IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received June 7, 2020, accepted June 30, 2020, date of publication July 14, 2020, date of current version July 24, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3009226

Multimodal System to Detect Driver Fatigue Using EEG, Gyroscope, and Image Processing

NAVEEN SENNIAPPAN KARUPPUSAMY[1] AND BO-YEONG KANG[2], (Member, IEEE)
[1]Naveenam Tech Private Ltd., Coimbatore 641062, India
[2]School of Mechanical Engineering, Kyungpook National University, Daegu 41566, South Korea

Corresponding author: Bo-Yeong Kang (kby09@knu.ac.kr)

This work was supported by the National Research Foundation of Korea under Grant NRF-2019R1A2C1011270.

ABSTRACT Sleepiness detection system that evaluates driver's sleepiness level has always been the primary interest of researchers. There are a number of systems like electroencephalography-based sleepiness detection system (ESDS), vehicle based sleepiness estimator system, image acquisition technology and bio-mathematical models to detect drowsiness of drivers. However there has been less research on hybrid of these systems that detect sleepiness of drivers. In order to overcome the above limitation we propose a neural network based hybrid multimodal system that detects driver fatigue using electroencephalography(EEG) data, gyroscope data and image processing data. It was found that the proposed hybrid system performed well with a detection accuracy of 93.91% in identifying the drowsiness state of the driver.

INDEX TERMS Deep neural networks, driver fatigue detection, electroencephalography module, gyroscope module, multimodal system, tensorflow, vision module.

I. INTRODUCTION

Sleepiness or drowsiness of the driver while driving has been the major factor for many accidents according to surveys. The US Foundation for Traffic Safety conducted in 2010 had analysed police reports estimated that 16.5% of fatal crashes involved a drowsy driver. A nationally representative telephone survey showed that 41.0% of drivers admit to have ''fallen asleep or nodded off'' while driving at some point in their lives [1], [2]. Mental fatigue is also one of the major causes of serious accidents, especially in transportation and aviation area which is believed to account for 20% - 30% of all traffic accidents [3]–[5]. This often occurs to drivers who do not get enough sleep, under medication, shift workers, people with sleep disorders and commercial drivers where lot of concentration is required for longer working hours. In the time of emergency the response of the driver is highly important while driving however drowsiness of the driver reduces the reaction time drastically thereby increasing the danger level [6], [7]. It also makes driver's pay less attention to the road and affects the driver's ability to make good decisions.

The associate editor coordinating the review of this manuscript and approving it for publication was Muhammad Asif.

Sleep can be categorised in three main stages according to the American Academy of Sleep Medicine as wakefulness, non-rapid eye movement(NREM) and rapid eye movement (REM) [8]. The NREM can further be classified into three stages as N1- transition from awake to sleep(drowsy), N2-light sleep and N3 deep sleep [9]. This study on driver drowsiness is mainly concentrated on the N1 stage of NREM. The term driver fatigue has been defined as a state of reduced mental alertness, which impairs performance of a range of cognitive and psychomotor tasks, including driving [10], [11]. The terms ''drowsy'' and ''sleepy'' are synonymous with each other and the term ''fatigue'' has been used by researchers in the previous studies instead of ''sleepiness'' [9], [12]. There are many methods proposed to evaluate driver's sleepiness level but these methods could be categorised into five main categories.

The first category for driver sleepiness detection is based on ''subjective measures''. The Karolinska sleepiness scale (KSS) is frequently used for evaluating subjective sleepiness levels where the participating driver has to rate his own sleepiness level during the experiment [13]. Although most of the experiments showed that this method performed well when comparing with the performance of other methods like EEG, a lot of care was taken by the person conducting the experiment [14], [15]. It was also found that the subjects

participating in the experiment often misjudged their own sleepiness level and they were distracted by the attention towards providing feedback to the experimenter.

The second category for driver-drowsiness detection is based on ''Vehicle based systems''. The common warning signs shown by the driver while driving under the influence of drowsiness are drifting from current lane and hitting a rumble strip on the side of the road. In the vehicle based systems the vehicle parameters like deviation from lane, steering wheel movement, force applied on the brake, usage of accelerator and driving behaviour pattern are monitored [16], [17]. Although these vehicle data provide a detailed view of the driver's status, it is greatly influenced by the current driving conditions like weather, traffic, vehicle's performance and driver's emotional status. Hence any change in the monitored metrics may not be due to drivers drowsiness.

The third category for driver sleepiness detection is based on ''driver behavioural based measures''. Several behavioural signs like yawning, facial touching, blinking frequently, eye closure, head posture and a combination of these signs are exhibited by the driver while he is experiencing drowsiness. These symptoms exhibited by the driver are detected using a camera which uses image acquisition and processing technology [18]–[21]. There has been a lot of improvements in the image acquisition and processing technology, however the illumination of the subject, surrounding area, glare from glasses and reflection from mirrors affect the accuracy of prediction. It is also impossible to detect changes if the subject's face goes out of the field of view of the camera.

The fourth category for detection of driver's sleepiness is based on ''human physiological signal based systems''. The physiological features of body are always correlated to the driver's driving behaviour. Several physiological signals like the electrocardiogram (ECG), electrooculogram (EOG), electromyogram(EMG) and EEG were used to study the drowsiness state of the driver [22]–[24]. These signals change according to the alertness state of the driver. Among the above physiological signals EEG was the most commonly studied and used for fatigue. The robustness of the EEG signals depends on the quality of the EEG electrodes, electrode localization, features and classifiers that are used and are most discriminative in classifying the different levels of sleep [12].

The fifth category for detection driver sleepiness is based on ''human physiological signal based systems''. The above methods for detecting drowsiness has its own advantages and limitations. However a combination of one or more of the above methods could complement the limitations in each of its approaches. There are only a few researches conducted to detect driver drowsiness by hybrid techniques [9], [12]. Researches are also moving towards a new generation of driver monitoring systems within the context of Internet of Vehicles System where smart cars collect information from the driver, the road, the car, and the surrounding cars to process the information for integrated safety [25], [26].

Hence in this paper we propose a hybrid multimodal system to detect driver fatigue using EEG for human physiological signal based system, gyroscope data of head motion and image process for behavioural based measures while the entire system was programmed using Google's TensorFlow. TensorFlow is an open source software library for machine learning and deep learning which used flexible numerical computation core and can be used across various platforms like desktop, mobile and edge devices. The experimental results showed that our proposed system performed better in detecting driver fatigue thus proving that the hybrid systems perform better than the individual systems.

II. RELATED WORKS

Driver fatigue, sleepiness, distraction and inattentiveness are some of the studies that were extensively carried out by researchers in this field as cited in these literature reviews [9], [12], [27]. The usage of wearables has also increased drastically in recent years and a significant amount of research has been carried out with the use of these sensors for monitoring human activity as found in [28]. A number of driver's alert systems are currently in use like the ''Driver Alert System'' in Volvo and Ford cars, ''Driver Fatigue Detection System'' in Volkswagen and Skoda cars, where the driver is alerted when driver fatigue is detected based on the changes in driving pattern [29]. However the system is not redundant as numerous factors cause the system to fail like the driving through rutted road surfaces and dust on cameras. Numerous sensors and methods were used by researchers to detect drowsiness, hence in this section we are only discussing multimodal or fusion of different sensors to predict driver fatigue.

Multimodal fatigue measures like surface electromyography(sEMG), EEG, seat interface pressure, blood pressure, heart rate and oxygen saturation level were used to evaluate driver fatigue using different statistical analysis like normality test, Friedman test, linear regression analysis and Wilcoxon-signed rant test [30]. During the study it was found that early onset of fatigue was detected by multimodal measures with the significant change in pressure distribution on thigh and buttocks regions. The drivers behaviours associated with visual and cognitive distractions both separately and jointly were studied with the perceptual scores. They indeed were used to define regression models with elastic net regularization and binary classifiers [31]. Dangerous driving intensity (DDI) framework was introduced wherein fuzzy sets were optimized using particle swarm optimization for modelling driver, vehicle and lane attributes [32]. The results obtained from the DDI analysis are in favourable agreement with those obtained from the perception study.

As EEG was one of the widely researched topics in detecting driver drowsiness. There were a few studies conducted by combining EEG with other sensors. EOG, ECG and EEG signal readings were measured simultaneously during simulated driving task, where the subject's status was classified as awake and drowsy state using Fisher's linear discriminant analysis [33]. In another method forehead EOG was used to extract various information on eye movement

TABLE 1. Important research papers that use multimodal signals in their study for objective other than fatigue detection.

Analysis methods	Bio-signal Type	Objectives	Results
SVM [3]	Respiratory and eight channel EEG are gathered	Drivers vigilance level	The SVM(desktop PC) shows an accuracy of 98.6% while that of mobile phones shows an accuracy of 98.5%
SVM [37]	ECG and EEG	Driver Drowsiness	The results show that a single EEG channel (specifically either O2 or P7) was able to detect driver drowsiness effectively, when combined with an ECG channel, providing more than 80% accuracy.
SVM and AdaBoost [38]	Eye gaze, head orientation, diameter of pupil and heart rate (RRI)	Cognitive distraction detection	The cognitive load for conversation and arithmetic of SVM was found to be 91.7% and 92% while that for Adaboost was found to be 93.0% and 92.2%
Standard statistical techniques and high dimensional clustering [39]	ECG, EEG, EOG, skin impedience at lower arm, skin temperature, pulse and oxygen saturation in the blood, respiration frequency and head movement	Lane tracking behaviour analysis	The relationship between different behavioural symptoms of fatigue and driving were studied like sleep and accidents, eye and eyelid movement with lane tracking behaviour. The results showed there were significant relation between all of the factors.
Modality fusion with temporal dependency [34]	Forehead EOG and EEG	Vigilance Estimation	The results show that the posterior EEG and forehead EOG have important complementary characteristics.
Multi class kNN classifier and different regression models [40]	Camera, microphone array, brake, gas acceleration, vehicle speed, steering wheel angle, frontal camera, driver camera and GPS.	Driver distraction	The coefficient of determination for this model is $R^2 = 0.53$, which corresponds to a correlation of $p = 0.728$.

TABLE 2. Important research papers that use multimodal signals in their study for fatigue detection.

Analysis methods	Bio-signal Type	Objectives	Results
Bayesian Network with Hidden Markov Models (HMMs) [36]	Arm position, gaze estimation, eye closure, face orientation, facial expressions, voice activity detection, speech analysis, heart rate, clutch, brake, gas, and steering wheel.	Fatigue and distraction assessment	Accuracy without prior information 97.3% and with prior information is 98.4%.
Dynamic bayesian network with first order HMMs [41]	Sleep time, sleep quality, circadian rhythm, work environment, eye movement, ECG signals and EEG signals.	Driver fatigue recognition	The average square error with EEG and ECG is 1.2629 which showed better performance than with the absence of the above signals.
Artificial Neural Networks(ANN) [42]	Vehicle parameter data which is steering angle and eye closure data	Fatigue detection	Accuracy for truck study with only steering data 85%, steering and eye data 88%, Car study only steering data 86%, steering and eye data 92%.

features and combined with EEG as EOG contains complementary characteristics for vigilance estimation [34]. EEG and respiration signals of the driver in time and frequency domain were recorded in real time with a Bluetooth based mobile technology and analysed using support vector machine(SVM) [3]. Real-time drowsiness detection is hard as high number of data is received from various sensors and the method used to predict real time drowsiness is not light enough or it requires high computing power. Hence in this method various optimized indicators based on driver's physical and driving performance measures are obtained from Advanced Driver Assistance Systems (ADAS) in simulated driving conditions and a new method of ground-truth generation based on a supervised Karolinska Sleepiness Scale(KSS) was used [35]. In order to combine various sensors researchers extracted audio, color video, depth map, heart rate, steering wheel and pedal position information which were then processed according to three modules of vision, audio and other signals. The output from these modules were then combined with contextual information and a dedicated Bayesian network was designed to predict both fatigue and distraction [36].

Table.1 provides information on the multimodal data acquisition and processing techniques used for achiving objectives other than driver fatigue detection. In Table.2 the multimodal techniques used for driver fatigue detection are explained along with the results.

From the literatures we can see that multimodal driver fatigue performs better and are reliable than individual signal based prediction [43]. However due to high number of data received from different sensors and the complexity of the algorithms developed, required large computing power and sophisticated signal transmission techniques. Hence we propose a multimodal system that predicts driver fatigue from EEG, gyroscope and image processing data using Neural networks.

III. MULTIMODAL SYSTEM ARCHITECTURE

In order to build a more robust and reliable driver fatigue detection system multimodal system to detect driver fatigue using Deep Neural Networks (DNN) was proposed. The system consists of multimodal time series data that is acquired from EEG, gyroscope and image processing systems from the driving simulator platform. This data is then processed individually in three seperate modules where the EEG module predicts driver drowsiness, gyroscope module predicts driver head activity and vision module predicts driver facial behavioural signs. The predictions from these modules are then given as input to the multimodal driver fatigue detection DNN which processes the data and then predicts the

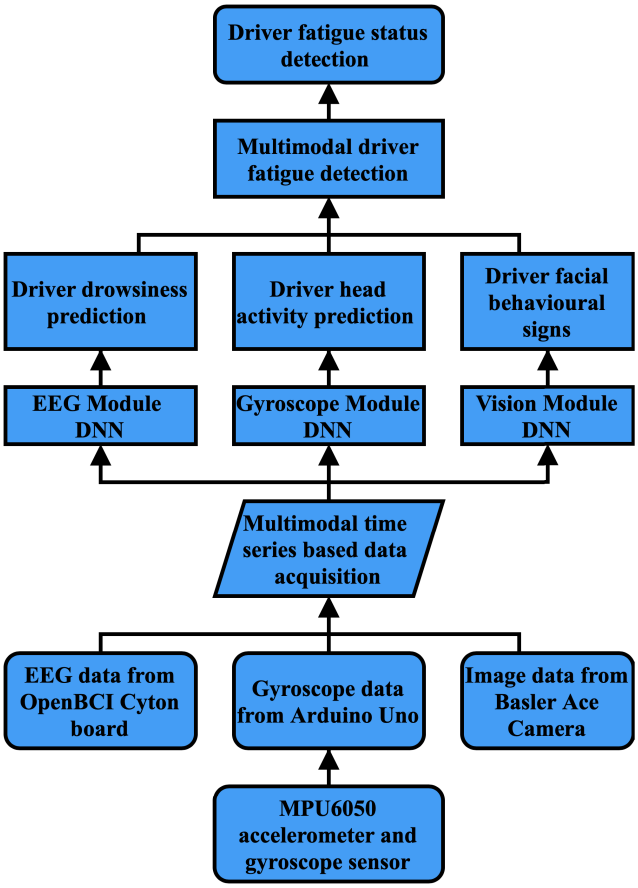


FIGURE 1. Methodology of the proposed multimodal system to detect driver fatigue.

driver fatigue status. The methodology of this proposed multimodal system to predict driver fatigue is as shown in Fig.1. Each module comprises of a seperate NN which is designed according to the data type that is acquired from each sensor and predicts a value according to its drowsiness detection system which is explained in the following sections. The main advantage of the proposed multimodal driver fatigue detection system is that it not only predicts if the driver is experiencing fatigue state but also the physiological and behavioural characters exhibited by the driver under fatigue. It also helps in adding additional modules based on the type of sensor data being introduced. The advantage in the use of individual modules like EEG module(EM) enables an accurate detection of drowsiness and awake state experienced by drivers with the use of DNN which is developed according to the specific datatype. Similarly the head motion of the driver could be found using Gyroscope module(GM) DNN and face behavioural signs could be found with the help of vision module(VM) NN. Fatigue in drivers is mainly caused due to prolonged driving over monotonous and repetitive environments, while sleepiness or drowsiness occurs in professional drivers during end hours of all night driving shift [11]. In the proposed method, fatigue is detected with the help of behavioural signs exhibited by the driver like rapid blinking, nodding of head, yawning while eyes closed, frequently being distracted by adjusting driver console, eyes getting closed and yawning by using gyroscope sensor and image processing. EEG signals are used to detect the transition between awake and sleep or the drowsy stage. Behavioural signs like yawning, yawning while eyes closed and rapid blinking are mostly triggered involuntary actions due to fatigue or sleepiness. The drawback of one system is complemented by the advantage of other system. The data acquisition technique are explained in the following sections.

In order to effectively detect driver fatigue it is essential to detect driver drowsiness effectively. The sleep wake cycle is one of the major factors of driver fatigue along with other factors like health, physical injuries, regular food intake that contribute to physical fatigue of a driver. Some of the factors that contribute to the mental fatigue are monotonous driving, boredom, and so on. Hence for this research two states of the driver are considered, which are i) normal driving state, where the driver has had proper prior sleep and is not physically ill in any way. The other state where the driver is under ii) fatigue driving state, where he has not had a minimum of 8 hours of sleep during the normal sleep cycle within 24 hours, is physically ill or is experiencing any one of the factors contributing to fatigue. Therefore the two main categories considered are normal driving state and fatigue driving state of the driver.

There are many researches conducted to measure the variables like performance, perceptual, electrophysiological, physiological and biochemical measures which are associated with fatigue but drowsiness is considered as the main indicator of fatigue present in a driver and in previous studies mentioned by the researchers in [11]. As for the case of the multimodal research the EEG, gyroscope and image processing data is considered. In the case of EEG the physiological data from the driver is categorised into awake and drowsy state of the driver. In the case of gyroscope the behavioural signs of drowsiness exhibited by the driver are categorised into nodding state where rapid head movement is exhibited by the driver along with frequent changes made to the driver console like adjusting the volume of music, changing the climate controls of the car, and so on. This behaviour is interpreted by the gyroscope present in the driver headset for tracking the head pose of the driver by means of section numbers where the field of vision of the driver is divided into different sections as shown in Fig.2. Section 5 is the region located where the angle of rotation of the head is about 20 degree to either side along the x axis and y axis from the mean position. The other sections are consecutively divided as shown in Fig.2. Hence, the head pose of the driver is categorised according to section numbers from 1 to 9 and 0 for regions outside this zone or for signal loss.

In the case of image processing behavioural signs of drowsiness like yawning, eyes closed and yawning while eyes closed are detected. While all of the above signs of drowsiness from EEG, gyroscope and image processing may happen even in normal driving state but the likelyhood and frequency with which it happens increases with the level of fatigue

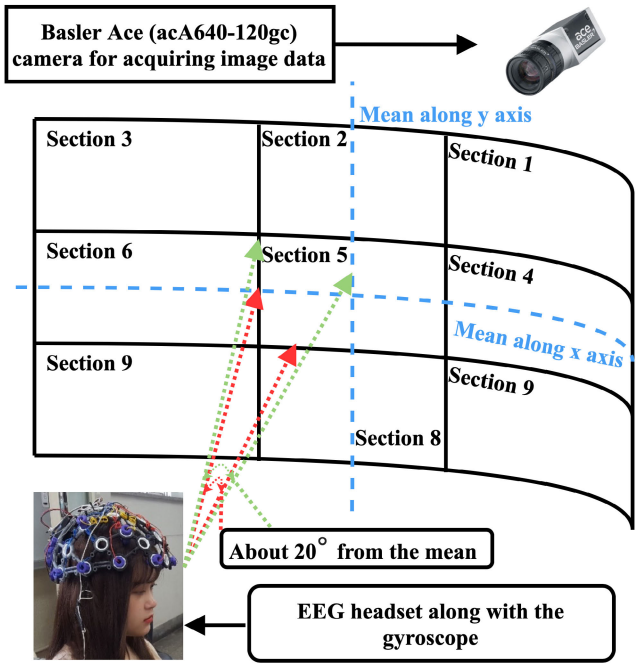


FIGURE 2. Multimodal System to detect driver fatigue.

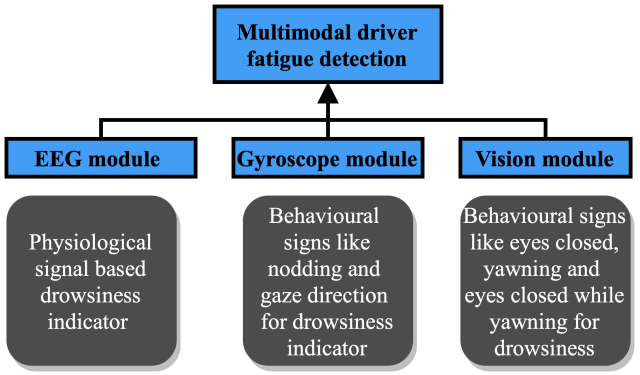


FIGURE 3. Multimodal based indicator of driver fatigue.

experienced by the driver [42]. Therefore for multimodal driver fatigue detection system a deep neural network that uses the prediction from EEG module, gyroscope module and vision module as inputs is proposed as shown in Fig.3. These individual modules contain neural networks that are specifically designed to the data type being recorded from the sensors.

The raw data for the multimodal fatigue detection is collected with the help of five different test subjects who are aged between 21 and 30. All the test subjects are license holders and regular drivers who normally commute by car. The data was recorded at two different times in a driving simulator platform, the first is when the test subject is in a normal driving state and the second is when the test subject is in a fatigue state where the reading is recorded between 3 pm and 5 pm when the driver is more prone to drowsiness. In order to record the fatigue state the test subjects had less than

TABLE 3. Predictions from EEG, gyroscope and vision module.

Module Name	Description of Prediction	Value of Prediction
EEG module	Awake	0
	Drowsy	1
Gyroscope module	Outside the sections	0
	Section Numbers	1 to 9
Vision module	Normal	0
	Yawning	1
	Eyes closed	2
	Eyes Closed while yawning	3
	Absence of data	4

6 hours of sleep the previous night in the 24 hours cycle and were deprived of their sleep. The raw data from the different sensors attached to the driver headset were collected with the help of LabVIEW software and processed using the EEG, gyroscope and vision modules. The module's name along with the description of prediction and the value of prediction is as shown in Table.3.

The raw data was recorded under five different scenarios. In the case of first scenario the raw data from the test subject is recorded under normal driving condition while performing the act of checking the time while driving. In the case of second scenario the raw data is recorded while the test subject is in fatigue state and is requested to perform the act of yawning and eyes closed while yawning. In the case of third scenario the raw data is recorded while the test subject is in fatigue state and is requested to perform the act of nodding by moving the head up, down left and right. During this scenario they were also asked to change the temperature of the car in the console. The above actions of yawning and nodding only happen in drivers when they are experiencing extreme fatigue. It may not occur to some drivers as it is highly dependant on individual's behaviour. Hence these actions were simulated by the test subjects while the scenarios were recorded [36]. The fourth scenario was recorded while the test subject is in fatigue state and is extremely drowsy. The drowsy and fatigue state of the driver were also verified with the help of videos recorded from the session. In the fifth scenario the data was recorded while the driver is experiencing extreme fatigue and falls asleep while driving in the simulator. The scenario description and the status of the driver are given in Table.4.

The multimodal data that is recorded from the scenarios are analysed with the help of the three individual modules and their predictions are found out individually. In the case of results from the vision module, the predicted values are stored according to the corresponding time and if there is an absence of data then it is tagged as "4" for vision module. All the data from the first scenario are tagged 0 except for the data that has detected drowsiness from the EEG module which is tagged as "1". Hence by tagging the first scenario as "0" we have recorded all the characteristics from EEG, gyroscope and vision module for safe driving in normal state of the driver. So by tagging the data as "1" that has shown drowsiness from the EEG sensor, we are recording the

TABLE 4. Status of driver along with the description of scenario.

Scenario Number	Driver status	Description of Scenario
1	Normal state	Simulating the act of checking the time and adjusting the air-con while driving
2	Fatigue state	Simulating the act of yawning and eyes closed while yawning while driving
3	Fatigue state	Simulating the act of nodding, checking time and adjusting the aircon while driving
4	Fatigue state	Driving while extremely drowsy which is verified with the help of videos recorded during the session
5	Fatigue state	Drowsy to sleeping with eyes closed while driving [42]

TABLE 5. Multimodal tagging information with respect to scenario.

Scenario Number	Driver status	Tagging description based on drowsiness predicted with EEG module	Tagging value
1	Normal state	Awake while driving	0
		Drowsy while driving	1
2	Fatigue state	Awake while driving and performing the act of yawning	0
		Drowsy while driving and performing the act of yawning	2
3	Fatigue state	Awake while driving and performing the act of nodding	0
		Drowsy while driving and performing the act of nodding	3
4	Fatigue state	Awake while driving	0
		Drowsy while driving	4
5	Fatigue state	Awake while driving	0
		Sleeping while driving with eyes closed	5

characteristics of drowsiness even in normal state of the driver, as drowsiness can happen at any point of time.

In the case of the second scenario the predicted data that was recorded during the fatigue state and performing the act of yawning of the driver is tagged as ''2'' when the driver is detected as experiencing drowsiness with the predictions of EEG module and all the other data are tagged as ''0''.

As for the case of third scenario predicted data was recorded when the driver is under fatigue stage and performing the act of nodding. Hence, based on the predictions from the EEG module if the driver is experiencing drowsiness then all the data are tagged as ''3'' and the other data are tagged as ''0''.

Similarly for the fourth scenario the data is recorded while under fatigue state and the driver is found to be extremely drowsy. Hence based on the EEG predictions if the driver is found to be drowsy then the data is tagged as ''4'' while the other datas are tagged as ''0''.

Finally as for the fifth scenario the datas are recorded when the driver is in fatigue state and is sleeping while performing the act of driving. In this case based on the predictions of EEG module, if the driver is found to be drowsy then the data is tagged as ''5'' while all the other data are tagged as ''0''.

The multimodal tagging information with respect to scenario in which the data is recorded is given in Table.5. The combinations of these information in tagging gives a detailed information on the fatigue characteristics with respect to the behavioural signs or physiological signs expressed by the driver. It also gives information on the drowsiness experienced by the driver even in the case of normal state of the driver. Prolonged continuous driving even in normal state may lead to fatigue condition. Hence by tagging datas based on these properties the characteristics of datas from each of the sensors are found out and also the predictions gives us an inside view of the behaviour expressed by the driver while experiencing drowsiness. Scenarios 4 and 5 also give an insight on the behavioural characteristics of the driver when he is experiencing drowsiness in normal state and fatigue state while performing the act of driving.

A. FATIGUE DETECTION BASED ON THE PROPOSED MULTIMODAL DNN

The structure of the proposed DNN is as shown in Fig.4. The initial raw data from the multimodal data acquisition using driver monitoring platform as explained in the following sections of this chapter is processed individually in their respective EEG module (EM), gyroscope module (GM) and vision modules (VM) for prediction. The predicted results are also time series based and given as input to the proposed DNN.

The input consists of three neurons and the number of previous time instances is chosen as three because the safe reaction time required for the driver to react in case of emergency is three seconds [44]. This history of the previous time instances transfers the information of the previous time predictions. If the number of previous time instances are increased then the sensitivity of prediction is decreased and if the number of previous time instances are decreased then the sensitivity of prediction is increased [42]. This data is given as input to the recurrent neural network(RNN) which consists of 400 long short-term memory cells(LSTM) with one dimension. The output from the RNN are then given as input to the NN. The NN also consists of 400 neurons. The NN consists of three hidden layers and all the hidden layers utilise the rectifier linear unit (ReLU) along with dropout function. The output layer of the NN utilises sigmoid function(SF) to classify the six different physiological and behavioural signs of fatigue and normal state of the driver. All layers of the DNN except the RNN utilises the Xavier initilization (XI) with dropout function and adam optimization (AO) for fast optimization in all the layers of the DNN.

The cost and accuracy of prediction for the proposed DNN were calculated. The cost of the prediction is calculated by finding the mean of the sigmoid cross entropy with logits. The

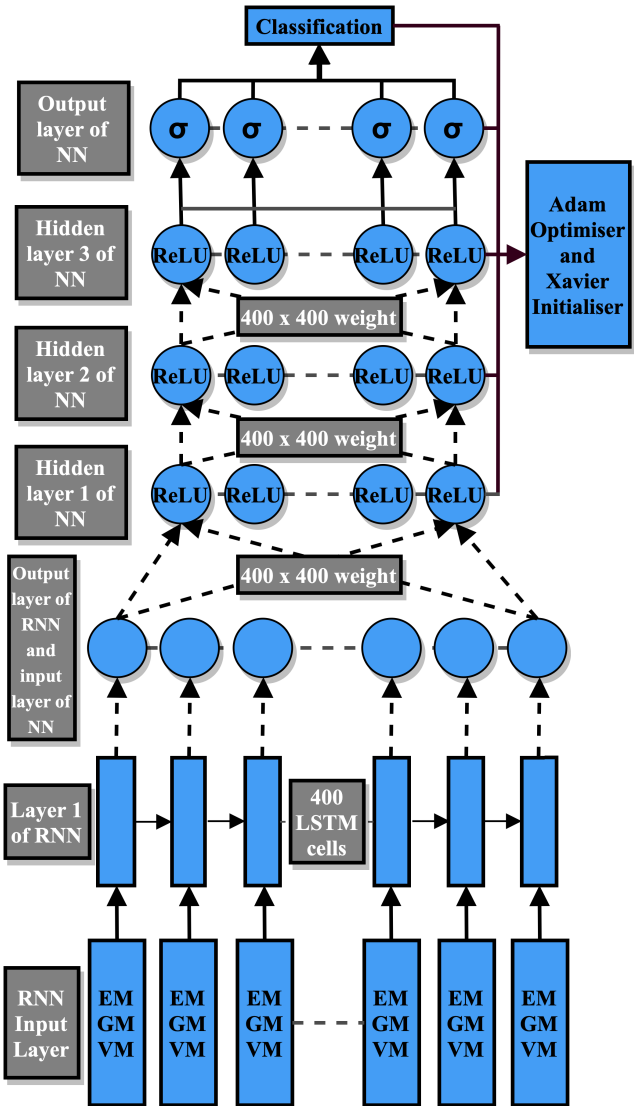


FIGURE 4. Structure of the proposed deep neural network for multimodal classification. The dashed arrow indicate connections where dropout is applied and the solid lines indicate connections where dropout is not applied.

sigmoid cross entropy with logits measures the probability error in discrete classification tasks or multilabel classification in which each class is independent and not mutually exclusive. Accuracy of prediction is calculated by finding the mean between the predicted value of the proposed multimodal DNN and the actual value of prediction. The tagging information or the actual value of prediction at that instance is taken as the mode of the tagged data of all the previous time instances that are considered and is converted to one hot method of tagging to equalise the tagged values so that it does not affect the results by misrepresenting to the DNN that greater the value of tagging then greater is the importance of the tag. Thus with the cost and accuracy functions the performance of the proposed multimodal DNN could be verified. In order to verify the performance of proposed DNN the input from one of the modules was left for testing. A combination of EM along with GM and EM along with VM were considered for analysing the performance of the proposed DNN where only the number of input nodes is changed from three to two. As EEM is considered as the most important dataset for predicting fatigue it is considered in both the combination of modules [11]. The tensorboard generated by the proposed multimodal DNN is as shown in Fig.5

B. EEG MODULE

1) EEG HEADSET

EEG is one of the most widely accepted and used methods for measuring human physiological signal based systems as it is a good indicator of the transition between wakefulness and sleep [27]. It can also detect the transition between different stages of sleep. The inhibitory and excitatory postsynaptic potentials generated by communication between cortical nerve cells are summated in the cortex and extended to the scalp surface which are recorded by the EEG electrodes. There are two types of recording the EEG signals. The monopolar method records the voltage difference between an active electrode placed on the scalp with the reference electrode placed on the ear lobe. The bipolar electrode records the voltage difference between two active scalp electrodes. An amplitude of 10 μ V to 100 μ V along with frequency ranging from 1 Hz to 100 Hz is usually measured with EEG signals. Electrode placements are devised based on the relationship between the location of an electrode and the underlying areas of the cerebral cortex. According to the constitution of the international federation of societies for electroencephalography and clinical neurophysiology, the standard and widely used 10-20 electrode placement system was developed. However in recent times the development of multi-channel EEG hardware systems has lead to the development of higher density electrode settings such as 10-10 and 10-5 systems. For our study we have used the 8 channel 10-20 electrode placement system as shown in Fig.6.

The electrodes used for signal acquisition are dry spikey and non spikey electrodes. The spikey electrodes are placed in the regions where there is lot of hair like the C3, C4, P7, P8, O1 and O2 while the non-spikey electrodes are placed in the region of forehead like the FP1 and FP2 as shown in Fig.6. The ground electrode is placed on the right earlobe while different characters refer to each lobe region which is described as follows: frontal lobe(F), temporal lobe(T), central lobe(C), parietal lobe(P) and Occipital lobe(O). These electrodes where then fixed to the Open BCI Mark IV headset. The OpenBCI Mark IV headset was 3D printed in house using a Finebot FB-Z420 3D printer and assembled. We used an OpenBCI Cyton board which is a 32 Bit 8 channel device that is capable of measuring and recording electrical activity produced by the EEG, EMG and EKG. This board uses ADS1299 ADC biopotential measurement integrated chip developed by Texas Instruments. The wires from these electrodes where connected to the OpenBCI Cyton Board to measure and record the EEG signals. The final image

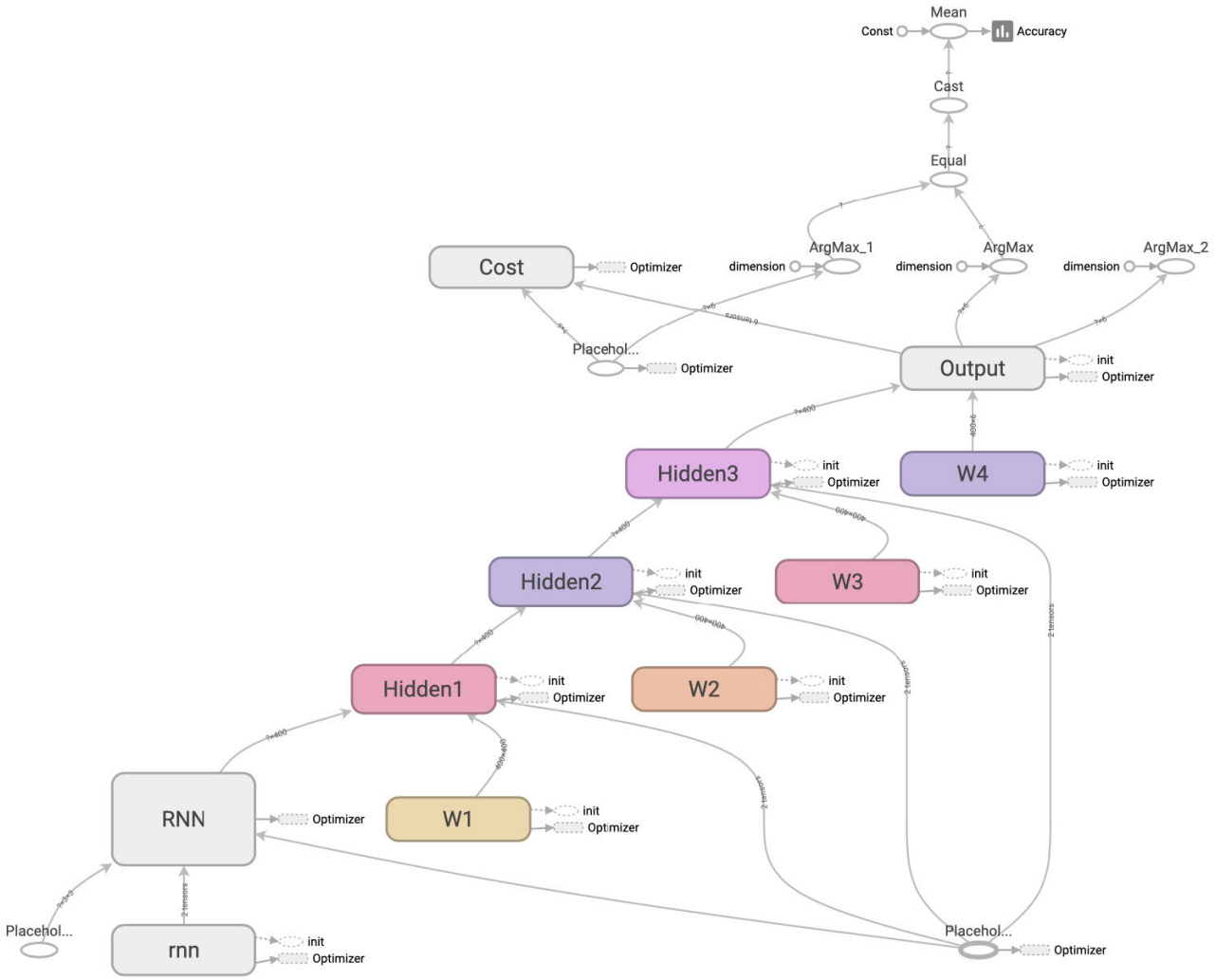


FIGURE 5. Tensorboard graph of the proposed multimodal DNN.

of the assembled headset is as shown in Fig.7. The signal from the Cyton board is then transferred to the computer via OpenBCI Dongle which is capable of communication with RFD22301 radio module.

2) EEG BASED INDICATOR OF FATIGUE

The measured EEG signal can be divided into alpha, beta, delta and theta sub-bands based on their frequency and amplitude. Alpha waves (8-13 Hz) are predominately found in adults who are awake but relaxed with their eyes closed. It is normally associated to N1 stage of sleep where it is the transition from awake to sleep(drowsy). When the subjects open their eyes or if they are disturbed through external factors the amplitude of alpha waves get diminished. Hence alpha waves are the most associated ones to detect drowsiness or fatigue. Beta waves (13-30 Hz) are associated with individuals who are alert, aroused and excited. Delta waves (1-5 Hz) and Theta waves (4-8 Hz) are normally associated with adults who are in deep sleep. The occurrence of these low frequency signals increase and alpha waves diminish as the subjects move from light sleep or drowsy state to deep sleep.

The main aim of this EEG module is to detect the occurrence of alpha waves effectively and efficiently predict the onset of sleep by analysing the signals using continuous wavelet transform (CWT) and to verify the system using fast Fourier transform(FFT). When the driver falls into micro-sleep or drowsiness the amplitude of alpha waves increases and this phenomenon is well observed in the occipital region or occipital lobe where the O1 and O2 electrodes are present. This phenomenon may occur while the eyes are briefly closed or when the rate of blinking is also higher which can be observed by the fluctuation of signals from the FP1 and FP2 electrodes.

3) EEG DATA COLLECTION, ANALYSIS AND TAGGING USING CWT AND FFT

The 8 channel EEG data required for analysis were collected from three test subjects. The signals were recorded during late

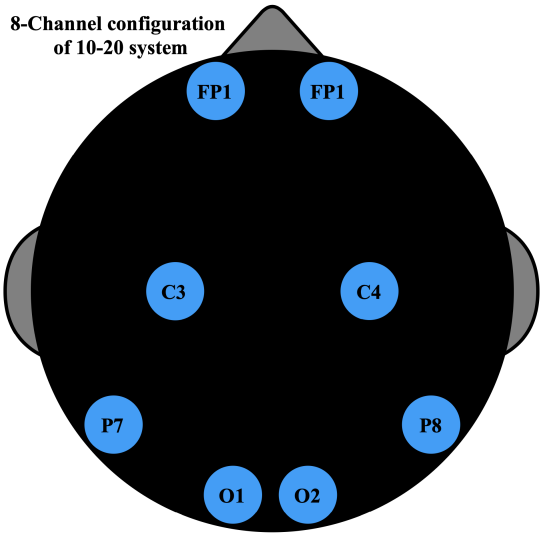


FIGURE 6. 8 channel 10-20 EEG electrode placement system.

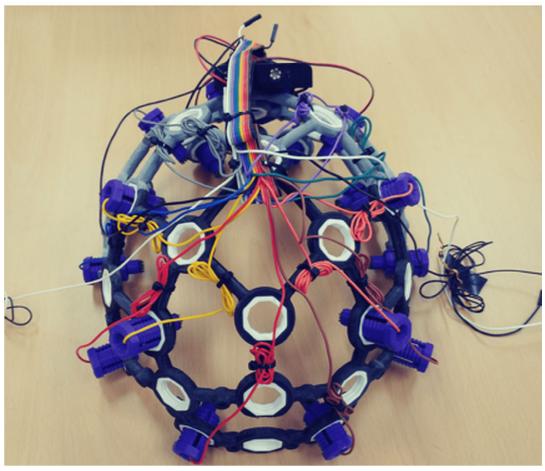


FIGURE 7. EEG Headset used for measuring EEG signals.

night from 11 pm to 2 am in a simulated environment when the drivers are most prone to sleep. The test subjects have prior driving experience and are regular drivers who normally commute by car. The consumption of alcohol was prohibited from 24 hours prior to conducting the test so as to minimise external factors from influencing the test results.

The signals from the occipital regions of O1 and O2 were analysed using CWT. Recently wavelet transformation is a popular method for analysing signals that often exhibit abrupt changes like the real world signals. FFT does not represent abrupt changes efficiently because it represents the data as a sum of sign waves and they are not localised in time and space. Hence unlike FFT which can only perform frequency analysis, CWT enables simultaneous analysis of time and frequency. Wavelet coefficients are obtained by correlating the signal with a wave of finite duration and energy or a wavelet. The mother wavelet is a reference wavelet, whose coefficients are evaluated for the entire range of dilation and translation factors while shifting continuously along the time scale for evaluating the set of coefficients at all instances of time [45]. Morlet function was chosen as the mother wavelet for our application as there was high similarity with EEG signals.

TABLE 6. Tagging information for EEG Module.

Module name	Description of prediction	Value of prediction
EEG Module	Awake	0
	Drowsy	1

The wavelet coefficients in the 10 to 13 Hz interval, which is the high frequency band of alpha waves are extracted. We selected 5 μ V range which is the mean value for detecting the microsleep region and if the average value of the previously calculated amplitudes is greater than 5 μ V, the corresponding section is tagged as '1' for sleep and vice versa as '0' for awake as shown in Table. 6. Similarly the EEG raw data is analysed using FFT where the mean value of alpha wave was analysed for every interval of 1 second and similar parameters for tagging were followed as for that of CWT. Thus the tagged data from CWT along with the 8 Channel EEG will be given as inputs for the deep neural network (DNN) programmed in Tensorflow for detecting drowsy or awake condition while the tagged data from FFT will be used to verify the performance the proposed DNN.

4) FATIGUE DETECTION BASED ON PROPOSED EEG MODULE DNN

The structure of the proposed DNN is as shown in Fig. 8. The proposed DNN was developed in TensorFlow as it processes the data's as Tensors and retains the initial information throughout the processing.

As the data type from EEG is a time series data recurrent neural network (RNN) was initially used and the output was given as input to the next layer of neural network (NN). The raw data from the EEG is given as input to the RNN which consists of 200 long short-term memory cells(LSTM) with one dimension. The LSTM also utilises a dropout function as it greatly reduces overfitting in LSTM's. The LSTM are well suited for time series based data classification, processing and prediction. The output from the RNN are gathered and are given as input to the NN. The NN consists of 100 nodes or neurons and the dimension is reduced from 200 to 100 when it is connected from RNN to NN. The NN consists of three hidden layers and all the hidden layers use rectifier linear unit (ReLU) along with the dropout function. The trends or features pattern from the data are captured with the help of activation function like ReLU which is a deciding parameter.

The use of ReLU enables quick training compared to other activation functions and is not subjected to the problem of vanishing gradient. The only problem is that the mean output is not zero and it introduces a bias for the next layer of the

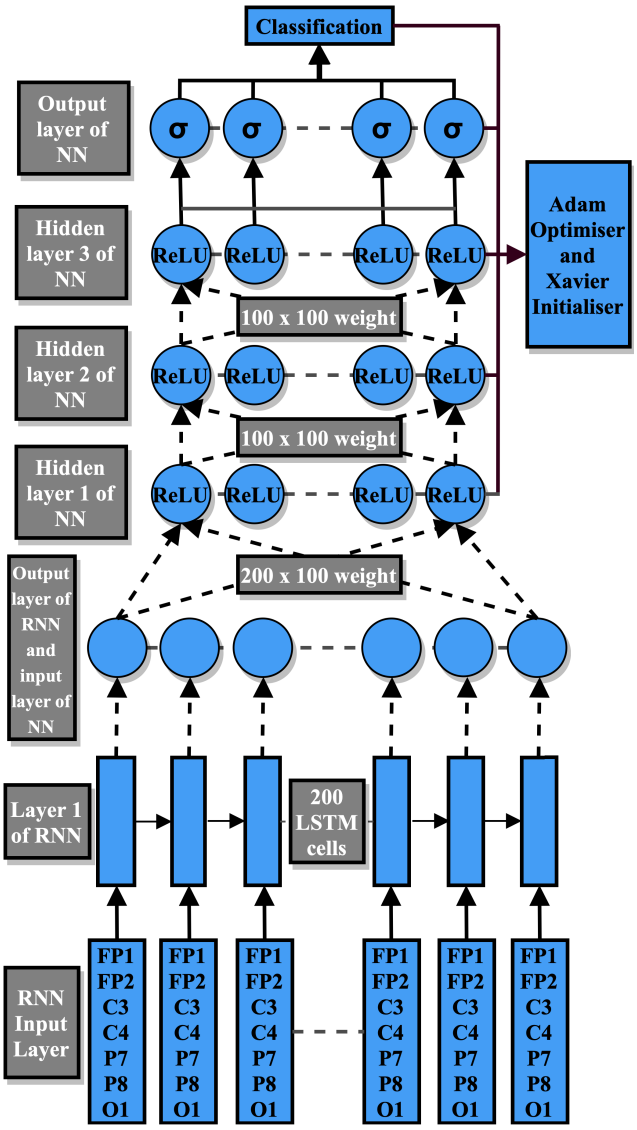


FIGURE 8. Structure of Proposed Deep Neural Network. The dashed arrow indicate connections where dropout is applied and the solid lines indicate connections where dropout is not applied.

neural network where it can slow down the learning process. Thus after three hidden layers the neurons are connected to the output layer where one node is present. The output layer of the NN uses sigmoid function (SF) to classify the fatigue state of the driver. The SF consists of a "S" shaped curve or a sigmoid curve. The SF has similarities with the step function or threshold when the value is a large positive number the output of SF is near to 1.

All layers of the deep neural network except the RNN uses Xavier initialization(XI) with dropout function. The XI automatically determines the scale of weight initialization based on the number of input and output neurons. The XI helps the data's to reach deep into the network. The XI also makes sure that the weights are distributed efficiently for faster learning and it works with almost all types of neurons. Adam optimizer(AO) also known as adaptive moment estimation was used for fast optimization in all the layers of the DNN. The learning rates are calculated and the momentum changes are also stored in the AO. The first moment mean and the second moment the un-centered varience of the gradients respectively are calculated and those values are used to update the parameters in the AO which usually outperforms the other optimization techniques.

The cost and accuracy of prediction for the DNN were calculated. The cost function is defined as given in the below equation:

$$H(x) = \frac{1}{e^{Wx+b}} \tag{1}$$

$$c(H(x), y) = \begin{cases} -log(H(x)) & \text{if y} = 1, \\ -log(1 - H(x)) & \text{if y} = 0. \end{cases} \tag{2}$$

$$Cost(W) = \frac{1}{m} \sum_{c} (H(x), y) \tag{3}$$

where $H(x)$ is sigmoid function, W is weight, x is input, b is bias, y is our prediction, and m is total input number. Thus $Cost(W)$ gives the cost for predicting driver fatigue based on each epoch. Accuracy of prediction is calculated by finding the mean between the predicted value of the DNN and the actual value of prediction. Thus with the cost and accuracy functions the performance of the proposed DNN could be verified.

C. GYROSCOPE MODULE

1) GYROSCOPE SYSTEM

The gyroscope sensor was used for head pose estimation as it provides robust data unlike image processing based head pose estimation. Usually head orientation in image processing is measured with the help of facial features hence when the facial features are not visible to the camera then the system does not work. The variation in lighting also effects the head pose estimation. Hence in order to avoid such problems we have used gyroscope sensor for head pose estimation. The gyroscope sensor used for our proposed system is the Arduino based MPU-6050 Accelerometer plus Gyroscope sensor. The sensor contains hardware for 16-bits analog to digital conversion therefore it can record the x, y and z axis channels at the same time. It can measure the three axis accelerometer, gyroscope, roll, pitch and yaw data. The gyroscope sensor is then connected to the Arduino Uno board which used to stream real time data to the computer.

2) GYROSCOPE MODULE BASED INDICATOR OF FATIGUE

The gyroscope module is used to detect driver fatigue based on the behavioural signs. A driver under fatigue normally exhibits behavioural signs like nodding the head frequently, distraction by operating different consoles in the car and swaying of the head [46]. Hence for our study we have considered the above factors where the motion of the head is tracked using the gyroscope sensor.

TABLE 7. Tagging information for gyroscope module.

Module Name	Description of Prediction	Value of Prediction
Gyroscope module	Outside the sections	0
	Section Numbers	1 to 9

3) GYROSCOPE DATA COLLECTION, PREPROCESSING AND TAGGING

The gyroscope sensor connected to the Arduino Uno board is used to stream real time data to the computer. Matlab is then used to record the real time data and a visualization model was developed to verify the orientation of the head and the observed raw data. The volunteer driver's head motion was then recorded according to the scenario explained as follows. The field of vision of the volunteer driver is divided into different sections as shown in Fig.2. The sections are divided according to the angle of rotation of the head from the mean x and y axis. The intersection of the mean x and y axis are fixed at the normal orientation of the head or when the volunteer driver is looking straight at the road. The mean x axis is the line drawn along the horizontal rotation of the head and the mean y axis is the line drawn along the normal rotation of the head along the vertical axis. The sections are divided at an angle of 20 degrees rotation along the mean x and y axis on either side. The consecutive sections are divided with reference to the central section or section 5. The method of data collection was chosen as the angle of rotation of the head is the same for all the volunteer drivers, only the field of vision of the drivers change according to their body physique. The test subject is then made to look at the sections while the raw data was recorded separately for each sections and are tagged according to the section number as described in Table.7. The raw data recorded consists of roll, pitch, yaw, accelerometer and gyroscope data for three axis. Around 5000 instances each containing 9 data types for each section was recorded. The mean, standard deviation (SD) and principal component analysis(PCA) were then calculated seperately for accelerometer data's, gyroscope data's, for roll, pitch and yaw data's along with the combination of all the above data's separately each of these instances. Therefore the head activity data set along with tagging information of the section was created with 22 different data types. Hence the raw data that consisting of 9 data types were then converted to 21 different data types along with tagging information. In order to verify the head activity data the collected data was verified using fine Gaussian support vector machine(SVM) in classification learner toolbox provided in MATLAB. All the data's of head activity dataset are used as predictors except the section information which was used as response. The result from fine Gaussian SVM was used to verify the redundancy of the data collected.

4) FATIGUE DETECTION BASED ON THE PROPOSED GYROSCOPE MODULE NN

The head activity data is given as input to the proposed NN. The structure of the proposed vision module neural network

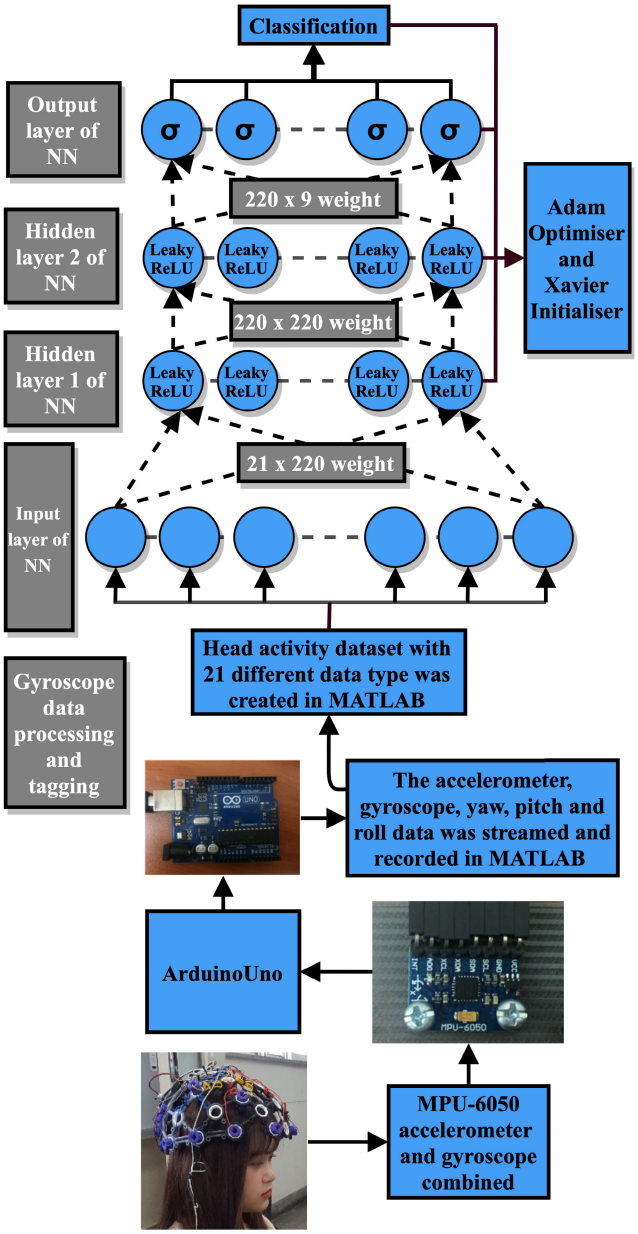


FIGURE 9. Structure of the proposed gyroscope module DNN. The dashed arrow indicates droupout function.

system is shown in Fig.9. The input layer of DNN consists of 21 input nodes each representing the values of mean, SD and PCA calculated for each raw data type. The output layer consists of 9 output nodes where each node represents each section of the drivers visible region. There are two hidden layer with 220 neurons between the input and output layers of the DNN. All the layers of the proposed DNN utilize leaky ReLU as activation function except the output layer which uses sigmoid function. All the layers of the DNN utilizes the XI and dropout function. AO was used for all layers of the DNN as it enables fast optimization.

The cost and accuracy of prediction for the proposed NN were calculated. The cost of prediction was calculated

by finding the mean of sigmoid cross entropy with logits and accuracy of prediction is calculated by finding the reduced mean between the predicted section number of the proposed gyroscope module NN and the actual value of section number. Thus with the cost and accurasy functions the performance of the proposed vision module NN could be verified.

D. VISION MODULE

1) CAMERA SYSTEM

The use of image processing has been in practice since the inception of research into driver fatigue analysis. For our research we used the Basler Ace (acA640-120gc) camera fixed with a compact fixed focal length 8.5 mm lens (781414-01 lens). We used the Pylon viewer software provided by the Basler to acquire the images. The Basler Ace camera is equipped with a charged-coupled device (CCD) based Sony ICX618 sensor capable of capturing pictures with a 658 x 492 pixel in size and of portable network graphics (PNG) format. The camera was connected to the computer with a Gigabit Ethernet (GigE) cable and then the pylon viewer software was configured to capture images at the rate of 5 frames per second for 60 seconds. The camera was mounted on the driving simulator platform as shown in Fig.10. The images collected were used to detect the driver's behavioural based measures.

2) VISION MODULE BASED INDICATOR OF FATIGUE

There are a number of behavioural signs that a driver exhibits when he is drowsy under a fatigue condition. A driver who is drowsy displays characteristic facial movements like rapid blinking, swinging their head in random directions, yawning frequently and a combination of one or all of the above signs. The main aim of this vision module is to detect these behavioural signs with the help of vision based approach. For our study we concentrated on the behavioural signs of eyes closed, yawning and eyes closed while yawning.

3) IMAGE DATA COLLECTION, PREPROCESSING AND TAGGING

The images were collected at the rate of 5 frames per second for 60 seconds under each scenario and they were categorised according to the scenarios of normal, yawning, eyes closed and eyes closed while yawning. These scenarios were recorded from three test subjects and the scenarios were also recorded with the test subjects wearing glasses. In order to establish a base line, normal driving of the test subjects with and without glasses were also recorded. Hence a total of eight scenarios with approximately 300 images for each scenario and hence for three subjects a total of 6000 images were collected. The face of the test subjects are detected using the Dlib facial landmark detector. Dlib is a C++ open-source cross platform software library and also supports

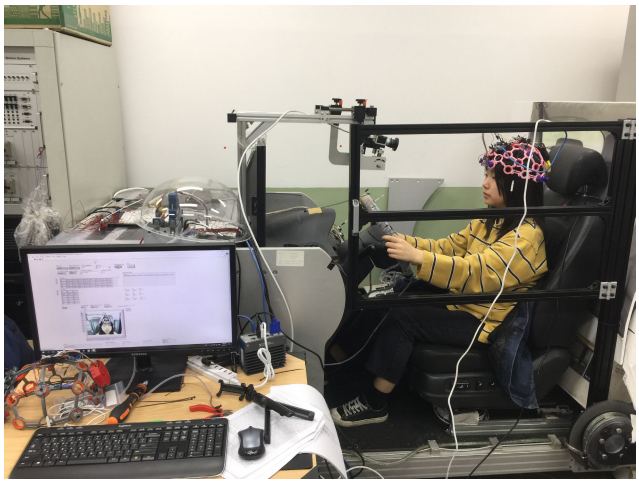


FIGURE 10. Driving simulator platform along with the volunteer subject.

python API. The face is detected with the help of Dlib's frontal face detector(DFFC). The DFFC was made using histogram of oriented gradients feature combined with the concept of linear classifier, and image pyramid and sliding window detection scheme. The DFFC gives an output of bounding box around the detected face and this output image was then used to detect the facial landmarks on the detected face using the Dlib's shape predictor(DSP) along with the ''shapepredictor-68-face-landmark.dat''. The DSP used the bounding box to align itself to predict the facial features in the image. The Dlibs shapepredictor-68-face-landmark.dat is the pre-trained Dlib model for face landmark detection which was trained on the i.bug 300-W face landmark dataset [47]. The output from the DSP gives the information of the facial landmark features. This information is then used to crop and resize the image evenly around the detected facial features.

The resized image grouped into two groups of 128 x 128 pixel size and 256 x 256 pixel size, which are categorised into four main categories of normal, yawning, eyes closed and yawning while eyes closed. Two groups of images were selected to compare and validate the results. These categorised images are again processed using DFFC and DSP using the shapepredictor-68-face-landmark.dat which detects the facial landmarks. These facial landmarks are then used to detect the distance between the nodes of left eye, right eye and mouth. The information from each image are then tagged into 15 columns where the first column detects the status of the driver as 0 for awake, 1 for yawning, 2 for eyes closed, 3 for eyes closed while yawning and 4 for absence of data as in Table.8. Columns two to five indicate the distance of left eye nodes, columns six to nine indicate the distance of right eye nodes and columns 10 to 15 indicate the distance of mouth nodes. The node information for facial feature landmark is shown in Fig.11. and the information on tagging distance between nodes is given in Table.9. The structure of the proposed vision module neural network system is shown in Fig.12.

TABLE 8. Vision module driver status tagging information.

Module Name	Description of Prediction	Value of Prediction
Vision module	Normal	0
	Yawning	1
	Eyes closed	2
	Eyes Closed while yawning	3
	Absence of data	4

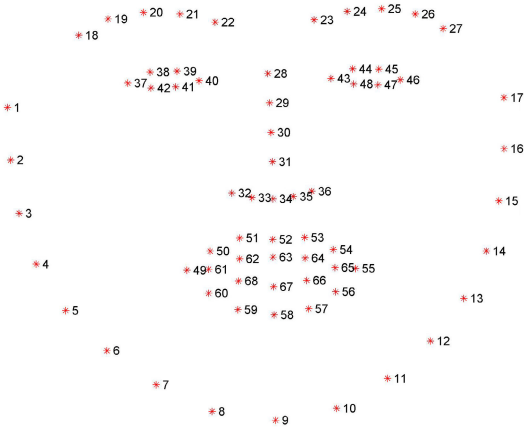


FIGURE 11. The 68 point markup used for tagging facial feature information.

TABLE 9. Tagged data information.

Column Number	Tagging information
1	Driver status
2	Distance between nodes 38 and 42 in x direction
3	Distance between nodes 38 and 42 in y direction
4	Distance between nodes 39 and 41 in x direction
5	Distance between nodes 39 and 41 in y direction
6	Distance between nodes 44 and 48 in x direction
7	Distance between nodes 44 and 48 in y direction
8	Distance between nodes 45 and 47 in x direction
9	Distance between nodes 45 and 47 in y direction
10	Distance between nodes 62 and 68 in x direction
11	Distance between nodes 62 and 68 in y direction
12	Distance between nodes 63 and 67 in x direction
13	Distance between nodes 63 and 67 in y direction
14	Distance between nodes 64 and 66 in x direction
15	Distance between nodes 64 and 66 in y direction

4) FATIGUE DETECTION BASED ON THE PROPOSED VISION MODULE NN

The tagged data is then given as input to the NN. The input layer of the NN consists of 14 input nodes for each of the tagged distance between the nodes information. The output layer of the NN consists of 4 output nodes each once representing the status of the driver. There are 100 hidden neurons between the input and output layers of the NN. The activation function used in the input layer of NN is the leaky ReLU. The reason for using leaky ReLU instead of ReLU function is that there is a possibility for the weights to be updated in such a way that the neuron never gets activated when a large gradient is passing through ReLU. Hence there is a problem of ReLU units dying during training process which also called as dying ReLU problem. Thus leaky ReLU overcomes such problem

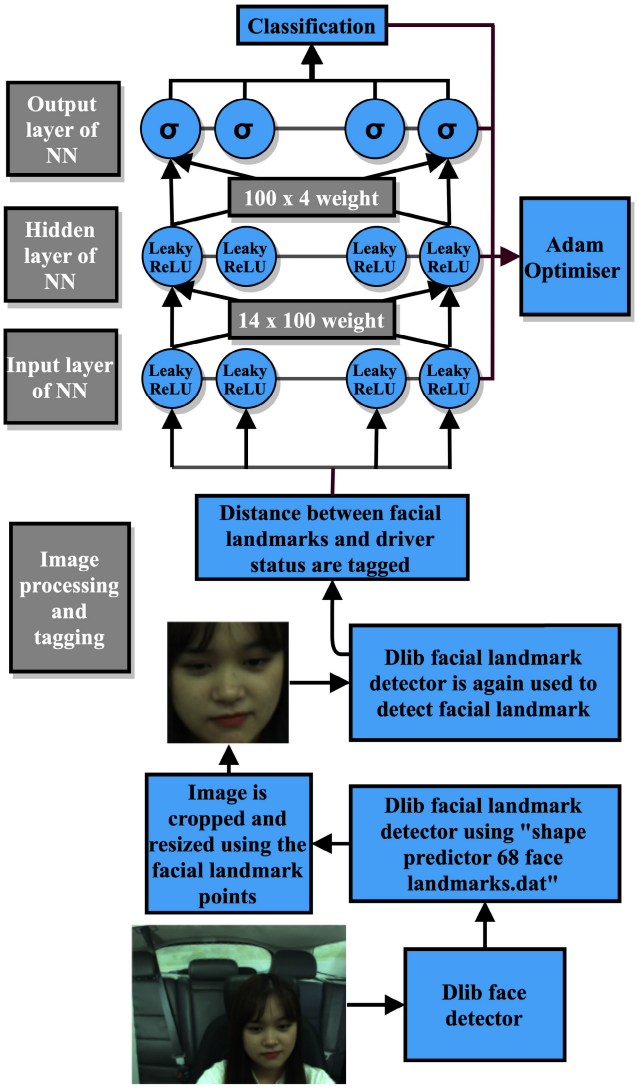


FIGURE 12. Structure of proposed Vision module neural network.

by having a small negative slope of 0.01 or so instead of the function being zero.

All layers of the NN use batch normalization(BN) which accelerates deep network training by reducing the internal covariant shift. In BN the normalization is done for each training mini-batch thus making the process of normalization a part of the model architecture [48]. It enables the use of higher learning rates and also acts as a regularizer. The output layer of the NN uses SF to classify the facial drowsiness state of the driver. AO was also used for fast optimization.

The cost and accuracy of prediction for the proposed NN was calculated. The cost of the prediction is calculated by finding the mean of sigmoid cross entropy with logits. The sigmoid cross entropy with logits measures the probability error in discrete classification tasks or multilabel classification in which each class is independent and not mutually exclusive. For clarity the function for calculation logistic loss is given as follows.

$$Logistic\ loss = x - x * z + log(1 + exp(-x)) \quad (4)$$

TABLE 10. 8 Channel EEG data represented in microvolts of x10^3 along with Tagging value.

Sl.No	Sample Index	Fp1	Fp2	C3	C4	P7	P8	O1	O2	Tagged Value
1	0	72.88	80.09	78.69	19.55	117.94	28.14	49.17	51.72	0
2	1	72.89	80.09	78.60	19.55	117.85	28.14	49.19	51.51	0
3	2	72.86	80.07	78.65	19.53	117.97	28.13	49.18	51.91	0
•	•	•	•	•	•	•	•	•	•	•
•	•	•	•	•	•	•	•	•	•	•
•	•	•	•	•	•	•	•	•	•	•
1535	254	72.95	79.99	78.83	19.50	117.63	28.13	49.17	67.16	1
1536	255	72.97	80.00	78.82	19.52	117.54	28.14	49.19	66.80	1
1537	0	72.96	79.99	78.74	19.52	117.49	28.14	49.22	66.65	1
1538	1	72.94	79.98	78.76	19.50	117.59	28.12	49.20	67.02	1
•	•	•	•	•	•	•	•	•	•	•
•	•	•	•	•	•	•	•	•	•	•

TABLE 11. EEG module driver status tagging information.

Test Subject Number	Number of 8 x 255 instances of raw data collected
1	187
2	199
3	153

where $x = logits$ and $z = labels$. Logits is basically the raw output of the final neuron layer of the NN which returns the raw values of prediction. This raw values of prediction is then used to calculate the mean value of cross entropy with sigmoid function which gives the cost of prediction. Accuracy of prediction is calculated by finding the mean between the predicted value of the proposed vision module NN and the actual value of prediction. Thus with the cost and accuracy functions the performance of the proposed vision module NN could be verified.

IV. RESULTS AND DISCUSSION

The results of each of the modules are first discussed individually and then the combined multimodal approach to detect driver fatigue with EEG, gyroscope and image processing data will be discussed.

A. EEG MODULE

An EEG signal of around 5 min was recorded for each test subject with a sample rate of 250 Hz. MATLAB was used for programming the CWT, FFT and tagging was performed on the raw dataset. The graph of the CWT for the EEG tagged data for one of the test subjects is as shown in Fig.13. The CWT graph shows the time frequency analysis of the EEG signals for the O1 and O2 electrode along with the magnitude information. Fig.13 shows that there exists a frequency fluctuation in the 10 to 13 Hz interval which can be seen in the graph. A sample of the 8 Channel EEG data along with tagging information is shown in Table.10. The same data was tagged using FFT for comparing the results with the CWT analysis. The FFT graph for the same test subject is shown in Fig.14(a) and the corresponding tagging information is shown in Fig.14(b).

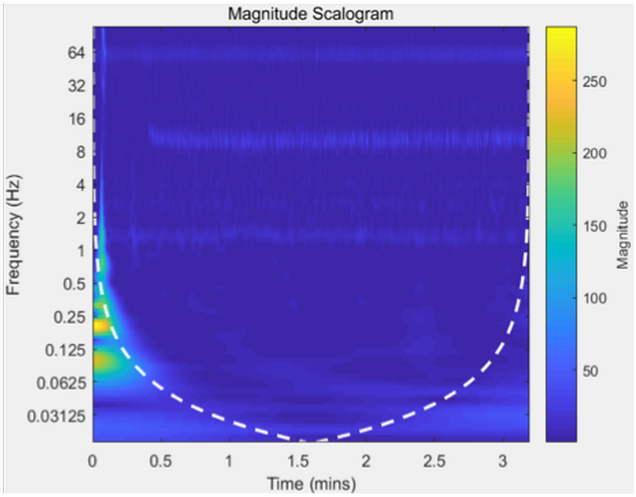
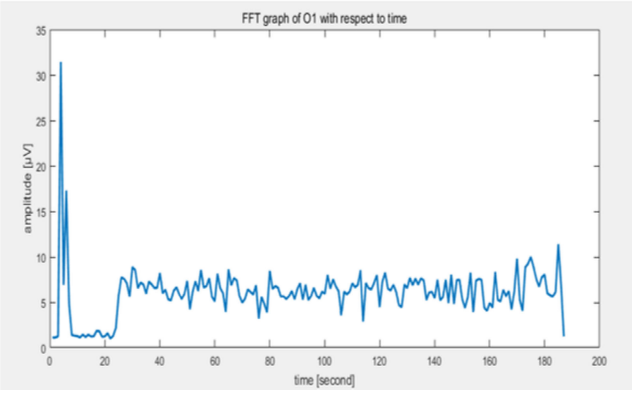
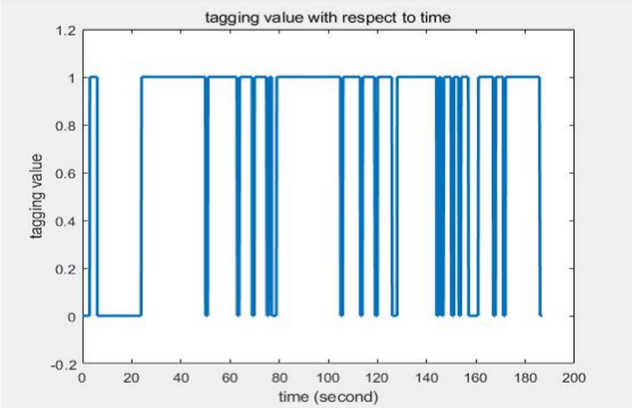


FIGURE 13. EEG signal CWT analysis with Morse wavelet.



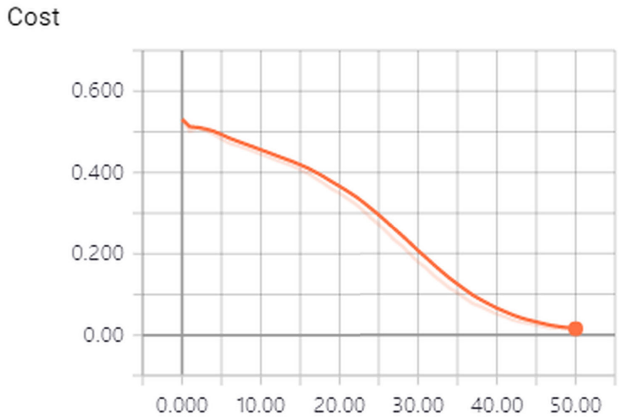
(a) FFT graph of O1 with respect to time



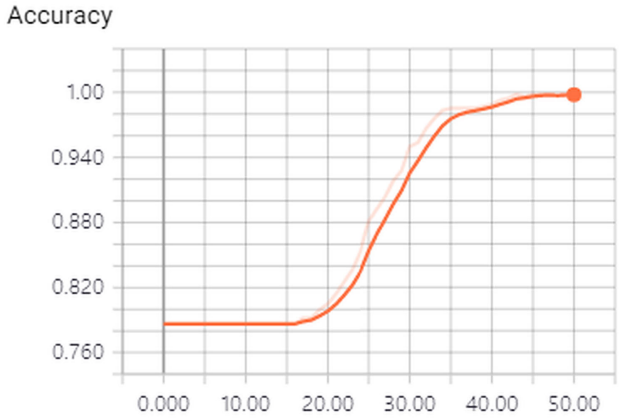
(b) Tagging value of FFT data with respect to time

FIGURE 14. Information from FFT data.

The 8 channel EEG data is then given as input the DNN. We used Window 10 OS, Python 3.6.4, and Tensorflow 1.8.0 in an Intel Core i5-6400 CPU processor with NVIDIA GeForce GT 730 graphical processing unit. The training data is the variation of EEG data with 539 datasets of 8 x 255 size along with information of tagging that was collected from the three test subjects as shown in Table.11.

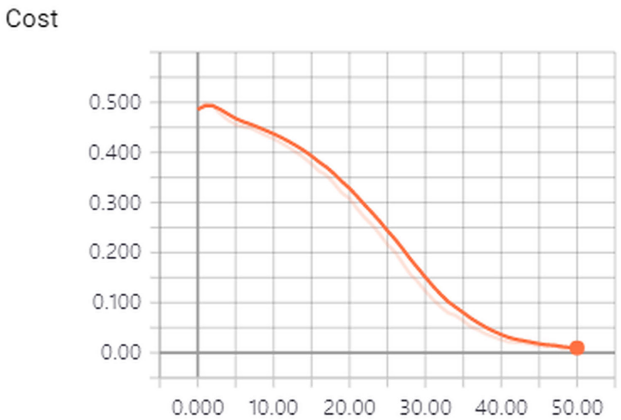


(a) Cost of prediction for every epoch (Cost of prediction along y-axis and epoch number along x-axis)

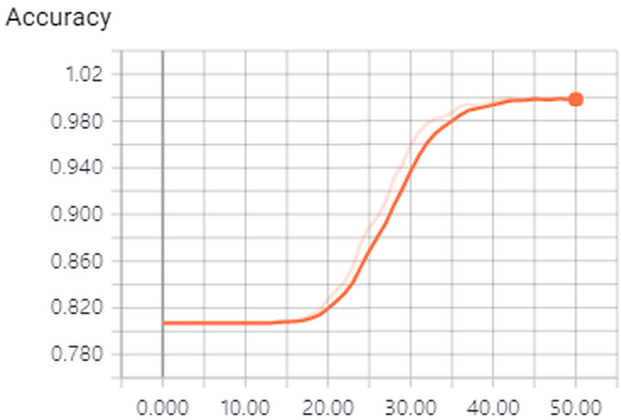


(b) Accuracy of prediction for every epoch (Accuracy of prediction along y-axis and epoch number along x-axis)

FIGURE 15. Results of the proposed DNN for FFT data.



(a) Cost of prediction for every epoch (Cost of prediction along y-axis and epoch number along x-axis)



(b) Accuracy of prediction for every epoch (Accuracy of prediction along y-axis and epoch number along x-axis)

FIGURE 16. Results of the proposed DNN for CWT data.

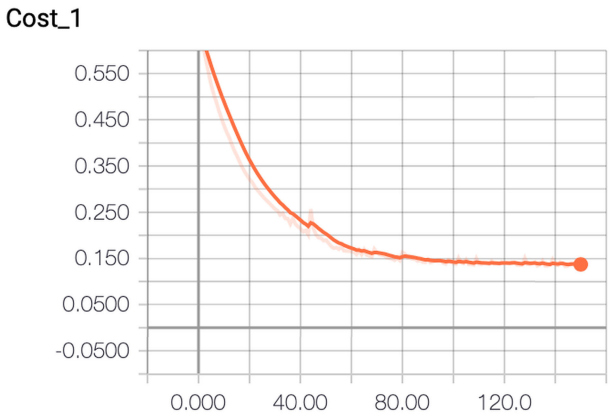
In the learning process 10-fold cross validation was used for each and every epoch and a total of 50 generations were studied to find out the cost and accuracy of prediction of the proposed DNN. A dropout value of 1 was used to train and test the dataset for every 10 generations. It was found that the proposed DNN's final learning cost function value was about 0.06 and the accuracy of detection for the whole tagged FFT data was about 98.11% as shown in Fig.15. A final learning cost function value of 0.01 and an accuracy of detection for the whole tagged CWT data was found to be 99.62% for the proposed DNN as shown in Fig.16.

There are two lines in the graph as shown in Fig.15 and Fig.16. The thicker orange line is the smoothed values, and the lighter orange line is the actual values which were logged. This smoothing can be useful for displaying the overall trend when the summary logging frequency is higher i.e. after every training step rather than after every epoch as in this example. The actual value being displayed in the graph do not change even if smoothing value is changed as the logit values from the last layer of the neural network are used to display the behaviour of the neural network. From Fig.15 and Fig.16 we can see that the slope of cost decreases and the slope of accuracy increases as the learning progresses. This shows that the proposed DNN performs well in predicting the drowsiness state of the driver. The results also showed that the proposed DNN with CWT tagged data performed better than the FFT tagged data. This shows that our new proposed method of detecting drowsiness using DNN and CWT tagged data can be used for detecting drowsiness from EEG data.

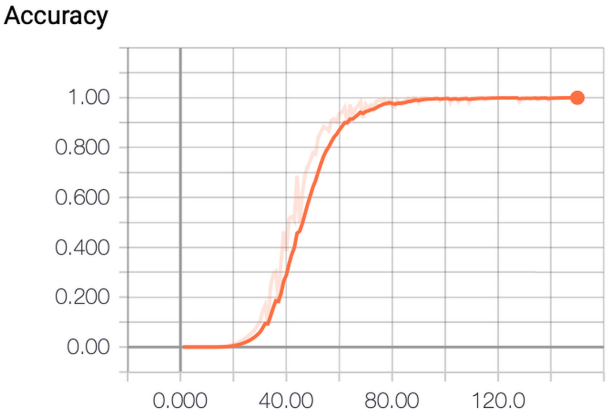
B. GYROSCOPE MODULE

The gyroscope data for each section is recorded using the test subjects with a baud rate of 115200. The rate at which the information is transferred in a communication channel is called the baud rate. The recorded raw gyroscope data consists of three axis accelerometer, gyroscope, roll, pitch and yaw data. Hence a total of 5000 instances, each containing 9 data types for each section was recorded using MATLAB.

The raw data consisting of 45000 x 9 instances is then converted into the head activity data by calculating the mean, SD and PCA for each of the instance along with section information therefore the head activity data thus



(a) Cost of gyroscope data prediction for every epoch (Cost of prediction along y-axis and epoch number along x-axis)



(b) Accuracy of gyroscope data prediction for every epoch (Accuracy of prediction along y-axis and epoch number along x-axis)

FIGURE 17. Results of the proposed gyroscope module DNN for tagged head activity dataset.

(a) Test subject driving in normal condition

(b) Test subject yawning while driving

(c) Test subject eyes closed while driving

(d) Test subject eyes closed and yawning while driving

FIGURE 18. Screen shot of the raw image dataset of test subject collected for different scenarios.

obtained consists of 45000 x 21 data types. In order to verify the head activity data classification learner toolbox from MATLAB was used and the fine Gaussian SVM was trained

TABLE 12. Vision module driver status tagging information.

Test Subject Number	Reading Glasses	Scenario Type	Number of Images collected
1	Yes	Normal	299
		Yawning	299
		Eyes closed	299
		Eyes closed while yawning	295
1	No	Normal	299
		Yawning	299
		Eyes closed	299
		Eyes closed while yawning	297
2	Yes	Normal	299
		Yawning	299
		Eyes closed	298
		Eyes closed while yawning	297
2	No	Normal	299
		Yawning	298
		Eyes closed	299
		Eyes closed while yawning	297
3	No	Normal	298
		Yawning	298
		Eyes closed	299
		Eyes closed while yawning	299

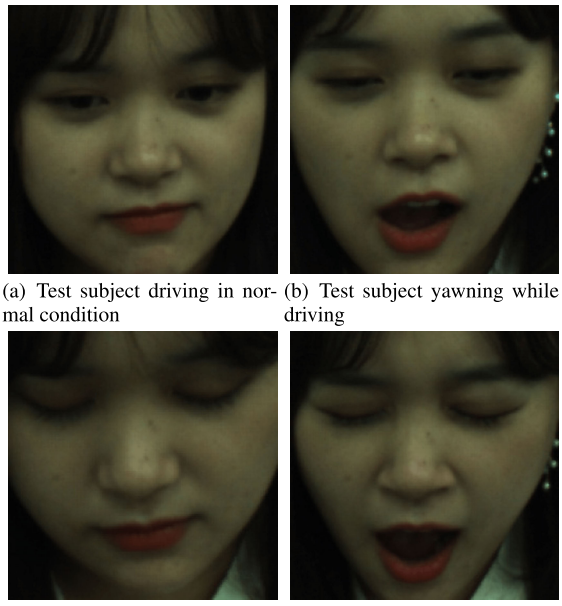
TABLE 13. Cropped image dataset information.

Scenario Type	Number of images cropped and resized categorically
Normal	1482
Yawning	1440
Eyes closed	922
Eyes closed while yawning	1353

with 10-fold cross validation. The results of the fine Gaussian SVM showed 98.9% accuracy in classifying the sections.

The head activity data is then given as input to the gyroscope module DNN. Windows 10 OS, Python 3.6.4, and Tensorflow 1.8.0 in an Intel Core i5-6400 CPU processor was used again with NVIDIA GeForce GT 730 graphical processing unit. The training dataset is a variation of gyroscope head activity dataset of 45000 x 21 size along with tagging information of nine sections. In the learning process 10-fold cross validation was used for every epoch and a total of 150 generations were studied to find out the cost and accuracy of prediction of the proposed gyroscope module DNN. A dropout value of 0.9 was used to train and test the dataset for every 10 generations. It was found that the proposed gyroscope module DNN's final learning cost function value was 0.1347 and the accuracy of detection for the whole tagged human activity dataset was found to be 99.97% as shown in Fig.17.

From Fig.17(a) one can see that the slope of cost decreases and from Fig.17(b) it can be observed that the slope of accuracy increases as the learning progresses. This shows that the proposed gyroscope module DNN performs well in predicting the head motion of the driver.



(a) Test subject driving in normal condition

(b) Test subject yawning while driving

(c) Test subject eyes closed while driving

(d) Test subject eyes closed and yawning while driving

FIGURE 19. Screen shot of the cropped image dataset of test subject collected for different scenarios.

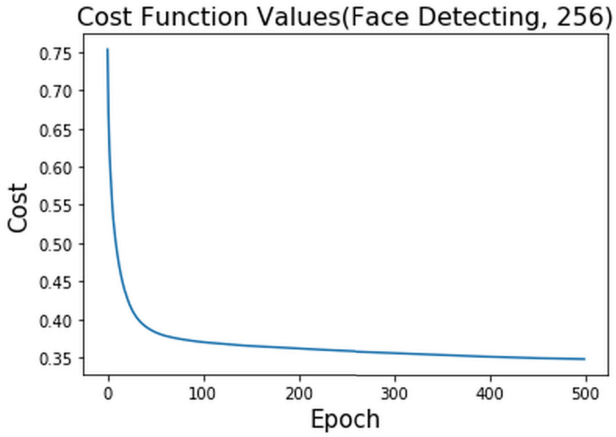
TABLE 14. Sample of tagged value for image database.

Sl.No	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•
•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•
3058	2	0	-5	1	-4	-1	-5	-1	-4	0	-2	0	-2	0	-2
3059	2	1	-5	1	-6	0	-5	0	-5	-1	-1	0	-2	0	-2
3060	2	1	-5	1	-5	-1	-5	-1	-5	0	-2	1	-2	0	-2
3061	3	-1	-4	-2	-4	-1	-4	-1	-4	0	-6	0	-7	-1	-7
3062	3	-1	-5	-1	-6	0	-5	-1	-5	0	-10	-1	-10	-1	-10
3063	3	1	-5	0	-6	0	-6	-1	-6	-1	-5	-1	-6	-1	-6
3064	3	2	-5	2	-6	1	-6	1	-6	2	-1	1	0	2	0
3065	3	2	-7	2	-7	0	-6	1	-6	2	-3	2	-3	1	-3
3066	3	2	-7	1	-7	0	-7	0	-7	2	-4	2	-3	1	-3
3067	3	1	-7	2	-7	1	-8	1	-7	2	-3	2	-3	2	-3
3068	3	2	-7	1	-7	0	-8	1	-6	2	-3	2	-3	2	-2
3069	3	2	-6	1	-6	1	-6	1	-7	2	-3	1	-2	1	-2
3070	3	0	-5	0	-5	0	-6	1	-5	0	-3	0	-4	0	-4
•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•
•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•
•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•

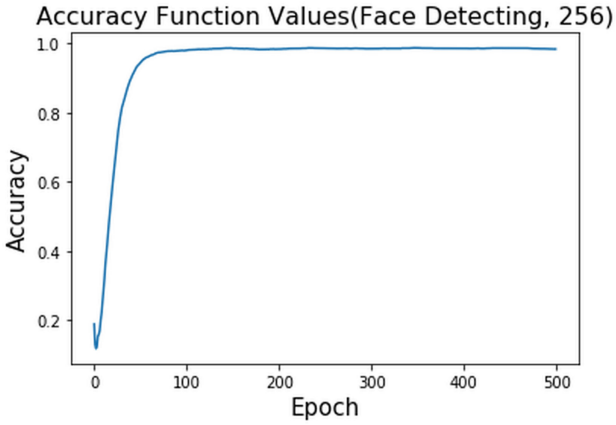
C. VISION MODULE

The image datasets consisting of four scenario of normal, yawning, eyes closed and eyes closed while yawning were recorded from three test subjects who were with and without reading glasses for recoding the dataset as shown in Table.12. Hence a total of 5696 images were collected for the 4 different datasets. A screen shot of a test subject with raw image dataset of 658 x 492 pixel size collected for normal, yawning, eyes closed and eyes closed while yawning is shown in Fig.18.

The collected dataset are then processed and categorised according to the four scenarios using the DFFC and DSP along with ''shapepredictor-68-face-landmark.dat''. The processed images consists of face of the test subject cropped and



(a) Cost of image data prediction for every epoch (Cost of prediction along y-axis and epoch number along x-axis)



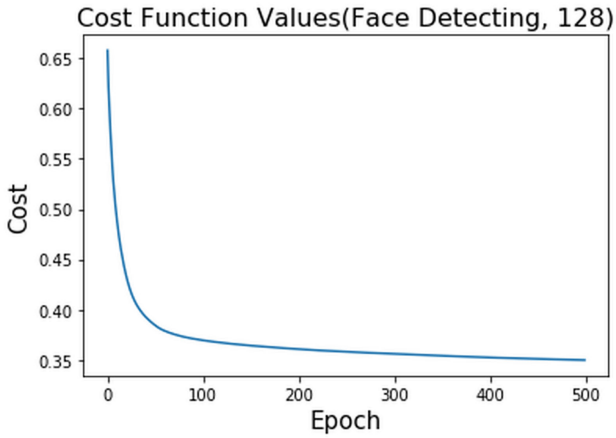
(b) Accuracy of image data prediction for every epoch (Accuracy of prediction along y-axis and epoch number along x-axis)

FIGURE 20. Results of the proposed vision module NN for tagged datase of 256 x 256 pixel image's.

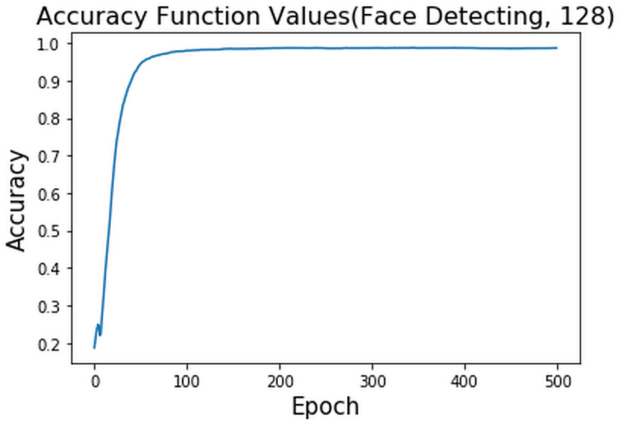
categorised into two datasets consisting of 4000 images of different image sizes which are used for comparing and verifying the results. The cropped images are of 128 x 128 pixel and 256 x 256 pixel sizes and each category of the dataset consists of four scenarios as given in Table.8. A screen shot of the same test subject as shown in Fig.18 with cropped image of the detected face with image dataset of 256 x 256 pixel size collected for eyes closed while yawning is shown in Fig.19.

The collected image dataset is again processed using the DFFC and DSP using the shapepredictor-68-face-landmark.dat which detects the facial landmarks. The distance between the facial landmarks is calculated and the information is tagged for each image along with the driver status information. A total of 4000 images from each image size dataset were tagged and the details of driver status tagging is given in Table.12, while a sample of the tagged data is shown in Table.14. The number of tagged data for each scenario is as shown in Table.13.

The 15 column image tagged data is then given as input to the vision module NN, where the first column



(a) Cost of image data prediction for every epoch (Cost of prediction along y-axis and epoch number along x-axis)



(b) Accuracy of image data prediction for every epoch (Accuracy of prediction along y-axis and epoch number along x-axis)

FIGURE 21. Results of the proposed vision module NN for tagged datase of 128 x 128 pixel image's.

represents the status of the driver. We again used the same Window 10 OS, Python 3.6.4, and Tensorflow 1.8.0 in an Intel Core i5-6400 CPU processor with NVIDIA GeForce GT 730 graphical processing unit. The training data is the variation of 14 columns of corresponding node distance tagged information of around 8000 images from the two different cropped image size datasets.

In the learning process 10-fold cross validation was again used for each and every epoch and a total of 500 generations were studied to find out the cost and accuracy of prediction of the proposed vision module NN. It was found that the proposed vision module NN with with tagged data from 256 x 256 pixel image dataset showed an average learning cost of 0.347 and the accuracy of prediction was found to be 98.34% as shown in Fig.20. While the tagged data from 128 x 128 pixel image dataset showed an average learning cost of 0.3497 and an accuracy of prediction was found to be 98.41% as shown in Fig.21. From Fig.20 and Fig.21 we can see that the slope of cost of prediction decreases as the number of epochs increases and the accuracy of prediction increases as the learning progresses. This shows that the proposed vision based NN performs well in predicting the drowsiness state of the driver. The results also showed that the tagging data from 128 x 128 pixel image dataset showed better performance than the tagging data from 256 x 256 pixel image dataset.

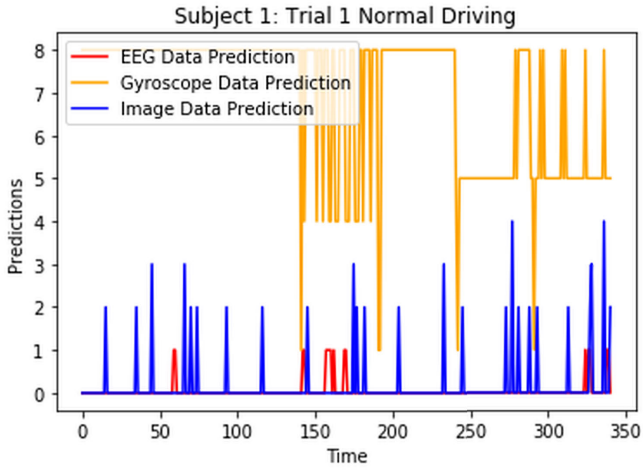
TABLE 15. Size of of the dataset for each test subject after prediction by the individual neural network modules according to each scenario.

Test Subject Number	Scenario Number	Size of the predicted dataset
1	1	341
	2	196
	3	196
	4	252
	5	196
2	1	184
	2	137
	3	192
	4	142
	5	132
3	1	147
	2	147
	3	144
	4	183
	5	142
4	1	243
	2	146
	3	142
	4	189
	5	145
5	1	208
	2	144
	3	150
	4	142
	5	123

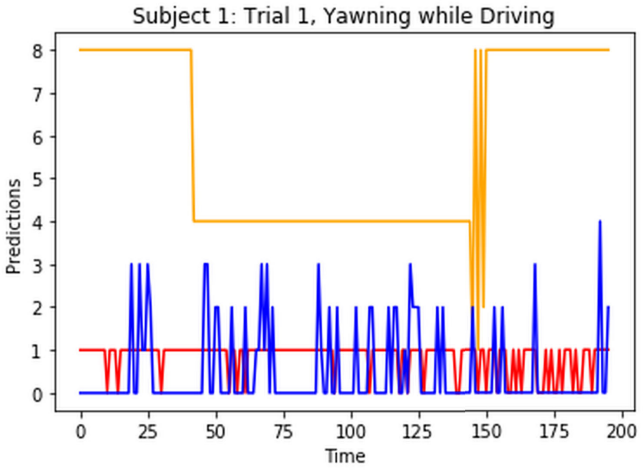
D. MULTIMODAL SYSTEM

The multimodal dataset consists of datas recorded from five test subjects under five scenarios as described in Table.4. The raw datas of EEG, gyroscope and image from the five scenarios are processed individually in their respective EEG module(EM), gyroscope module(GM) and vision module(VM) neural networks. The predictions from these individual modules contain the values as described in Table.5. These predictions are also time series based and contain the status and behavioural information of the driver. The size of the dataset for each test subject after prediction by the individual neural network modules according to the scenario are given in Table.15. The predictions from EEG module, gyroscope module and vision module of Subject 1 for different scenarios is plotted as graph and the relationship between each of them can be seen in Fig.22. The characteristics and relationship between the predictions can also be observed in Fig.22.

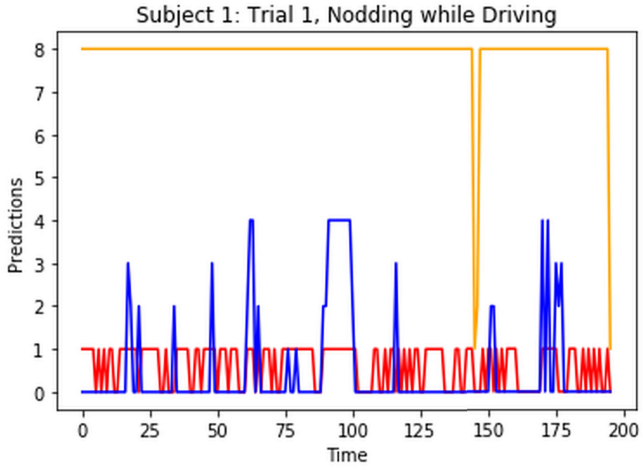
The data is tagged for predicting six classes as given in Table.5 using one hot method of tagging as it avoids misinterpreting the information to the DNN that if greater the tagging number then greater is its importance. Windows 10 OS, Python 3.6.4 and Tensorflow 1.8.0 in an Intel Core i5-6400 CPU processor with NVIDIA GeForce GT 730 graphical processing unit is used for training and testing



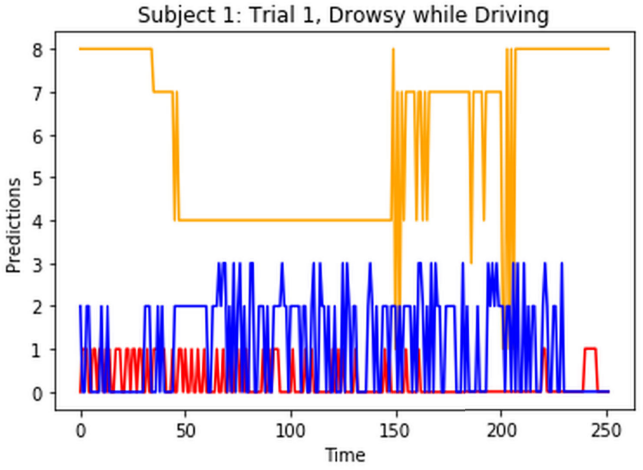
(a) Scenario 1(Normal state and Normal driving)



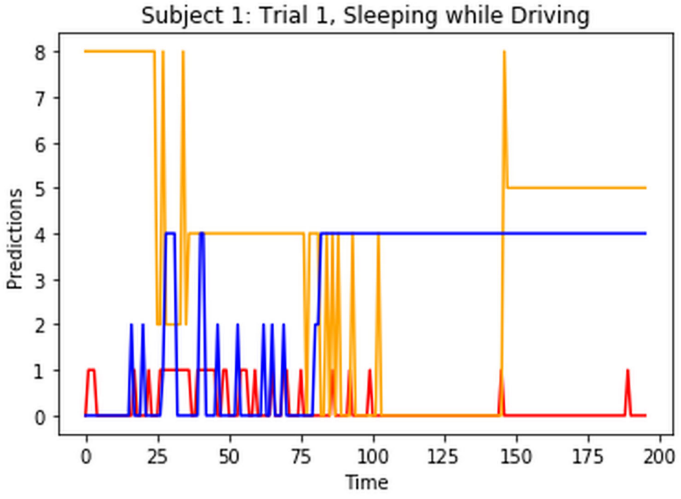
(b) Scenario 2(Fatigue state and yawning while driving)



(c) Scenario 3(Fatigue state and nodding while driving)



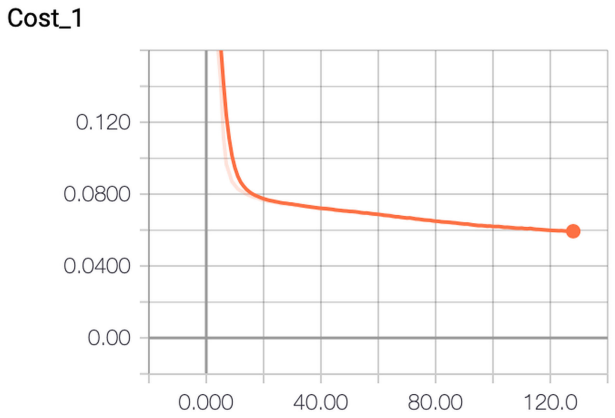
(d) Scenario 4(Fatigue state and extremely drowsy while driving)



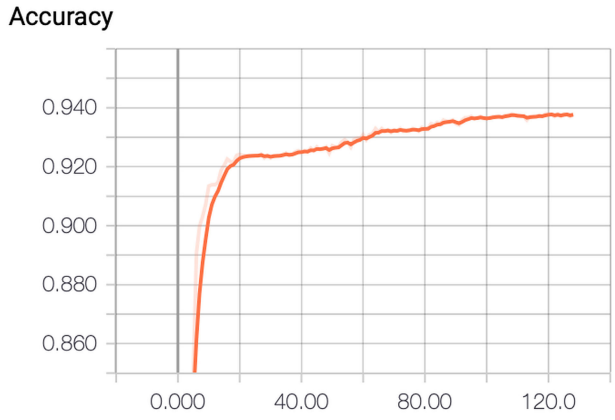
(e) Scenario 5(Fatigue state and drowsy to sleeping while driving)

FIGURE 22. **Predictions from EEG module, gyroscope module and vision modue of subject 1 for different scenarios with red representing EEG data prediction, yellow representing gyroscope data prediction and blue representing image data prediction.**

the proposed multimodal DNN. From Table.15 one can observe that a total of 4363 instances that contain prediction from EM, GM and VM are given as input to the proposed multimodal DNN along with tagging information. In the learning process 10-fold cross validation was used for each and every epoch and a total of 150 generations were studied

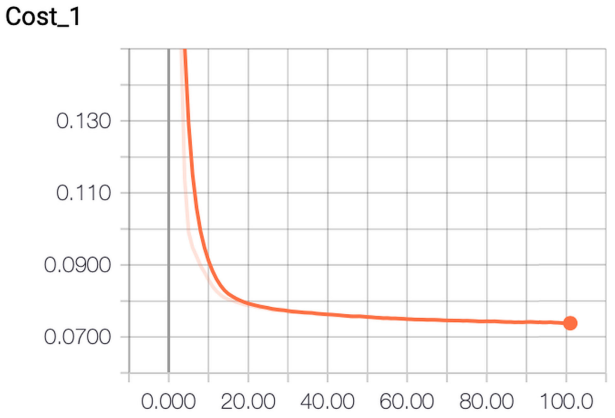


(a) Cost of multimodal data prediction for every epoch consisting of for EM, GM and VM data (Cost of prediction along y-axis and epoch number along x-axis)

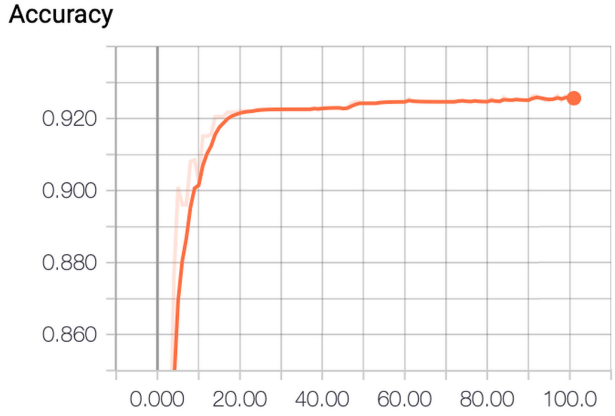


(b) Accuracy of image data prediction for every epoch (Accuracy of prediction along y-axis and epoch number along x-axis)

FIGURE 23. Results of the proposed multimodal DNN for EM, GM and VM data for 3 seconds history of time instances.



(a) Cost of multimodal data prediction for every epoch considering only EM and VM data (Cost of prediction along y-axis and epoch number along x-axis)



(b) Accuracy of multimodal data prediction for every epoch considering only EM and VM data (Accuracy of prediction along y-axis and epoch number along x-axis)

FIGURE 24. Results of the proposed multimodal DNN considering only EM and VM data for 3 seconds history of time instances.

TABLE 16. Performance of the proposed multi-modal DNN for combination of different datasets.

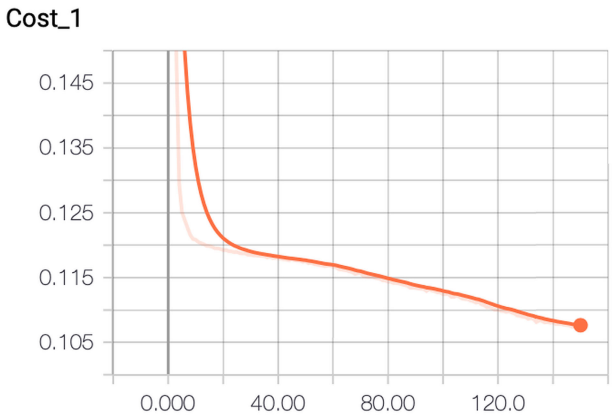
Module prediction datasets considered	Average Learning Cost	Average Learning Accuracy
EM and VM	0.07260	92.58%
EM and GM	0.10740	86.74%
EM, GM and VM	0.05781	93.91%

to find out the cost and accuracy of prediction of the proposed multimodal DNN. It was found that the proposed multimodal DNN, with tagged data showed an average learning cost of 0.05781 and the accuracy of prediction was found to be 93.91% [49] as shown in Fig.23.

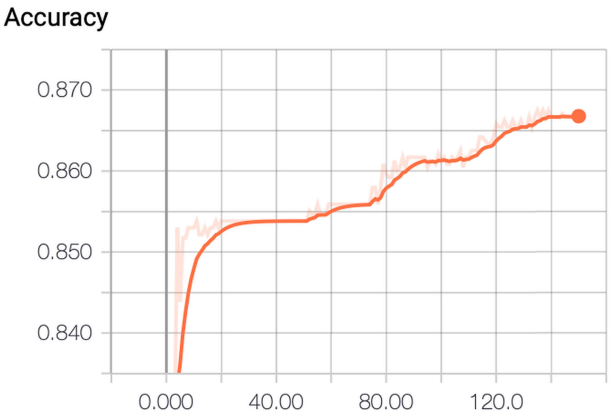
From Fig.23 one can see that the slope of cost of prediction decreases as the number of epochs increases and the accuracy of prediction increases as the learning progresses. This shows that the proposed multimodal DNN performs well in predicting the fatigue state of the driver.

In order to compare the performance and the behaviour of the proposed multimodal DNN the same dataset containing only the EM and VM, along with EM and GM were given as input along with the same tagging information. The below Fig.24 shows the performance of the proposed multimodal DNN with the EM and VM. The performance of the proposed multimodal DNN with EM and GM is shown in Fig.25

From Fig.24, the average learning cost and accuracy of prediction for the proposed multimodal DNN with EM and VM tagged data was found to be 0.07260 and 92.58%. While the average learning cost and accuracy of prediction for the proposed multimodal DNN with EM and GM tagged data was found to be 0.10740 and 86.74% as shown in Fig.25. Fig.24 and Fig.25 also show that slope of cost of prediction decreases as the number of epochs increases and the accuracy of prediction increases as the learning progresses. This shows the proposed multimodal DNN with reduced sensor information can also be used for predicting the fatigue status of the driver. Thus from Table.16 the proposed multimodal DNN with EEG module (EM), gyroscope module (GM) and vision module (VM) showed a high accuracy of 93.91% and is considered for the detection of driver fatigue.



(a) Cost of multimodal data prediction for every epoch considering only EM and GM data (Cost of prediction along y-axis and epoch number along x-axis)



(b) Accuracy of multimodal data prediction for every epoch considering only EM and GM data (Accuracy of prediction along y-axis and epoch number along x-axis)

FIGURE 25. Results of the proposed multimodal DNN considering only EM and GM data for 3 seconds history of time instances.

V. CONCLUSION

In this paper we have presented a framework where the driver fatigue is predicted based on the normal and fatigue scenarios while driving.

The EEG data collected is influenced by the degree of stress felt and the influence of surrounding environments like noise, temperature and other distractions. Even if the experiment is repeated with the same test subject, the value of EEG data will be different. Therefore it is necessary to collect more real world EEG data affected by real world environments. In this paper a new DNN model is proposed and used which can predict the drowsiness state of the driver using CWT tagged data. The results of the CWT tagged data showed better performance than the conventional FFT tagged data set. From the accuracy of prediction, it is observed that the proposed model can be utilised for detecting the N1 stage where the transition happens from awake to drowsy.

The image data collected is largely influenced by lighting and the presence of facial features in the frame. Allthough the system is robust in detecting the status of the driver when the facial features are present within the frame of the system. If the facial features are not present then the system is unable to detect the status of the driver. Hence this disadvantage is overcome with the use of other modules in detecting drowsiness. The proposed vision module NN performed well in detecting the status of the driver and the results from the 128 x 128 pixel tagged data exhibits better performance than the 256 x 256 pixel tagged data. Therefore from the accuracy of prediction it is confirmed that the trained vision module NN can be utilised for detecting drowsiness based on driver behavioural measures.

The gyroscope data collected helps in dividing the field of vision of the driver into various sections. This method helps in determining the head motion of the driver as the reason for head movement involves lot of real world factors. The drivers head pose helps in determining the drowsiness based on driver behavioural measures and the data streamed from the gyroscope is more robust compared to the image data. The decision from the gyroscope module NN complements the vision module NN. The proposed gyroscope module NN performed well in detecting the head pose or the head activity of the driver. Therefore from the accuracy of prediction one can see that the trained gyroscope module NN can be utilised for detecting drowsiness based on driver behavioural measures.

The trained neural networks from individual modules show high accuracy and low cost of prediction, their utilization for the multimodal fatigue detection can be realised. The multimodal prediction dataset consists of information of drowsiness in normal and fatigue state based on physiological and behavioural measures from the three different sensors. Table. 16 shows that as the number of sensors considered for prediction increases the accuracy of prediction increases. Therefore from the accuracy of predictions it is observed that the trained multimodal DNN with EEG module, gyroscope module and vision module is used for detection of driver fatigue.

New modules can be added to the system as the number of sensors are increased and the use of custom built NN for each of the modules ensures high accuracy for each type of dataset. The same system can be applied to other environments like manufacturing industry where the safety of the machine tool operator can be improved by changing only the learning parameters of vision and defining the safe sections from a gyroscope parameter while operating machine tool. Thus the proposed multimodal system detects driver fatigue using EEG, gyroscope and image processing using TensorFlow.

REFERENCES

[1] B. C. Tefft, "Asleep at the wheel: The prevalence and impact of drowsy driving," in *Proc. AAA Found. Traffic Saf.*, 2010.

[2] B. C. Tefft, "Prevalence motor vehicle crashes involving drowsy drivers, United States 2009–2013," in *Report Prepared for the American Automobile Association (AAA) Foundation for Traffic Safety*, 2014.

[3] B.-G. Lee, B.-L. Lee, and W.-Y. Chung, "Mobile healthcare for automatic driving sleep-onset detection using wavelet-based EEG and respiration signals," *Sensors*, vol. 14, no. 10, pp. 17915–17936, Sep. 2014. [Online]. Available: http://www.mdpi.com/1424-8220/14/10/17915

[4] C. Zhao, C. Zheng, M. Zhao, J. Liu, and Y. Tu, "Automatic classification of driving mental fatigue with EEG by wavelet packet energy and KPCA-SVM," *Int. J. Innov. Comput.*, vol. 7, no. 3, pp. 1157–1168, 2011.

[5] P. S. Rau, *Drowsy Driver Detection and Warning System for Commercial Vehicle Drivers: Field Operational Test Design, Data Analyses and Progress*. Washington, DC, USA: National Highway Safety Administration, Apr. 2005, pp. 1–7.

[6] R. Leproult, E. F. Colecchia, A. M. Berardi, R. Stickgold, S. M. Kosslyn, and E. Van Cauter, "Individual differences in subjective and objective alertness during sleep deprivation are stable and unrelated," *Amer. J. Physiol.-Regulatory, Integrative Comparative Physiol.*, vol. 284, no. 2, pp. R280–R290, Feb. 2003, doi: 10.1152/ajpregu.00197.2002.

[7] I. Lorenzo, J. Ramos, C. Arce, M. A. Guevara, and M. Corsi-Cabrera, "Effect of total sleep deprivation on reaction time and waking eeg activity in man," *Sleep*, vol. 18, no. 5, pp. 346–354, 1995, doi: 10.1093/sleep/18.5.346.

[8] R. B. Berry, R. Brooks, C. E. Gamaldo, S. M. Harding, C. Marcus, and B. Vaughn, "The AASM manual for the scoring of sleep and associated events," in *Rules, Terminology and Technical Specifications*. Darien, IL, USA: American Academy of Sleep Medicine, 2012.

[9] A. Sahayadhas, K. Sundaraj, and M. Murugappan, "Detecting driver drowsiness based on sensors: A review," *Sensors*, vol. 12, no. 12, pp. 16937–16953, Dec. 2012. [Online]. Available: http://www.mdpi.com/1424-8220/12/12/16937

[10] A. M. Williamson, A.-M. Feyer, and R. Friswell, "The impact of work practices on fatigue in long distance truck drivers," *Accident Anal. Prevention*, vol. 28, no. 6, pp. 709–719, Nov. 1996. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0001457596000449

[11] S. K. Lal and A. Craig, "A critical review of the psychophysiology of driver fatigue," *Biol. Psychol.*, vol. 55, no. 3, pp. 173–194, 2001. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0301051100000855

[12] R. P. Balandong, R. F. Ahmad, M. N. Mohamad Saad, and A. S. Malik, "A review on EEG-based automatic sleepiness detection systems for driver," *IEEE Access*, vol. 6, pp. 22908–22919, 2018.

[13] T. Åkerstedt and M. Gillberg, "Subjective and objective sleepiness in the active individual," *Int. J. Neurosci.*, vol. 52, nos. 1–2, pp. 29–37, Jan. 1990, doi: 10.3109/00207459008994241.

[14] M. Gillberg, G. Kecklund, and T. Åkerstedt, "Relations between performance and subjective ratings of sleepiness during a night awake," *Sleep*, vol. 17, no. 3, pp. 236–241, May 1994, doi: 10.1093/sleep/17.3.236.

[15] K. Kaida, M. Takahashi, T. Åkerstedt, A. Nakata, Y. Otsuka, T. Haratani, and K. Fukasawa, "Validation of the karolinska sleepiness scale against performance and EEG variables," *Clin. Neurophysiol.*, vol. 117, no. 7, pp. 1574–1581, Jul. 2006. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1388245706001428

[16] C. C. Liu, S. G. Hosking, and M. G. Lenné, "Predicting driver drowsiness using vehicle measures: Recent insights and future challenges," *J. Saf. Res.*, vol. 40, pp. 239–245, Aug. 2009. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0022437509000668

[17] M. Chai, S.-W. Li, W.-C. Sun, M.-Z. Guo, and M.-Y. Huang, "Drowsiness monitoring based on steering wheel status," *Transp. Res. D, Transp. Environ.*, vol. 66, pp. 95–103, Jan. 2019. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1361920917306582

[18] I.-H. Choi and Y.-G. Kim, "Head pose and gaze direction tracking for detecting a drowsy driver," in *Proc. Int. Conf. Big Data Smart Comput. (BIGCOMP)*, Jan. 2014, pp. 241–244.

[19] P. Smith, M. Shah, and N. da Vitoria Lobo, "Determining driver visual attention with one camera," *IEEE Trans. Intell. Transp. Syst.*, vol. 4, no. 4, pp. 205–218, Dec. 2003.

[20] T. Wang and P. Shi, "Yawning detection for determining driver drowsiness," in *Proc. IEEE Int. Workshop VLSI Design Video Technol.*, May 2005, pp. 373–376.

[21] Z. Jie, M. Mahmoud, Q. Stafford-Fraser, P. Robinson, E. Dias, and L. Skrypchuk, "Analysis of yawning behaviour in spontaneous expressions of drowsy drivers," in *Proc. 13th IEEE Int. Conf. Autom. Face Gesture Recognit. (FG)*, May 2018, pp. 571–576.

[22] M. Patel, S. K. L. Lal, D. Kavanagh, and P. Rossiter, "Applying neural network analysis on heart rate variability data to assess driver fatigue," *Expert Syst. Appl.*, vol. 38, no. 6, pp. 7235–7242, Jun. 2011. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0957417410013916

[23] M. B. Kurt, N. Sezgin, M. Akin, G. Kirbas, and M. Bayram, "The ANN-based computing of drowsy level," *Expert Syst. Appl.*, vol. 36, no. 2, pp. 2534–2542, Mar. 2009. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0957417408000675

[24] S. Hu and G. Zheng, "Driver drowsiness detection with eyelid related parameters by support vector machine," *Expert Syst. Appl.*, vol. 36, no. 4, pp. 7651–7658, May 2009. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0957417408006714

[25] A. Koesdwiady, R. Soua, F. Karray, and M. S. Kamel, "Recent trends in driver safety monitoring systems: State of the art and challenges," *IEEE Trans. Veh. Technol.*, vol. 66, no. 6, pp. 4550–4563, Jun. 2017.

[26] N. Dong, Y. Li, Z. Gao, W. H. Ip, and K. L. Yung, "A WPCA-based method for detecting fatigue driving from EEG-based Internet of vehicles system," *IEEE Access*, vol. 7, pp. 124702–124711, 2019.

[27] Y. Dong, Z. Hu, K. Uchimura, and N. Murayama, "Driver inattention monitoring system for intelligent vehicles: A review," *IEEE Trans. Intell. Transp. Syst.*, vol. 12, no. 2, pp. 596–614, Jun. 2011.

[28] P. Kumari, L. Mathew, and P. Syal, "Increasing trend of wearables and multimodal interface for human activity monitoring: A review," *Biosens. Bioelectron.*, vol. 90, pp. 298–307, Apr. 2017. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0956566316312295

[29] M. I. Chacon-Murguia and C. Prieto-Resendiz, "Detecting driver drowsiness: A survey of system designs and technology.," *IEEE Consum. Electron. Mag.*, vol. 4, no. 4, pp. 107–119, Oct. 2015.

[30] M. Jagannath and V. Balasubramanian, "Assessment of early onset of driver fatigue using multimodal fatigue measures in a static simulator," *Appl. Ergonom.*, vol. 45, no. 4, pp. 1140–1147, Jul. 2014. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0003687014000222

[31] N. Li and C. Busso, "Predicting perceived visual and cognitive distractions of drivers with multimodal features," *IEEE Trans. Intell. Transp. Syst.*, vol. 16, no. 1, pp. 51–65, Feb. 2015.

[32] J.-L. Yin, B.-H. Chen, K.-H.-R. Lai, and Y. Li, "Automatic dangerous driving intensity analysis for advanced driver assistance systems from multimodal driving signals," *IEEE Sensors J.*, vol. 18, no. 12, pp. 4785–4794, Jun. 2018.

[33] T. Nguyen, S. Ahn, H. Jang, S. C. Jun, and J. G. Kim, "Utilization of a combined eeg/nirs system to predict driver drowsiness," *Sci. Rep.*, vol. 7, p. 43933, Mar. 2017. [Online]. Available: https://search.proquest.com/docview/1903364010?accountid=11933

[34] W.-L. Zheng and B.-L. Lu, "A multimodal approach to estimating vigilance using EEG and forehead EOG," *J. Neural Eng.*, vol. 14, no. 2, Apr. 2017, Art. no. 026017.

[35] I. Daza, L. Bergasa, S. Bronte, J. Yebes, J. Almazán, and R. Arroyo, "Fusion of optimized indicators from advanced driver assistance systems (ADAS) for driver drowsiness detection," *Sensors*, vol. 14, no. 1, pp. 1106–1131, Jan. 2014. [Online]. Available: http://www.mdpi.com/1424-8220/14/1/1106

[36] C. Craye, A. Rashwan, M. S. Kamel, and F. Karray, "A multi-modal driver fatigue and distraction assessment system," *Int. J. Intell. Transp. Syst. Res.*, vol. 14, no. 3, pp. 173–194, Sep. 2016, doi: 10.1007/s13177-015-0112-9.

[37] M. Awais, N. Badruddin, and M. Drieberg, "A hybrid approach to detect driver drowsiness utilizing physiological signals to improve system performance and wearability," *Sensors*, vol. 17, no. 9, p. 1991, Aug. 2017, doi: 10.3390/s17091991.

[38] M. Miyaji, H. Kawanaka, and K. Oguri, "Driver's cognitive distraction detection using physiological features by the AdaBoost," in *Proc. 12th Int. IEEE Conf. Intell. Transp. Syst.*, Oct. 2009, pp. 1–6.

[39] M. Rimini-Doering, D. Manstetten, T. Altmueller, U. Ladstaetter, and M. Mahler, "Monitoring driver drowsiness and stress in a driving simulator," in *Proc. 1st Int. Driving Symp. Hum. Factors Driver Assessment, Training Vehicle Design, driving Assessment*, 2001, pp. 58–63.

[40] C. Busso and J. Jain, "Advances in multimodal tracking of driver distraction," in *Digital Signal Processing for In-Vehicle Systems and Safety*. New York, NY, USA: Springer, 2012, pp. 253–270.

[41] G. Yang, Y. Lin, and P. Bhattacharya, "A driver fatigue recognition model based on information fusion and dynamic Bayesian network," *Inf. Sci.*, vol. 180, no. 10, pp. 1942–1954, May 2010. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0020025510000253
[42] A. Eskandarian, R. Sayed, P. Delaigue, A. Mortazavi, and J. Blum, "Advanced driver fatigue research," U.S. Dept. Transp., Federal Motor Carrier Saf. Admin., Office Res. Anal., Washington, DC, USA, Tech.Rep. FMCSA-RRR-07-001, 2007.
[43] S. Kaplan, M. A. Guvensan, A. G. Yavuz, and Y. Karalurt, "Driver behavior analysis for safe driving: A survey," *IEEE Trans. Intell. Transp. Syst.*, vol. 16, no. 6, pp. 3017–3032, Dec. 2015.
[44] K. G. Hooper and H. W. McGee, "Driver perception-reaction time: Are revisions to current specification values in order," Transp. Res. Board, Washington,ăDC, USA, Tech. Rep. HS-036 165, 1983.
[45] D. P. Subha, P. K. Joseph, R. Acharya U, and C. M. Lim, "EEG signal analysis: A survey," *J. Med. Syst.*, vol. 34, no. 2, pp. 195–212, Apr. 2010, doi: 10.1007/s10916-008-9231-z.
[46] Y. Zhao, L. Görne, I.-M. Yuen, D. Cao, M. Sullman, D. Auger, C. Lv, H. Wang, R. Matthias, L. Skrypchuk, and A. Mouzakitis, "An orientation sensor-based head tracking system for driver behaviour monitoring," *Sensors*, vol. 17, no. 11, p. 2692, Nov. 2017. [Online]. Available: http://www.mdpi.com/1424-8220/17/11/2692
[47] V. Kazemi and J. Sullivan, "One millisecond face alignment with an ensemble of regression trees," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 1867–1874.
[48] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," 2015, *arXiv:1502.03167*. [Online]. Available: http://arxiv.org/abs/1502.03167
[49] N. S. Karuppusamy, "Multimodal driver fatigue detection system using deep neural networks," Ph.D. dissertation, School Mech. Eng., Kyungpook Nat. Univ., Daegu, South Korea, 2019.

NAVEEN SENNIAPPAN KARUPPUSAMY received the M.Tech. degree in machine design from Christ University, Bengaluru, India, in 2015, and the Ph.D. degree in mechanical engineering from Kyungpook National University, Daegu, South Korea, in August 2019. He has incorporated a company Naveenam Tech Private Ltd., India. His research interests include path planning for autonomous robots, evolutionary cooperative robots, industry 4.0, artificial intelligence, and deep learning for autonomous cars.

BO-YEONG KANG (Member, IEEE) received the B.S., M.A., M.S., and Ph.D. degrees from Kyungpook National University, Daegu, South Korea. She was with Seoul National University as a Research Professor. She also held a postdoctoral position with KAIST. She joined Kyungpook National University, in 2009, where she is currently an Associate Professor with the School of Mechanical Engineering. Her current research interests include artificial intelligence implementation for social and intelligent robots.