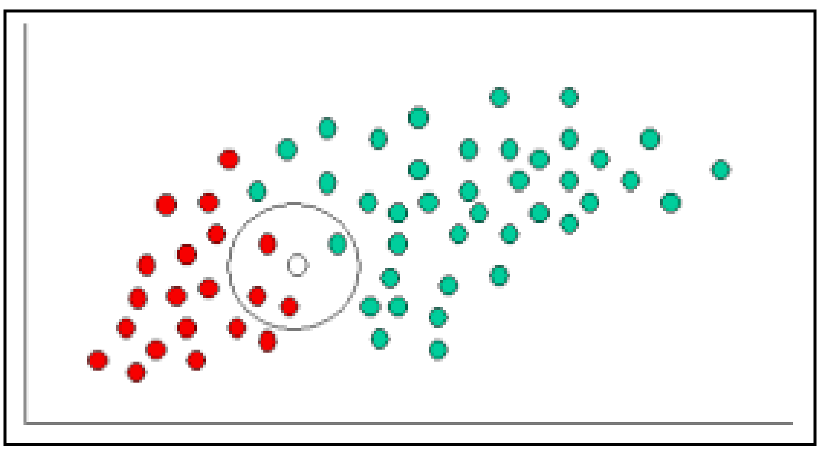

Fig. 8. An example of naïve Bayes showing red and green labels.

Having formulated our prior probability, we are now ready to classify a new object (WHITE circle) as shown in Fig. 8. Since the objects are well clustered, it is reasonable to assume that the more GREEN (or RED) objects in the vicinity of X, the more likely that the new cases belong to that particular color. To measure this likelihood, we draw a circle around X which encompasses a number (to be chosen a priori) of points irrespective of their class labels. Then we calculate the number of points in the circle belonging to each class label. From this we calculate the likelihood:

$$\text{Likelihood of } X \text{ given GREEN} \propto \frac{\text{Number of GREEN in the vicinity of } X}{\text{Total number of GREEN cases}}$$

$$\text{Likelihood of } X \text{ given RED} \propto \frac{\text{Number of RED in the vicinity of } X}{\text{Total number of RED cases}}$$

From the illustration above, it is clear that Likelihood of X given GREEN is smaller than Likelihood of X given RED, since the circle encompasses 1 GREEN object and 3 RED ones. Thus:

$$\text{Probability of } X \text{ given GREEN} \propto \frac{1}{40}$$

$$\text{Probability of } X \text{ given RED} \propto \frac{3}{20}$$

Although the prior probabilities indicate that X may belong to GREEN (given that there are twice as many GREEN compared to RED) the likelihood indicates otherwise; that the class membership of X is RED (given that there are more RED objects in the vicinity of X than GREEN). In the Bayesian analysis, the final classification is produced by combining both sources of information, i.e., the prior and the likelihood, to form a posterior probability using the so-called Bayes' rule.

$$\begin{aligned}
&\text{Posterior probability of } X \text{ being GREEN} \propto \\
&\ \text{Prior probability of GREEN} \times \text{Likelihood of } X \text{ given GREEN} \\
&= \frac{4}{6} \times \frac{1}{40} = \frac{1}{60} \\
&\text{Posterior probability of } X \text{ being RED} \propto \\
&\ \text{Prior probability of RED} \times \text{Likelihood of } X \text{ given RED} \\
&= \frac{2}{6} \times \frac{3}{20} = \frac{1}{20}
\end{aligned}$$

Finally, we classify X as RED since its class membership achieves the largest posterior probability.

5.3.2 Multilayer Perceptron Neural Network (MLP)

The following Fig.9. illustrates a perceptron network with three layers:

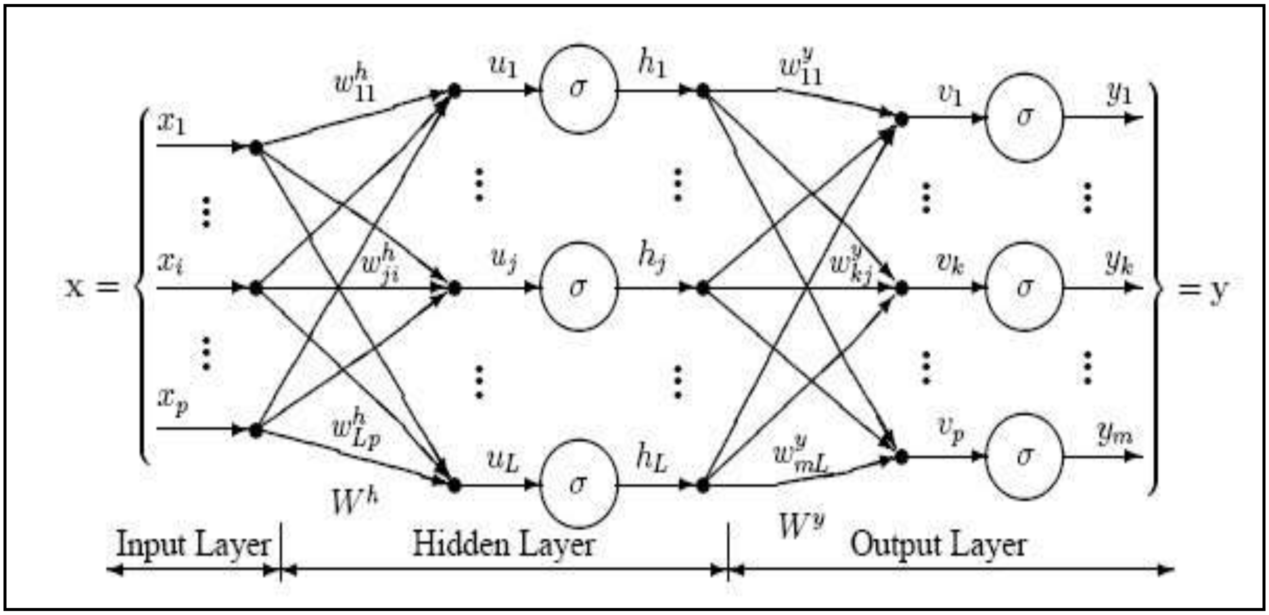

Fig. 9. A 3 layered perceptron network.

This network has an input layer (on the left) with three neurons, one hidden layer (in the middle) with three neurons and an output layer (on the right) with three neurons. There is one neuron in the input layer for each predictor variable. In the case of categorical variables, N-1 neurons are used to represent the N categories of the variable.

Input Layer — A vector of predictor variable values ($x_1 \ldots x_p$) is presented to the input layer. The input layer (or processing before the input layer) standardizes these values so that the range of each variable is -1 to 1. The input layer distributes the values to each of the neurons in the hidden layer. In addition to the predictor variables, there is a constant input of 1.0, called the bias that is fed to each of the hidden layers; the bias is multiplied by a weight and added to the sum going into the neuron.

Hidden Layer — Arriving at a neuron in the hidden layer, the value from each input neuron is multiplied by a weight (w_{ji}), and the resulting weighted values are added together producing a combined value u_j. The weighted sum (u_j) is fed into a transfer function, σ, which outputs a value h_j. The outputs from the hidden layer are distributed to the output layer.

Output Layer — Arriving at a neuron in the output layer, the value from each hidden layer neuron is multiplied by a weight (w_{kj}), and the resulting weighted values are added together producing a combined value v_j. The weighted sum (v_j) is fed into a transfer function, σ, which outputs a value y_k. The y values are the outputs of the network.

If a regression analysis is being performed with a continuous target variable, then there is a single neuron in the output layer, and it generates a single y value. For classification problems with categorical target variables, there are N neurons in the output layer producing N values, one for each of the N categories of the target variable.

The network diagram shown above is a full-connected, three layer, feed-forward, perceptron neural network. "Fully connected" means that the output from each input and hidden neuron is distributed to all of the neurons in the following layer. "Feed forward" means that the values only move from input to hidden to output layers; no values are fed back to earlier layers (a Recurrent Network allows values to be fed backward).

All neural networks have an input layer and an output layer, but the number of hidden layers may vary. When there is more than one hidden layer, the output from one hidden layer is fed into the next hidden layer and separate weights are applied to the sum going into each layer.

Training Multilayer Perceptron Networks

The goal of the training process is to find the set of weight values that will cause the output from the neural network to match the actual target values as closely as possible. There are several issues involved in designing and training a multilayer perceptron network:

- Selecting how many hidden layers to use in the network.
- Deciding how many neurons to use in each hidden layer.
- Finding a globally optimal solution that avoids local minima.
- Converging to an optimal solution in a reasonable period of time.
- Validating the neural network to test for overfitting.

5.3.3 k-NN (k-Nearest Neighbor)

Instance-based classifiers such as the kNN classifier operate on the premises that classification of unknown instances can be done by relating the unknown to the known according to some distance/similarity function. The intuition is that two instances far apart in the instance space defined by the appropriate distance function are less likely than two closely situated instances to belong to the same class.

The learning process

Unlike many artificial learners, instance-based learners do not abstract any information from the training data during the learning phase. Learning is merely a question of encapsulating the training data. The process of generalization is postponed until it is absolutely unavoidable, that is, at the time of classification. This property has led to the referring to instance-based learners as lazy learners, whereas classifiers such as feedforward neural networks, where proper abstraction is done during the learning phase, often are entitled eager learners.

Classification

Classification (generalization) using an instance-based classifier can be a simple matter of locating the nearest neighbor in instance space and labeling the unknown instance with the same class label as that of the located (known) neighbor. This approach is often referred to as a nearest neighbor classifier. The downside of this simple approach is the lack of robustness that characterizes the resulting classifiers. The high degree of local sensitivity makes nearest neighbor classifiers highly susceptible to noise in the training data.

More robust models can be achieved by locating k, where $k > 1$, neighbours and letting the majority vote decide the outcome of the class labeling. A higher value of k results in a smoother, less locally sensitive, function. The nearest neighbor classifier can be regarded as a special case of the more general k-nearest neighbors classifier, hereafter referred to as a kNN classifier. The drawback of increasing the value of k is of course that as k approaches n, where n is the size of the instance base, the performance of the classifier will approach that of the most straightforward statistical baseline, the assumption that all unknown instances belong to the class most frequently represented in the training data.

This problem can be avoided by limiting the influence of distant instances. One way of doing so is to assign a weight to each vote, where the weight is a function of the distance between the unknown and the known instance. By letting each weight be defined by the inversed squared distance between the known and unknown instances votes cast by distant instances will have very little influence on the decision process compared to instances in the near neighborhood. Distance weighted voting usually serves as a good middle ground as far as local sensitivity is concerned.

Example

To demonstrate a k-nearest neighbor analysis, let's consider the task of classifying a new object (query point denoted as blue dot) among a number of known examples. This is shown in the Fig. 10, which depicts the examples (instances) with the plus and minus signs and the query point with a blue dot. Our task is to estimate (classify) the outcome of the query point based on a selected number of its nearest neighbors. In other words, we want to know whether the query point can be classified as a plus or a minus sign.

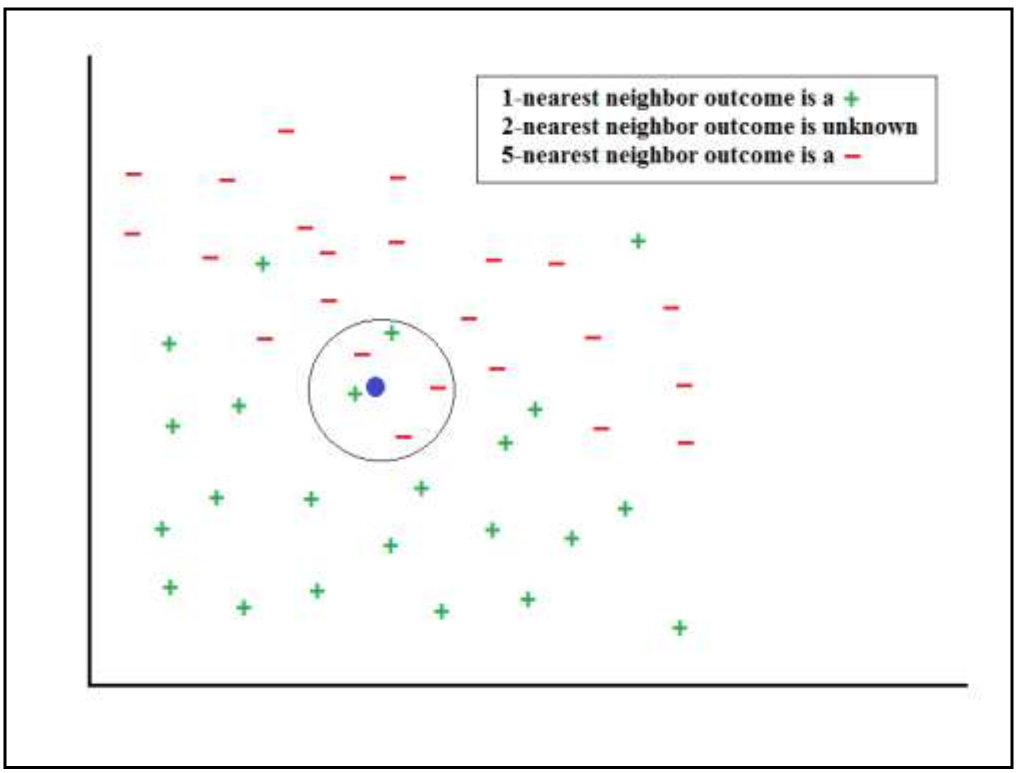

Fig. 10. An example of k-NN approach.

To proceed, let's consider the outcome of KNN based on 1-nearest neighbor. It is clear that in this case KNN will predict the outcome of the query point with a plus (since the closest point carries a plus sign). Now let's increase the number of nearest neighbors to 2, i.e., 2-nearest neighbors. This time KNN will not be able to classify the outcome of the query point since the second closest point is a minus, and so both the plus and the minus signs achieve the same score (i.e., win the same number of votes). For the next step, let's increase the number of nearest neighbors to 5 (5-nearest neighbors). This will define a nearest neighbor region, which is indicated by the circle shown in the figure above. Since there are 2 and 3 plus and minus signs, respectively, in this circle KNN will assign a minus sign to the outcome of the query point.

5.3.4 Support Vector Machines (SVM)

Support Vector Machines are based on the concept of decision planes that define decision boundaries. A decision plane is one that separates between a set of objects having different class memberships. A schematic example is shown in the illustration below. In this example, the objects belong either to class GREEN or RED. The separating line defines a boundary on the right side of which all objects are GREEN and to the left of which all objects are RED. Any new object (white circle) falling to the right is labeled, i.e., classified, as GREEN (or classified as RED should it fall to the left of the separating line).

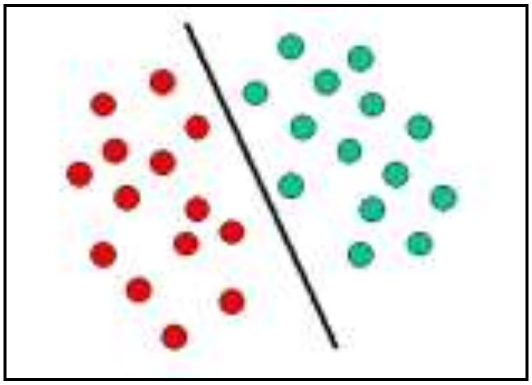

Fig. 11. An example of SVM approach.

The above is a classic example of a linear classifier, i.e., a classifier that separates a set of objects into their respective groups (GREEN and RED in this case) with a line. Most classification tasks, however, are not that simple, and often more complex structures are needed in order to make an optimal separation, i.e., correctly classify new objects (test cases) on the basis of the examples that are available (train cases). This situation is depicted in the illustration below. Compared to the previous schematic, it is clear that a full separation of the GREEN and RED objects would require a curve (which is more complex than a line). Classification tasks based on drawing separating lines to distinguish between objects of different class memberships are known as hyperplane classifiers. Support Vector Machines are particularly suited to handle such tasks.

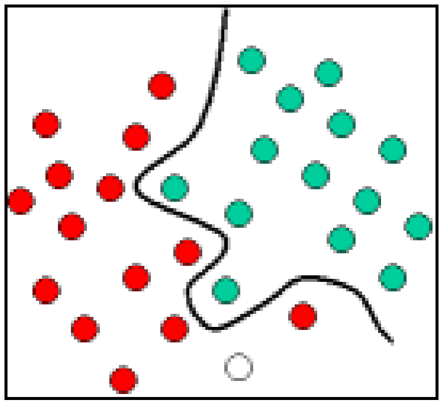

Fig. 12. An example of grouping of set in SVM approach.

The illustration below shows the basic idea behind Support Vector Machines. Here we see the original objects (left side of the schematic) mapped, i.e., rearranged, using a set of mathematical functions, known as kernels. The process of rearranging the objects is known as mapping (transformation). Note that in this new setting, the mapped objects (right side of the schematic) is linearly separable and, thus, instead of constructing the complex curve (left schematic), all we have to do is to find an optimal line that can separate the GREEN and the RED objects.

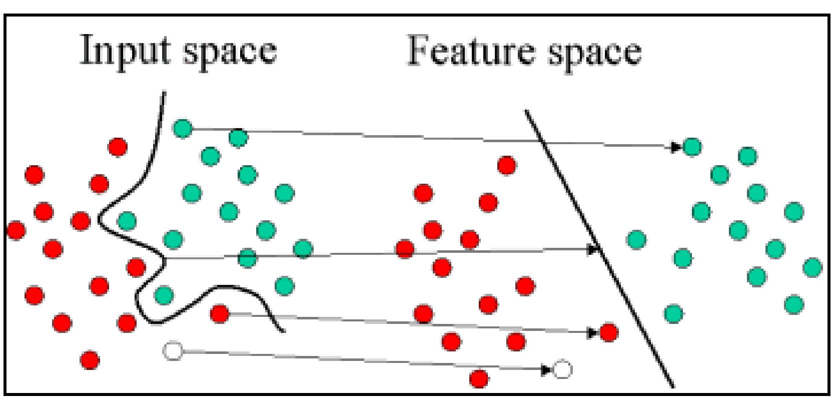

Fig. 13. The simple mapping process in SVM approach.

5.3.5 C4.5 algorithm (J48)

C4.5 is an algorithm used to generate a decision tree developed by Ross Quinlan. C4.5 is an extension of Quinlan's earlier ID3 algorithm. The decision trees generated by C4.5 can be used for classification, and for this reason, C4.5 is often referred to as a statistical classifier.

The J48 algorithm is WEKA's implementation of the C4.5 decision tree learner. The algorithm uses a greedy technique to induce decision trees for classification and uses reduced-error pruning

5.3.6 Decision trees (Random Forest)

Decision trees are a class of predictive data mining tools which predict either a categorical or continuous response variable. They get their name from the structure of the models built. A series of decisions are made to segment the data into homogeneous subgroups. This is also called recursive partitioning. When drawn out graphically, the model can resemble a tree with branches as shown by an example in Fig. 14.

Several tools fall into the category of decision tree including Classification and Regression Trees (C&RT), Chi Square Automatic Interaction Detector (CHAID), Random Forests and Boosted Trees. Each of these tools has unique qualities while sharing the principles of decision trees. C&RT and CHAID both build only one tree, while Random Forest and Boosted Trees build multiple.

A decision tree is comprised of nodes and splits to the data. The tree starts with all training data residing in the first node. An initial split is made using a predictor variable, segmenting the data into 2 or more child nodes. Splits can then be made from the child nodes. A terminal node is one where no more splits are made. Predictions are made based on the make-up of terminal nodes. To use a decision tree to make a prediction, the split decisions are followed until a terminal node is reached. This can be done manually by reviewing the tree graph. Software programs can make these predictions as well and for a large data set for deployment.

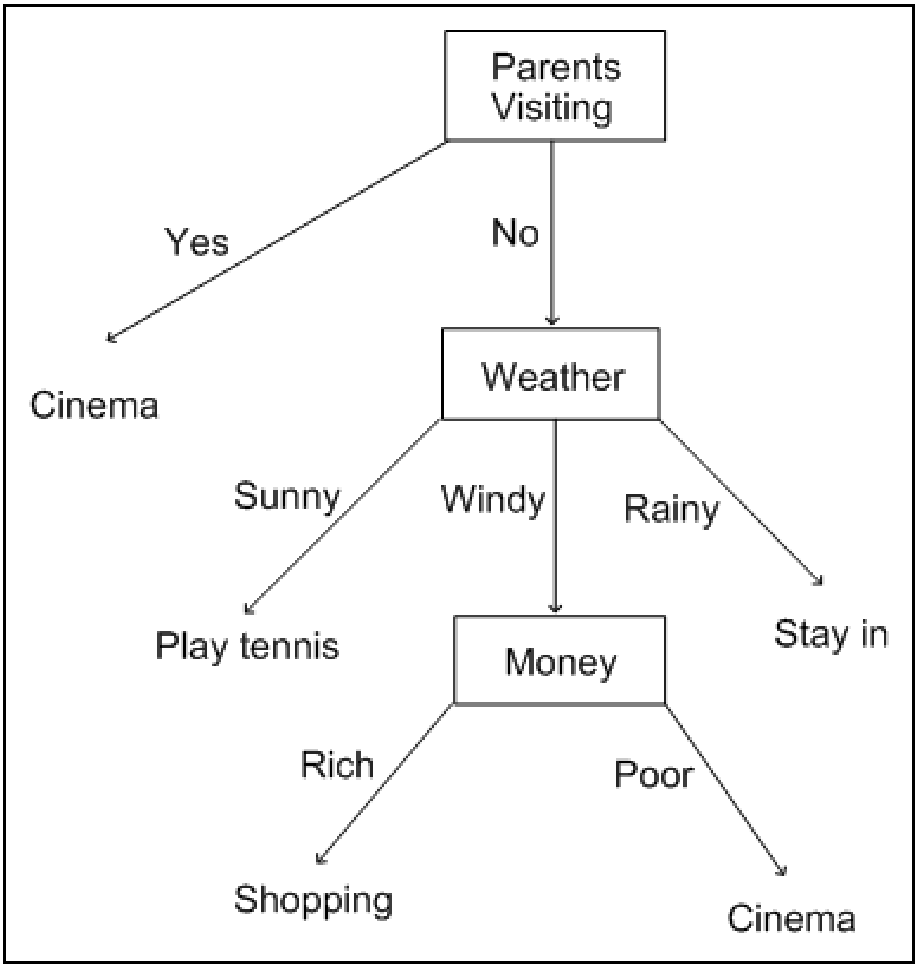

Fig. 14. A Simple Decision Tree Example

Decision trees offer many advantaged. One important advantage is the ease of interpretation of a decision tree. While the tree can be complex, involving a large number of splits and nodes, users can interpret the model. Additionally, making model predictions does not involve mathematical calculations as in General Linear Models. The predictions are based on decision rules. In classification problems, the user can specify misclassification cost. Decision trees tend to give good predictive accuracy and can allow for missing data in deployment.

Random Forest

A Random Forest consists of a collection or ensemble of simple tree predictors, each capable of producing a response when presented with a set of predictor values. For classification problems, this response takes the form of a class membership, which associates, or classifies, a set of independent predictor values with one of the categories present in the dependent variable.

Alternatively, for regression problems, the tree response is an estimate of the dependent variable given the predictors. The Random Forest algorithm was developed by Breiman. Random forests are an ensemble learning method for classification (and regression) that operate by constructing a multitude of decision trees at training time and outputting the class that is the mode of the classes output by individual trees.

Random Forest consists of an arbitrary number of simple trees, which are used to determine the final outcome. For classification problems, the ensemble of simple trees vote for the most popular class. In the regression problem, their responses are averaged to obtain an estimate of the dependent variable. Using tree ensembles can lead to significant improvement in prediction accuracy (i.e., better ability to predict new data cases).

5.4 Improvement of classification accuracy using ANN

A typical two-layer feedforward back propagation network having 4 inputs, 3 outputs, one hidden layer with 10 neurons and a output layer is used. Sigmoid function is used in hidden layer as shown in Fig. 15. The Neural Network Pattern Recognition Tool (nprtool) of MATLAB is employed to create and train a network, and evaluate its performance using mean square error and confusion matrices. The result obtained is shown in Table 11. It clearly illustrates that by using ANN, the percentage of classification accuracy increases by significant amount as compared to the results obtained by using threshold classification.

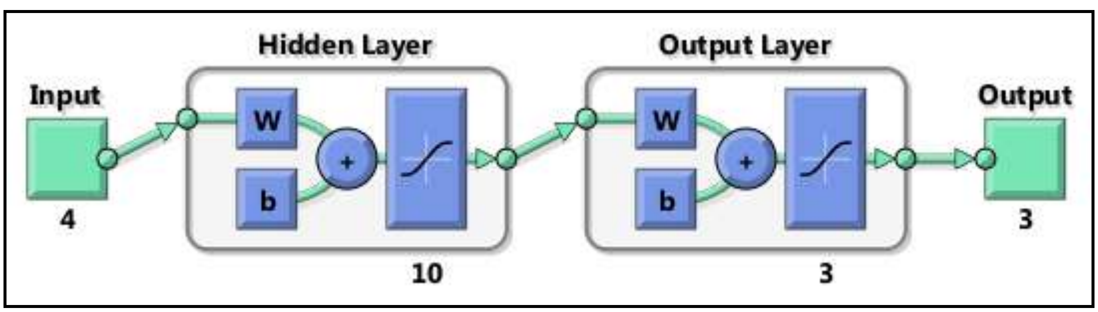

Fig. 15. A two-layer feed-forward network.

In summary, a single function ($Ƒ_4$) is sufficient for the stress classification in most of the drives. However, there may be some problems in data acquisition as in Drive06, Drive07, Drive10 and Drive12. This calls for an intelligent approach of using ANN for enhancing accuracy of classification.

Table 47
Percentage of Correctly Classified Instances after using Artificial Neural Network.

Drive No.	Individual threshold for each Drive	Using Artificial Neural Network (feedforward back propagation)	Confusion Matrix
Drive06	67.5 %	85 %	*Predicted Class* *Actual Class* $\begin{array}{cc} & \begin{array}{ccc} \text{LS} & \text{MS} & \text{HS} \end{array} \\ \begin{array}{c}\text{LS}\\ \text{MS}\\ \text{HS}\end{array} & \begin{bmatrix} 27 & 0 & 2 \\ 1 & 11 & 4 \\ 4 & 1 & 30 \end{bmatrix} \end{array}$
Drive07	68.7 %	82.5 %	*Predicted Class* *Actual Class* $\begin{array}{cc} & \begin{array}{ccc} \text{LS} & \text{MS} & \text{HS} \end{array} \\ \begin{array}{c}\text{LS}\\ \text{MS}\\ \text{HS}\end{array} & \begin{bmatrix} 24 & 0 & 2 \\ 0 & 13 & 5 \\ 2 & 5 & 29 \end{bmatrix} \end{array}$
Drive10	66.2 %	88.75 %	*Predicted Class* *Actual Class* $\begin{array}{cc} & \begin{array}{ccc} \text{LS} & \text{MS} & \text{HS} \end{array} \\ \begin{array}{c}\text{LS}\\ \text{MS}\\ \text{HS}\end{array} & \begin{bmatrix} 29 & 0 & 2 \\ 0 & 12 & 3 \\ 2 & 2 & 30 \end{bmatrix} \end{array}$
Drive12	63.7 %	77.5 %	*Predicted Class* *Actual Class* $\begin{array}{cc} & \begin{array}{ccc} \text{LS} & \text{MS} & \text{HS} \end{array} \\ \begin{array}{c}\text{LS}\\ \text{MS}\\ \text{HS}\end{array} & \begin{bmatrix} 26 & 0 & 5 \\ 0 & 6 & 9 \\ 1 & 3 & 30 \end{bmatrix} \end{array}$

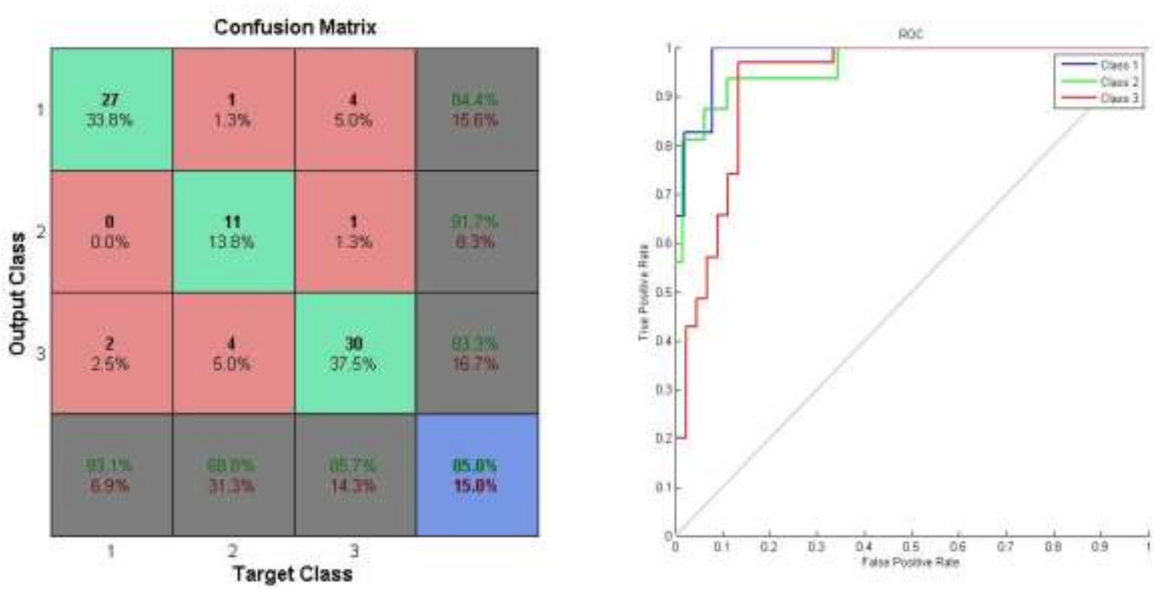

Fig. 16. Confusion Matrix and ROC Plot for Drive06.

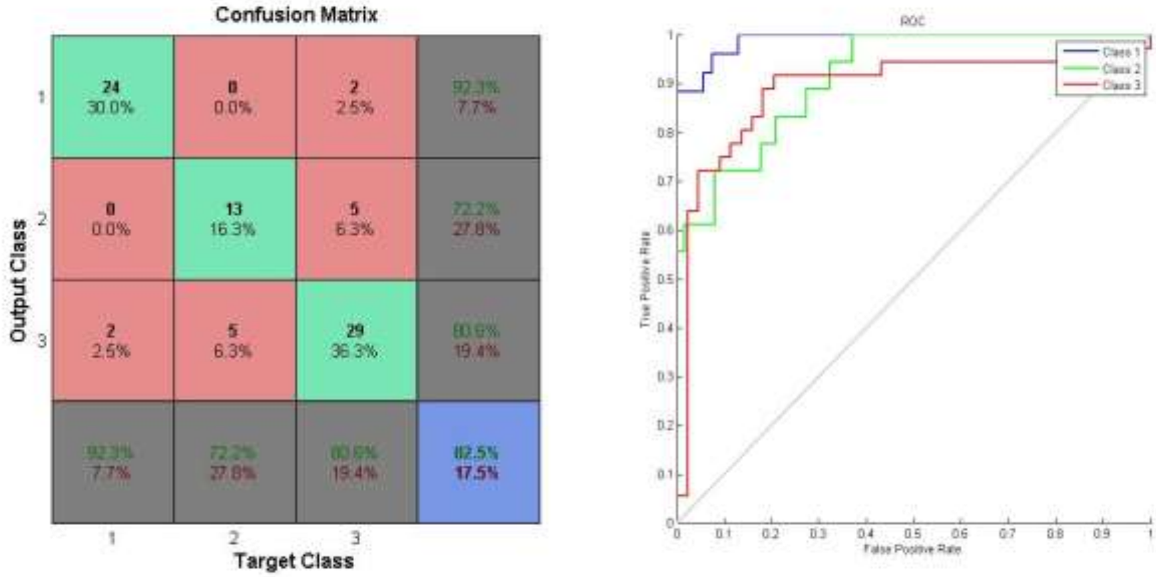

Fig. 17. Confusion Matrix and ROC Plot for Drive07.

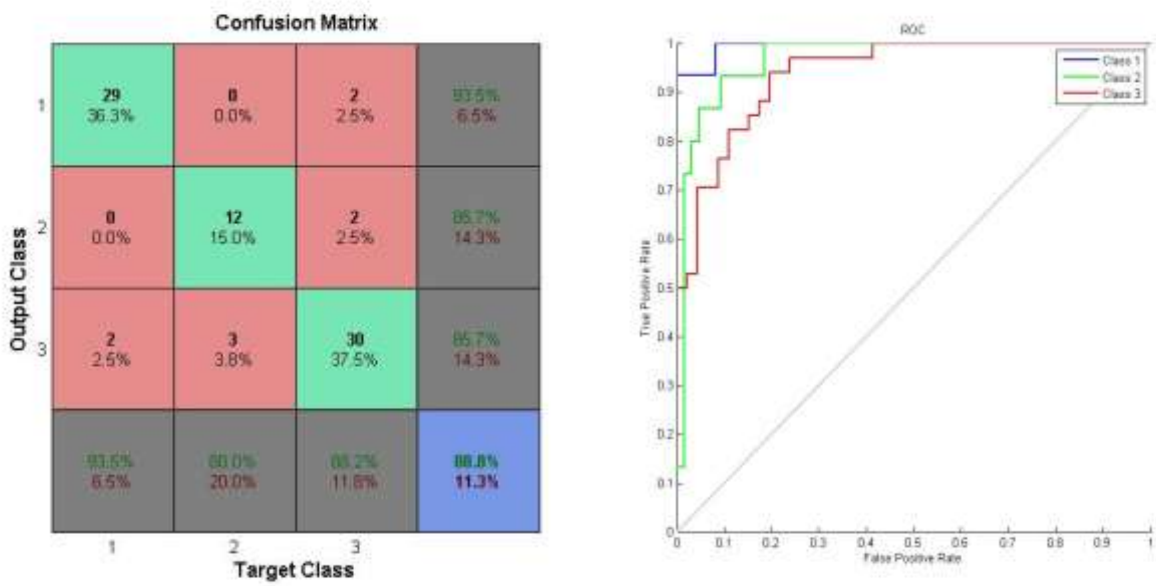

Fig. 18. Confusion Matrix and ROC Plot for Drive10.

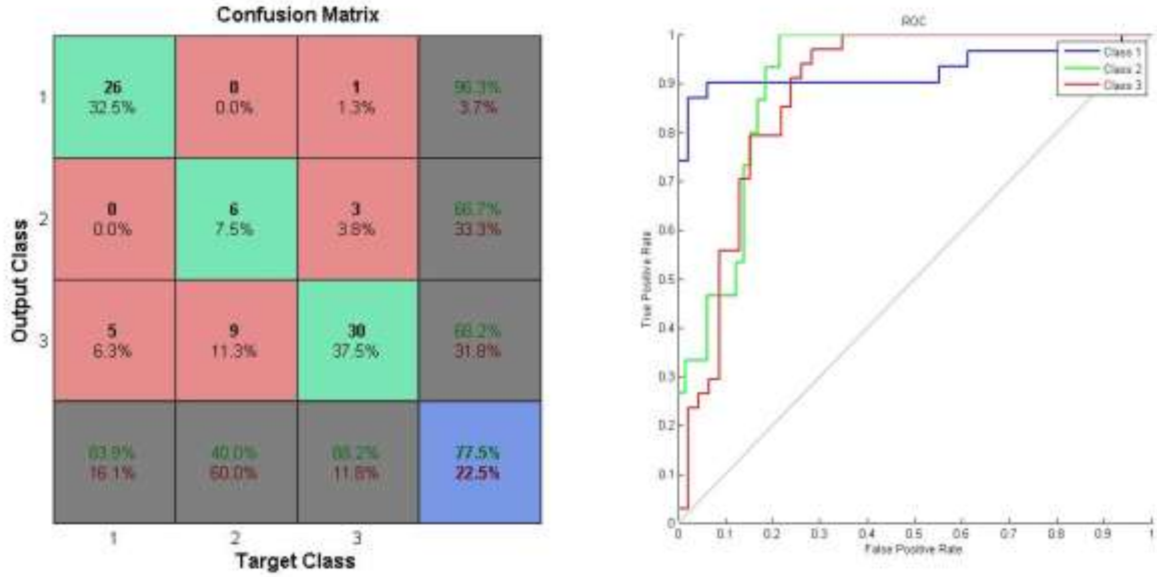

Fig. 19. Confusion Matrix and ROC Plot for Drive12.

Summary of Individual Threshold based classification of Drive05, Drive08, Drive09, Drive11, Drive15 and Drive16 and ANN based classification of Drive06, Drive07, Drive10 and Drive12 are shown in Table 48.

Table 48
Summary of Correctly Classified Instances after ANN Classification.

Drive No.	Classification Accuracy using Common Thresholds (3, 3.5)	Classification Accuracy using Individual thresholds and by ANN
Drive05	78.7 %	**80 %**
Drive06	61.2 %	**85 % (by ANN)**
Drive07	68.7 %	**82.5 % (by ANN)**
Drive08	81.2 %	**83.7 %**
Drive09	70 %	**81.4 %**
Drive10	60 %	**88.7 % (by ANN)**
Drive11	80 %	**80 %**
Drive12	62.5 %	**77.5 % (by ANN)**
Drive15	83.1 %	**83.1 %**
Drive16	75 %	**82.8 %**

5.5 Verification of the Classification accuracy

For verification of the classification accuracy achieved through Threshold Classification and ANN based classification, we have used "WEKA" software. Different supervised algorithmic approaches are applied on Drive06, Drive07, Drive10 and Drive12 database. The algorithms used are listed below:

1. Naïve Bayes
2. Simple Logistics
3. K-NN
4. J48 Decision Tree
5. Random Forest Decision Tree
6. Multilayer Perceptron (MLP)

The results obtained after application of algorithmic approach is shown in Table 49.

Table 49
Verification of Correctly Classified Instances using Algorithmic Classification.

	Simple Logistic	Naïve Bayes	MLP	k-NN				J48	Random Forest	Avg.
				1-NN	3-NN	5-NN	7-NN			
Drive06	86.3	85	80	85	90	85	83.8	86.3	88.8	**85.6**
Drive07	81.3	75	73.8	80	75	72.5	75	72.5	73.8	**75.4**
Drive10	82.5	75	75	81.3	83.8	82.5	80	73.8	77.5	**79.0**
Drive12	67.5	51.3	61.3	70	71.3	67.5	61.3	66.3	68.8	**65.0**

Table 49; indicate that classification accuracy achieved by "WEKA" is comparable to ANN classification accuracy. It may be concluded that for threshold based classification, the best suited physiological parameters are Mean Hand GSR, Mean Foot GSR, RR and Mean HR. Since the base physiological parameters are different for different drives, individual thresholds for each drive gives better classification accuracy as compared to the common thresholds. In case of some discrepancies in the data, classification accuracy may fall down. In that case, ANN based classifier is able to give better accuracy [32].

Chapter 6

CONCLUSION AND FUTURE SCOPE

6.1 Conclusion

Autonomic Nervous System (ANS) control all the vital involuntary muscles in human body like glands, blood vessels, heart, lungs, intestine etc. Whenever an automobile driver is subjected to fast changing environment (e.g. driving scenario) the ANS activity of automobile driver also changes according to it. As a result, the physiological parameters like galvanic skin response (GSR), electrocardiography (ECG), electromyography (EMG) of the trapezius muscle, and respiration rate (RR) also changes. This change in Heart rate (HR), EMG, GSR and RR can be recorded via sensors to detect the dynamic stress level of the driver in real-time. Thus stress detection systems can detect stress by using physiological signals like HR, EMG, RR, GSR etc. providing a precise output indicating to what extent an automobile driver is under a stress.

We accomplished following tasks in order to develop a stress level monitor:

i. Evaluate correlation of individual physiological parameters like GSR, ECG, EMG and RR with the traffic condition like highway, city, and rest.
ii. Analysis of the correlation identified the physiological parameters best suited to detect stress level.
iii. Derived a stress function of the selected/identified physiological parameters that gives better correlation with traffic conditions, as compared to any one physiological parameter taken in isolation.
iv. Based on this stress function a classifier have been developed that indicate the stress level of the driver as:
    a. Low
    b. Moderate
    c. High

We assumed that the psychological stress in driver is only on account of the traffic conditions which again depend on the terrain. Correlation analysis shows very strong linear relationship between Stress and Mean Hand GSR. This stress function acts as a direct indicator of stress level of automobile drivers.

6.2 Future Scope

Using the results of this research work, a microcontroller/field-programmable gate array (FPGA) based stress level indicator can be designed on real time monitoring of driver's mental stress. In case the driver is in low or moderate stress, he may be allowed to use navigational tools, else in case of high stress, the driver may be advised only to focus on driving and avoid multitasking, or take rest in between.

In future, we may want vehicles to be more intelligent and responsive, managing information delivery in the context of the driver's psychological state. Physiological sensing is one method of accomplishing this goal. Drivers could be allowed to make safe errors while talking on the cell phone or using visual navigation aids. But if a high-stress condition were detected using the algorithm on data, the driver distractions could be turned off until the driver recovered to a medium-stress level. These sensors can be tested in the laboratory and embedded into wearable and automotive systems to measure affective signals in the natural ambulatory environment. Finally, these findings may contribute toward progress in developing machines which can respond intelligently to human affect while simultaneously understanding human mental condition.

PUBLICATIONS FROM THIS RESEARCH

Following are the related publications from this research:

Published in: International Journal of Electronics Engineering (IJEE)

[1] Mandeep Singh and Abdullah Bin Queyam, "Stress Detection in Automobile Drivers using Physiological Parameters: A Review," *International Journal of Electronics Engineering*, vol. 5, no. 2, Dec 2013. [In Press]

[2] Mandeep Singh and Abdullah Bin Queyam, "Correlation between Physiological Parameters of Automobile Drivers and Traffic Conditions," *International Journal of Electronics Engineering*, vol. 5, no. 2, Dec 2013. [In Press]

[3] Mandeep Singh and Abdullah Bin Queyam, "A Novel Method of Stress Detection using Physiological Measurements of Automobile Drivers," *International Journal of Electronics Engineering*, vol. 5, no. 2, Dec 2013. [In Press]

REFERENCES

[1] The Green Flag Report on “Safe Driving Part Five: Driven to Distraction” (*Brake*, 2006).

Available at: http://www.brake.org.uk/facts/driver-stress.htm

PDF Available at: http://www.brake.org.uk/assets/docs/dl_reports/DLreport3-DISTRACTION-pt1-Dec11.pdf

[2] Healey J. A. & Picard R.W., “Detecting stress during real-world driving tasks using physiological sensors,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 6, no. 2, pp. 156-166, June. 2005.

[3] Shi Y., Ruiz N., Taib R., Choi E. & Chen F., “Galvanic skin response (GSR) as an index of cognitive load”, *CHI’07 extended abstracts on Human factors in computing systems, ACM, New York*, NY, USA, pp. 2651–2656, 2007.

[4] Book Review: Stern R. M., Ray W. J. & Quigley K. S., “*Psychophysiological Recording*”, (2nd ed.), New York: Oxford University Press, 2001.

[5] Rowe D. W., Sibert J. & Irwin D., “Heart Rate Variability: Indicator of User State as an Aid to Human Computer Interaction,” *In Proceedings of Conference on Human Factors in Computing Systems*, vol., no., pp. 480-487, 1998.

[6] Book Review: Cacioppo J. T., Berntson G. G., Larsen J. T., Poehlmann K. M. & Ito T. A., “*The Psychophysiology of Emotion*,” in Handbook of Emotions, J.M. Haviland-Jones, Editor. The Guilford Press: New York, 2000.

[7] Rigas G., Katsis C. D., Bougia P., Fotiadis D. I., “A reasoning-based framework for car driver’s stress prediction,” *Control and Automation, 2008 16th Mediterranean Conference on*, vol., no., pp. 627-632, 25-27 June 2008.

[8] Li Shiwu, Wang Linhong, Yang Zhifa, Ji Bingkui, Qiao Feiyan & Yang Zhongkai, “An active driver fatigue identification technique using multiple physiological features,” *Mechatronic Science, Electric Engineering and Computer (MEC), 2011 International Conference on* , vol., no., pp. 733-737, 19-22 Aug. 2011.

[9] In Cheol Jeong, Dong hee Lee, Shin Woo Park, Jae Il Ko & Hyung-Ro Yoon, "Automobile driver's stress index provision system that utilizes electrocardiogram," *Intelligent Vehicles Symposium, 2007 IEEE* , vol., no., pp. 652-656, 13-15 June 2007.

[10] Singh R. R., Conjeti S. & Banerjee R., "Biosignal based on-road stress monitoring for automotive drivers," *Communications (NCC), 2012 National Conference on*, vol., no., pp. 1-5, 3-5 Feb. 2012.

[11] Yong Deng, Zhonghai Wu, Chao-Hsien Chu & Tao Yang, "Evaluating feature selection for stress identification," *Information Reuse and Integration (IRI), 2012 IEEE 13th International Conference on* , vol., no., pp. 584-591, 8-10 Aug. 2012.

[12] Shi Y., Nguyen M. H., Blitz P., French B., Fisk S., De la Torre F., Smailagic A., Siewiorek D. P., Absi M., Ertin E., Kamarck T. & Kumar S., "Personalized Stress Detection from Physiological Measurements," *International Symposium on Quality of Life Technology*, 2010.

[13] Zhai J., Barreto A., "Stress Detection in Computer Users Based on Digital Signal Processing of Noninvasive Physiological Variables," *Engineering in Medicine and Biology Society, 2006. EMBS '06. 28th Annual International Conference of the IEEE*, vol., no., pp. 1355-1358, 30 Aug - 3 Sept 2006.

[14] Kumar M., Weippert M., Vilbrandt R., Kreuzfeld S. & Stoll R., " Fuzzy Evaluation of Heart Rate Signals for Mental Stress Assessment," *IEEE Trans. Fuzzy Syst.*, vol. 15, no.5, pp. 791-808, Oct. 2007.

[15] Book Review: Carlson R. Neil, "*Physiology of Behavior*", (11^{th} Ed.), New Jersey: Pearson Education, Inc., 2013, ISBN 978-0-205-23939-9.

[16] Wijsman J., Grundlehner B., Penders J. & Hermens H., "Trapezius Muscle EMG as Predictor of Mental Stress," *In: Wireless Health, WH 2010*, 5-7 October 2010, San Diego, California, 2010

[17] Shin J. W., Seong H. M., Cha D. I., Yoon Y. R. & Yoon H. R., "Estimation of stress status using biosignal and fuzzy theory," *Engineering in Medicine and Biology Society, Proceedings of the 20th Annual International Conference of the IEEE*, vol. 3, no., pp. 1393-1394, 29 Oct-1 Nov 1998.

[18] de Santos Sierra A., Avila C. S., Bailador del Pozo G. & Guerra Casanova J., "Stress detection by means of stress physiological template," *Nature and Biologically Inspired Computing (NaBIC), 2011 Third World Congress on* , vol., no., pp. 131-136, 19-21 Oct. 2011.

[19] Bundele M. M. & Banerjee R., "Detection of fatigue of vehicular driver using skin conductance and oximetry pulse: a neural network approach," *In Proceedings of the 11th International Conference on Information Integration and Web-based Applications & Services (iiWAS '09),* ACM, New York, NY, USA, pp. 739-744, 2009.

[20] Akbas A., "Evaluation of the physiological data indicating the dynamic stress level of drivers", *Scientific Research and Essays,* Vol. 6(2), pp. 430-439, 18 Jan. 2011.

[21] Anders Drachen, Lennart E. Nacke, Georgios Yannakakis, & Anja Lee Pedersen, "Correlation between heart rate, electrodermal activity and player experience in first-person shooter games," *In Proceedings of the 5th ACM SIGGRAPH Symposium on Video Games (Sandbox '10),* Stephen N. Spencer (Ed.). ACM, New York, NY, USA, pp. 49-54, 2010.

[22] Goldberger AL, Amaral LAN, Glass L, Hausdorff JM, Ivanov PCh, Mark RG, Mietus JE, Moody GB, Peng CK, Stanley HE. PhysioBank, PhysioToolkit, and PhysioNet: Components of a New Research Resource for Complex Physiologic Signals. Circulation 101(23):e215-e220 [Circulation Electronic Pages; http://circ.ahajournals.org/cgi/content/full/101/23/e215]; 2000 (June 13). PMID: 10851218; doi: 10.1161/01.CIR.101.23.e215

Available at: http://physionet.org/cgi-bin/atm/ATM/

[23] Microsoft. (2010). Microsoft Excel [computer software]. Redmond, Washington: Microsoft.

[24] Salahuddin L.; Kim D., "Detection of Acute Stress by Heart Rate Variability Using a Prototype Mobile ECG Sensor," *International Conference on Hybrid Information Technology*; vol. 2; p.453-459, 2006.

[25] Guanghua Wu, Guangyuan Liu & Min Hao, "The Analysis of Emotion Recognition from GSR Based on PSO," *Intelligence Information Processing and Trusted Computing (IPTC), 2010 International Symposium on* , vol., no., pp. 360-363, 28-29 Oct. 2010.

[26] Healey J. A., "Wearable and automotive systems for affect recognition from physiology", MIT Dept. of Electrical Engineering and Computer Science, Ph.D. thesis, 2000.

[27] Ikuo Homma & Yuri Masaoka, "Breathing rhythms and emotions," *Experimental Physiology*, vol. 93(9), pp. 1011-1021, 1 Sept. 2008.

[28] Mark Hall, Eibe Frank, Geoffrey Holmes, Bernhard Pfahringer, Peter Reutemann & Ian H. Witten (2009); The WEKA Data Mining Software: An Update; SIGKDD Explorations, Volume 11, Issue 1.

[29] Mandeep Singh and Abdullah Bin Queyam, "Stress Detection in Automobile Drivers using Physiological Parameters: A Review," *International Journal of Electronics Engineering*, vol. 5, no. 2, Dec 2013. [In Press]

[30] Mandeep Singh and Abdullah Bin Queyam, "Correlation between Physiological Parameters of Automobile Drivers and Traffic Conditions," *International Journal of Electronics Engineering*, vol. 5, no. 2, Dec 2013. [In Press]

[31] Web References: Nils J. Nilsson, *Introduction to Machine Learning* (online book)

Available at: http://ai.stanford.edu/people/nilsson/mlbook.html

PDF Available at: http://ai.stanford.edu/people/nilsson/MLBOOK.pdf

[32] Mandeep Singh and Abdullah Bin Queyam, "A Novel Method of Stress Detection using Physiological Measurements of Automobile Drivers," *International Journal of Electronics Engineering*, vol. 5, no. 2, Dec 2013. [In Press]