Article

Classifying the Cognitive Performance of Drivers While Talking on Hands-Free Mobile Phone Based on Innovative Sensors and Intelligent Approach

Boniface Ndubuisi Ossai [1], Mhd Saeed Sharif [1,*], Cynthia Fu [2,3], Jijomon Chettuthara Moncy [2], Arya Murali [4] and Fahad Alblehai [5]

[1] Intelligent Technologies Research Group, Department of Computer Science and Digital Technology (CDT), University of East London, London E16 2RD, UK
[2] Department of Psychological Science, University of East London, Water Lane, London E15 4LZ, UK
[3] Centre for Affective Disorders, Institute of Psychiatry, Psychology and Neuroscience, King's College London, De Crespigny Park, London SE5 8AF, UK
[4] School of Architecture Computing and Engineering, University of East London, London E16 2RD, UK
[5] Computer Science Department, Community College, King Saud University, Riyadh 11437, Saudi Arabia
* Correspondence: s.sharif@uel.ac.uk

Abstract: The use of mobile phones while driving is restricted to hands-free mode. But even in the hands-free mode, the use of mobile phones while driving causes cognitive distraction due to the diverted attention of the driver. By employing innovative machine-learning approaches to drivers' physiological signals, namely electroencephalogram (EEG), heart rate (HR), and blood pressure (BP), the impact of talking on hands-free mobile phones in real time has been investigated in this study. The cognitive impact was measured using EEG, HR, and BP data. The authors developed an intelligent model that classified the cognitive performance of drivers using physiological signals that were measured while drivers were driving and reverse bay parking in real time and talking on hands-free mobile phones, considering all driver ages as a complete cohort. Participants completed two numerical tasks varying in difficulty while driving and reverse bay parking. The results show that when participants did the hard tasks, their theta and lower alpha EEG frequency bands increased and exceeded those when they did the easy tasks. The results also show that the BP and HR under phone condition were higher than the BP and HR under no-phone condition. Participants' cognitive performance was classified using a feedforward neural network, and 97% accuracy was achieved. According to qualitative results, participants experienced significant cognitive impacts during the task completion.

Keywords: hands-free mobiles phones; wearable sensors; real-time driving; cognitive load; machine learning

1. Introduction

As a result of mobile phones' propensity to cause inattention and cognitive impairment, their use is restricted to hands-free mode while driving [1–4]. Nevertheless, during a crucial part of the conversation, when the leading vehicle slows down, there is a consequential chance of collision since the trailing driver may not be able to respond in a timely manner. Adult drivers who are not able to respond in a timely manner have higher heart rates, higher blood pressure, and altered neurophysiology [5–7]. The neurophysiological impact of talking on a hands-free mobile phone can be measured in a variety of ways. Some are based on heart-rate variability, skin conductance, skin temperature, and the drivers' respiration changes, whilst most are based on electroencephalography (EEG) technology, heart rate, and blood pressure. EEG signals are broadly used for comprehending how the human brain functions for different brain conditions. In addition to seizures, epilepsy,

encephalitis, brain tumors, memory loss, stress, and sleep disorders, EEG is also able to diagnose strokes and dementia [8].

EEG is one of the preferred techniques for capturing brain-generated electrical activity due to its high temporal resolution, low cost, portable nature, and low hardware complexity [8]. EEG has other applications in the areas such as bioengineering, psychology, psychophysiology, and video game biometric identification [8,9]. "Delta" (0.5–4 Hz), "Theta" (4–8 Hz), "Alpha" (8–12 Hz), "Beta" (13–30 Hz), and "Gamma" (more than 30 Hz) are the main frequency components of EEG signals [8,9]. The frequency bands that are represented in the dataset for this study are shown in Table 1. Studying the relationship between the theta band and the lower alpha band and a person's cognitive workload is the major focal point of this study.

Table 1. All EEG frequency bands.

EEG Wave	Freq. Band
Delta	0.01–4 Hz
Theta	4–8 Hz
Lower Alpha	8–10 Hz
Higher Alpha	10–12 Hz
Lower Beta	12–18 Hz
Higher Beta	18–30 Hz
Lower Gamma	30–40 Hz
Higher Gamma	40–50 Hz

Theta and alpha bands are strongly linked with focus and workload, according to most researchers who have evolved an interest in studying them either combined or as distinct units [10], both of which are notable indices of cognitive workload in a subject. Due to its direct correlation to increased task difficulty, the theta band (4–8 Hz) correlates closely with cognitive workload. The lower alpha band (8–10 Hz) exhibits a similar pattern when comparing the direct proportional relationship between rising attention and increasing workload. Theta band power increases with task difficulty, causing a greater workload, which ultimately increases lower alpha band power [10]. Therefore, we averaged the power spectra of theta waves and lower alpha waves to measure cognitive performance.

In this study, the authors hypothesize that the participant is deemed cognitively loaded and performs poorly on the task if the average of the amplitude of the theta and lower alpha EEG frequency bands are higher for a hard task than for an easy task, and if the BP and HR under phone condition are higher than the BP and HR under no-phone condition. Cognitive load relates to the amount of information that the working memory can hold at one time [11], and the driver can, therefore, be distracted. If the contrary is true, the subject's cognitive performance is seen as being good. The qualitative approach for this study focuses on the questionnaire, based on research on the cognitive load on drivers. The response provided by the drivers to the qualitative questionnaire provided the empirical evidence regarding their cognitive performance, which was used to verify the results of machine-learning techniques (quantitative method). Hence the hypothesis was validated. Several machine-learning algorithms were tested regarding the quantitative approach, and a neurophysiological model was developed that classified the cognitive impact of talking on hands-free mobile phones on drivers' performance (with an additional cognitive stressor).

The goal of the neural network used in this work is to classify drivers' cognitive performance using the neurophysiological indicators of drivers' cognitive load (HR, BP, EEG) into two classes (0 or 1) using a multi-layer perceptron neural network, where "0" represents "cognitively loaded" and "1" represents "not cognitively loaded". Artificial neural networks can be trained to find solutions as they consist of interconnected groups of neurons in a similar way to the biological nervous system. There are three types of

parameters that define them: firstly, the connections between the various layers of neurons. Secondly, the learning process that updates interconnection weights. Thirdly, the function that converts a neuron's weighted input into its output. A feedforward neural network algorithm begins with the inputs feeding in and the weighted sums of the inputs being added all together inside a hidden node. The hidden node's output is then biased in a specific direction [12].

An assessment of the literature showed that related studies have made extensive use of laboratory-driving simulator experiments. Without a question, this approach has some obvious flaws. Using a simulator experiment, the study [1] examined mobile phone usage characteristics, risk factors, compensatory strategies in use, and characteristics of high-frequency offenders. Four hundred eighty-four drivers (49.8% aged 17–25 and 50.2% aged 26–65) participated anonymously.

Furthermore, [2] examined the consequences of mobile phone use while driving, and the results showed that mobile phone use, regardless of the mode, affects driving behaviour. Nonetheless, driving-behaviour simulation only provides an approximation of reality and low-ecological validity [2]. Concerning sample characteristics, the participants' age range was 20–60 years, and most of the participants were young. A higher potential generalization of findings is evident in [1] compared to [2] because 49.8% were young and aged 17–25 in contrast with the case in [2] where most of the participants were young. Moreover, the age range of participants in [1] is more diverse than in [2].

An objectively equivalent hands-free phone task was used to assess heart rate and driving performance while late-middle-aged (51–66) and younger adults (19–23) engaged in a naturalistic task using a driving simulator [5]. However, based on the sample characteristics, it is uncertain what proportion of the sample represents late middle age and young adult, respectively. In contrast with the case in [2], the age group of drivers is more diverse in [5]. Hence, there is a higher potential generalization of outcome in [5] compared to [2]. The present study will seek to classify drivers' cognitive performance using the neurophysiological indicators of drivers' cognitive load due to talking on a hands-free mobile phone and driving in real time, considering all driver ages (18–80 years) as a complete cohort to maximize generalizability of the outcome in the context of age groups.

The upshot of the investigation into the frequency of mobile phone use by mode for making and answering calls showed that the proportion of drivers who reported making and answering calls (once a day or more) was higher amongst hands-free users, with 43% of hands-free users making calls and 47% answering calls, whereas these proportions were 17% and 21%, respectively, among handheld users [13] This manifests the need for more research into hands-free mode of mobile phone usage.

A scrutiny of related studies on the effect of added cognitive demands (phone tasks) while talking on a hands-free mobile phone in real-time driving using drivers of all age groups as a complete cohort has not been explored. The present study aims to fill in this gap. The authors aim to develop an intelligent model that classifies divers' cognitive performance while talking on a hands-free mobile phone and driving, using drivers of all age groups as a complete cohort. The foremost purpose of this study is to classify drivers' cognitive performance whilst talking on hands-free mobile phone during reverse bay parking using machine-learning approaches. To achieve this, several objectives were completed: measure the BP, HR, and EEG of drivers whilst they talk on hands-free mobile phone during reverse bay parking, collect and analyse the data, network modelling using Python, data training, testing, and validation.

2. Related Work

As a result of talking on a hands-free phone while driving, drivers are less likely to pay attention to the road. While driving, drivers' attention is divided between mobile phone use and driving tasks. The driving task may be interfered with. Consequently, the driver following may be unable to respond in time when the lead vehicle slows down [14].

Blood pressure and heart rate can be measured to examine the outcome of talking on a mobile phone while driving using cardiovascular reactivity (CV). The effects of talking on CV reactivity were studied in 60 participants with an average age of 19 years; 50% were male. Compared to the mobile phone task, participants' heart rates were remarkably lower with no task. A mean heart rate, systolic blood pressure (SBP), and diastolic blood pressure (DBP) were recorded electronically. The no-task heart rate was 82.69 beats per min, while the talking heart rate was 86.90 beats per min. For no task, SBP was 116.22 mmHg, whereas talking resulted in 118.93 mmHg. For DBP, no task was 65.30 mmHg. In contrast, talking resulted in a DBP of 66.81 mmHg. It was hypothesized that secondary tasks like talking on a mobile phone while driving might be as detrimental to the cardiovascular system as stress itself since CV reactivity has been linked to cardiovascular diseases such as heart attacks and strokes [7].

The EEG brain responses of a driver were examined in a study. According to test results, talking on a mobile phone while driving changed the EEG pattern. The driver used hands-free devices. A sample of 20 min of data was taken for the interval to ensure statistical significance. To clarify the change in the brain state, data of pre- and post-intervals of talking on the phone were measured consecutively, and the Alpha/Beta power ratio between normal driving and driving and talking on a mobile phone was compared from 20 min' consecutive measurements. Pre-20 min driving, in the first test, the mean Alpha/Beta power ratio was 0.7252. After 20 min of driving and talking on the phone, the first test was 1.2935. Post-20 min of driving, the first test was 0.8990 [6].

The application of machine learning in research has grown rapidly over the past three decades. Due to rapid technological advances, strict machine-learning algorithms, and the proliferation of vast data, machine learning has evolved significantly. However, due to a lack of training data, advanced machine-learning algorithms have not yet been widely applied to studies. Using BP, HR, and EEG signals, we explored a straightforward, efficient, and practical method of assessing the physiological impact of hands-free mobile phone use. BP, HR, and EEG analysis, feature extraction, and classification have been studied. BP, HR, and EEG data classification remains heavily dependent on feature extraction. Machine-learning-based BP prediction, HR estimation, and EEG classification have gained popularity in recent years [15].

Using pulse-transit time and pulse-waveform characteristics, a study developed a machine-learning-based model to predict systolic and diastolic blood pressure based on the shortcomings of conventional blood pressure-measurement methods and non-invasive continuous BP-measurement techniques. During the model-building phase, the genetic algorithm was used to optimize parameters. As a result of the experimental results, the proposed models were more accurate than conventional techniques and precisely captured the nonlinear relationship between characteristics and blood pressure, which had errors of 3.27 ± 5.52 mmHg for systolic BP and 1.16 ± 1.97 mmHg for diastolic BP. Ultimately, this study advanced the use of non-invasive continuous BP-estimation techniques [16].

Cardiovascular disorders are among the most prevalent and dangerous diseases worldwide [17]. Through artificial intelligence services, it is possible to enhance the diagnosis of these illnesses and even forecast their occurrence. For this study, four models were developed and trained. "PsychioNet Smart Health for Assessing the Risk of Events via ECG Database" was used to train the models to examine heart-rate variability features and forecast cardiovascular and cerebrovascular events. Support Vector Machines, Deep Neural Networks, and XGBoost all showed accuracy levels of 91.89%, 90.19%, and 89.50%, respectively, demonstrating the reliability of artificial intelligence in cardiology [17].

Several target emotions have been identified as having an impact on health and daily life, including boredom. Boredom was classified using EEG and galvanic response (GSR) in the study [18]. To inquire into the integrated effects of EEG and GSR on boredom classification, 28 participants used off-the-shelf sensors. A video clip designed to elicit boredom was used during data acquisition, as well as two video clips with entertaining content. Participants' questionnaire-based evidence of boredom was used to label samples.

The top three candidates were selected after training 30 models with 19 machine-learning algorithms. Validation of the final models was conducted through 1000 iterations of 10-fold cross validation after the hyperparameters were tuned. A Multilayer Perceptron model performed the best, achieving 79.98% accuracy. In addition, integrated EEG and GSR features were found to be associated with boredom. The results of this study can be used to build accurate affective computing systems and to understand boredom's physiological properties [18].

Several machine-learning approaches were evaluated in a mouse model for classifying EEG data of TBI (Traumatic Brain Injury). A variety of algorithms were assessed for classifying mild TBI data, extracted from data from the control group in wake stages for contrasting epoch lengths. These algorithms including decision trees (DT), random forests (RF), neural networks (NN), support vector machines (SVM), K-nearest neighbors (KNN), and convolutional neural networks (CNN). Alpha–theta power ratios and average power in different frequency sub-bands were used in machine-learning approaches. The results of this mouse model suggest that indistinguishable approaches can be used to detect TBI in humans. There was a higher level of reliability with the CNN method. To measure the reliability of each model, the average variance of all cross-validation experiments was calculated. CNN achieved 0.92%, while the KNN3, KNN5, and KNN7 achieved 1.93%, 1.94%, and 2.10%, respectively. As epoch length increased, CNN's performance increased, confirming its data-driven nature. Using 24 h recording data from nine mice, this study examines various machine-learning approaches to classify TBI. The model presented in this study requires further investigation and optimization since CNN is a data-intensive approach [19].

The study [20] also examined the literature on machine learning to identify some notable algorithms that are commonly used for analysis and prediction. Among these algorithms are Artificial Neural Networks (ANN), Random Forest, Support Vector Machines (SVM), Logistic Regression, K-means clustering, and K-Nearest Neighbor (KNN) [21–23]. SVMs (Support Vector Machines) have been proven to be very effective. SVM divides datasets into groups based on subspaces. A support vector is a set of points that lie close to all groups. As a result of its high precision and low computation power, this algorithm is widely used. A decision tree is generally used as a nonlinear classifier. It can also classify and train large amounts of data quickly and easily. Before reaching the confirmed class, each decision has a class. In this model, a branch represents a test result, a node represents a feature tested using transformation rules, and a leaf represents a class category [20].

K-nearest Neighbor (KNN) is also widely used in supervised learning to solve classification and regression problems. As well as being able to handle moderate amounts of data with good speed, it is easy to use and understand. Using the KNN method, an item is classified based on the majority vote of its nearest K-Neighbors, which can also be applied to supervised machine-learning problems. The algorithm handles moderate amounts of data quickly; it is straightforward to use and easy to understand. By using this method, items are categorized according to their nearest k-neighbors. The entire dataset is often used as a training set [20].

Artificial Neural Networks are among the most effective data-mining and analysis tools. Three layers of neurons exist: an input layer, a hidden layer, and an output layer. Artificial Neural Networks outperform rival methods in categorizing data because of their positive predictive value for detecting epilepsy. In a study in which an automated neural network classifier with 97.8 to 99.6% accuracy was developed, it was possible to determine whether a subject was normal or epileptic. In another study, ANNs were used to classify a driver's cognitive workload based on their ECGs. The classification model was satisfactory for both learning and testing data (95% and 82%, respectively) [12,24].

This study examined BP, HR, and EEG as physiological indicators of cognitive impact. The EEG of participants for hard tasks were compared with the EEG of participants for easy tasks, while the BP and HR under phone condition were compared with the BP and HR under no-phone condition to determine if the participants were cognitively loaded. The BP,

HR, and EEG signals were analyzed using SVM, KNN, decision trees, and feedforward neural networks. A higher level of task demand increases BP, HR, and EEG values, making BP, HR, and EEG arguably one of the most studied cognitive-impact indicators [25–27].

Based on physiological signals such as BP, HR, and EEG, this study focused on ML classification of drivers' cognitive loads. The qualitative approach for this study focused on the questionnaire, based on research on the cognitive load on drivers. The answers provided by the drivers to the qualitative questionnaire were used to find the ground truth regarding their cognitive performance. After testing several machine-learning algorithms, these answers were used to verify the results of machine-learning techniques (quantitative method).

It is expected that the subjects' theta and lower alpha EEG frequency bands for hard tasks will increase and exceed those for easy tasks. It is also expected that the subjects' BP and HR under the phone condition will be higher than their BP and HR under the no-phone condition [28–30]. As a result, the following hypothesis was developed: "When a subject's average amplitude of theta and lower alpha EEG frequency bands are higher for a hard task than for an easy task, and if the BP and HR under phone condition are higher than BP and HR under no-phone condition, the subject is considered cognitively loaded, resulting in poor performance. On the other hand, the subject is not cognitively loaded, resulting in good performance".

3. Methodology

The authors applied SVM, KNN, decision tree, and feedforward neural network machine-learning techniques to the data from EEG, HR, and BP signals. EEG is a measurement of potentials that reflect the electrical activity of the human brain. The brain waves recorded from the scalp have small amplitude usually measured in microvolts (μV) [30]. By comparing the amplitudes, one can determine which value is higher.

3.1. Subject Selection

The participants were healthy drivers of different age categories. Sixteen participants participated. However, the data from five participants were eliminated because the EEG device had failed to record during the experimental sessions. A total of 214 simulated data points were generated and applied in this study. This number also includes data from 11 participants [31]: five males and six females. The participants' ages range from 18 to 89 years, standard deviation 16.8, and mean age 42.9. Informed consent was provided by the contributors. The experiment included two tasks (easy and hard) for each participant. In addition, questionnaires were filled out by participants indicating their subjective experiences with workload. The questionnaire contained driver's age, gender, and driver group such as novice, experienced, or elderly.

The questionnaire form comprises various boxes. Each box represents a weighted value in percentage, with 0% being the lowest and 100% the highest. Each participant completed the questionnaire directly after the experiment by ticking the box which they think best represents the cognitive load from their experience during the experiment. This method was applied extensively to measure the cognitive performance of the participants whilst the tasks were done ("0" represents no impact, while "100" represents max impact) [32]. The sample questionnaire form is illustrated in Appendix A.

3.2. Experiment Protocol

The authors collected data during controls, easy tasks, and hard tasks using non-invasive devices. The EEG signal was collected using a NeuroSky MindWave EEG headset during the experiment, as described in Section 3.5. EEG values were collected from the device using the eegID application. Participants' blood pressures and heart rates were collected using an Omron blood-pressure monitor.

Control task involves reverse bay parking without phone use. Participant's baseline heart rate, blood pressure, and brain-wave activity were measured before the control task

driving began. During this task, participant's brain wave activity was recorded continuously. Heart rate and blood pressure readings were taken again within the experimental time frame [7,26].

The phone task involves reverse bay parking while talking on the phone. This task was split into two parts, namely: (1) easy task (2) hard task (one trial per task for each participant). The easy task procedure is as follows:

(a)   The experimenter presses the power button on the phone.
(b)   Participant makes a voice call by saying "Experiment".
(c)   The phone confirms and dials the number associated with "Experiment".
(d)   A pre-recorded message is played. The message is: "Count from 50 up to 200".
(e)   The participant will begin to reply to the message from the car park gate as they drive towards the bays to do bay parking.

A similar procedure applies for hard tasks, but with the following message: "Count backwards from 100, taking away 3 each time". Our hypothesis was designed based on the increment in the level of task difficulty in the order from easy task to hard task during the experiment, which allows critical evaluation of the changes in divers' performance due to the task difficulty increment from easy to hard. Hence, task randomization was not applied. We have used standardization process in which the procedure performed in the research was kept the same [1–4]. Great attention was taken to keep all elements of the procedure identical, so reliable results are measured. To minimize any remote possibility of order effects, adequate care was taken to ensure that participants had no prior knowledge of the tasks before the experiments began. One participant per experimental session was used to avoid anticipation.

3.3. Data Collection and Data Description

As mentioned above, during the experiment, EEG signals were collected using the NeuroSky Mindset headset. The MindWave Mobile 2 headset was first paired to a mobile phone for them to connect to each other. The connection is wireless via Bluetooth. The headset was first turned on to make sure the Blue LED light is on, and the pairing instruction was followed. The name of the application is eegID. Recording was started only once the poor signal value reaches "0", as instructed by the users' guide. Data were exported to the experimenter's research phone once each recording was completed.

Numeric values were recorded by the device, and these numbers represent the wearer's mental state; for example, their attention level. For the main data file on an Excel spreadsheet, the first column shows the timestamp of each sample in milliseconds (ms), while the second column represents the poorSignal value ("0"). The following columns are ordered from left to right: the subject's eegRawValue, attention levels, meditation levels, blink strength, delta band, theta band, alphaLow band, alphaHigh band, betaLow band, betaHigh band, gammaLow band, gammaMid band, tagEvent, and location.

Neurosky MindSet EEG headset places one electrode on top of a person's forehead, and the other electrode clipped to the left earlobe. It records numerical numbers based on their own algorithms. Algorithms are primarily applied to attention and meditation values. In this case, instead of using Volts-squared per Hz to calculate power spectrum band power, the device uses its own representation, which results in using a different unit for its final output. Consequently, the numeric data it gathers from the initial voltage measurement is transformed and rescaled several times before it is recorded. Therefore, no conventional unit has been assigned to the values. They are simply called ASIC_EEG_POWER units and are used only to output the numeric values from the Neurosky MindSet EEG headset. As described above, since they have their own units, the readings can only be meaningful when compared among themselves. Consequently, comparing them directly with, for example, the values output by another EEG system would not be meaningful or accurate. Their current output form can be used to determine whether each band is increasing or decreasing over time, and how strong it is compared to other bands [10,33].

OMRON M7 Intelii IT Blood Pressure Monitor was used to collect the HR and BP data. It also has the benefit of Bluetooth capacity. Similarly, it was paired to a mobile phone for them to connect to each other. The name of the application is "OMRON connect". The cuff was wrapped around the participant's upper arm, so it is level with the participant's chest, ensuring the tubing falls over the front centre of the participant's arm. The end of the cuff is pulled so the sensor is correctly placed and evenly tight around the arm. The cuff inflates when the start button on the unit is pressed. The readings are taken once the cuff reaches maximum and stops inflating, and the figures on the display are steady. All measurements were transmitted directly via Bluetooth to the experimenter's research phone during the experiment. The device records numeric numbers, representing the participant's BP and HR, respectively. For the main data file on an Excel spreadsheet, the first column shows the date, while the second column represents the recording times. In order from left to right, the following columns are listed: the subject's Systolic BP in mmHg, Diastolic BP in mmHg, and HR in bpm.

3.4. Feature Extraction and Data Processing

To ensure accurate and reliable readings, this study's experiment was done such that the EEG recorded continuously during the tasks, while the HR and BP readings were recorded within the experimental time frame [7,26]. A brief description of the steps, and how EEG, BP, and HR readings of each participant were taken is as follows: participants' prescribed bay parking procedure was explained and practiced for 15 min. Participant rested for 5 min [34]. The experimenter measured baseline blood pressure and baseline heart rate, respectively. Experimenter set the EEG (Mindwave) app to record continuously. Participant drove from car park gate and completed bay parking without phone use in the car. Experimenter stopped the EEG (Mindwave) app from recording. Participant's blood pressure and HR were measured again. Participant rested for 5 min [34]. Just before the easy task began, the experimenter set the EEG (Mindwave) app to record continuously. Participant drove from car park gate and completed bay parking while talking on the phone. Participant's EEG recording was saved. Participant's blood pressure and HR were recorded. Participant rested for 10 min [34]. Just before the hard task began, the experimenter set the EEG (Mindwave) app to record continuously. Participant drove from car park gate and completed bay parking while talking on the phone. Participant's EEG recording was saved. Participant's blood pressure and HR were recorded. All the EEG frequency bands are represented in the particular dataset for this study, which creates eight groups based on their input values as shown in Table 1 below [10]. Figures 1 and 2 below show pictures of the experimental testing site.

Figure 1. Experimental testing site entrance.

Figure 2. Experimental testing site car park.

Considering that this experiment uses an EEG device that already provides the EEG power spectrum dataset, the analysis, therefore, tilts towards creating a training dataset by combining the identified values. As mentioned above, theta and alpha bands are strongly linked with focus and workload [10], both of which are significant indicators of workload in a subject. Due to its direct correlation to increased task difficulty, theta band correlates closely with workload. The lower alpha band exhibits a similar pattern when comparing the direct proportional relationship between rising attention and increasing workload. In the presence of a challenging task, theta band power increases, resulting in an increased workload and a subsequent increase in lower alpha band power [10]. As a result, the authors used averaged power spectra of theta waves and lower alpha waves as inputs when establishing the subjects' cognitive performance in this study.

The EEG readings were recorded in blocks. Each reading iterated about 512 times, after that a different reading records and iterated for the same number of times, and it carries on in this pattern down the line, generating numerous blocks of readings. Each block had duplicate readings. Rather than adding the readings in every block and finding the average per block, which will be daunting and time inefficient, the authors have simply taken a reading from each block (same as average) and computed the total reading for all the blocks and then calculated the average [10]. This calculation was done for theta band extracted during hard task and easy task, respectively. The computation was repeated for lower alpha band extracted during hard task and easy task, respectively [10].

The calculated figures were further converted to two decimal places, respectively, to make them useable as inputs for the machine-learning classification [10]. Additionally, input values include the driver's average heart rate for Task 1 and Task 2 (phone condition), driver's average blood pressure for Task 1 and Task 2 (phone condition), driver's age, and gender. Predicted output from the classifier is the binary class depicting whether the driver is cognitively loaded or not. Class 1 represents "not cognitively loaded", whereas Class 0 represents "cognitively loaded". By calculating and using the average of each set of theta and lower alpha samples, respectively, for analysis, this helped to create a standardized set of data because during the experiment, the number of theta and lower alpha samples recorded by the EEG device for each participant varied from one participant to the other [10]. This is because the bay parking duration was not the same across the participants. The experiment was designed to allow the participants to complete the parking task while talking on the phone. Therefore, the authors averaged the samples as mentioned above [10].

3.5. Measurement Process Flow

The flowchart in Figure 3 illustrates the sequence of actions, movements, and decision points within the system, providing a detailed overview of the project. The workflow has employed 214 subjects. Data were recorded while driving with and without hands-free

mobile phone, respectively. The project was designed to ensure data was recorded three times without fail, namely: control, easy task, and hard task according to the first decision box on the flow chart. As described above in the "Data Processing" section, considering that this experiment had used an EEG device that already provides the EEG Power Spectrum dataset, the analysis, therefore, tilted towards creating a training dataset by combining the identified values. Therefore, the averaged power spectra of theta waves and lower alpha waves were computed and converted to two decimal places, respectively, to make them usable as inputs for the machine-learning classification. Further data processing includes averaging driver's heart rate for Task 1 and Task 2 (phone condition), and averaging driver's blood pressure for Task 1 and Task 2 (phone condition).

For the ML classification, features selection was done by selecting the most critical variables and eliminating the redundant and irrelevant ones. The variables are as follows: theta band for easy task, theta band for hard task, lower alpha band for easy task, and lower alpha band for hard task. Additionally, extracted values include the driver's average heart rate for Task 1 and Task 2 (phone condition), driver's average blood pressure for Task 1 and Task 2 (phone condition), driver's age, and gender. After extensive literature survey, four algorithms were selected for the ML classification. The models created followed a step-by-step process as follows:

(a) Importing libraries such as matplotlib, which supports graph plotting and pandas for making dataset into data-frame.
(b) Importing dataset
(c) Processing dataset such as removing null values and converting categorical values to integers such as "1" and "0".
(d) Data analysis and correlation.
(e) Model creation, training, testing, validation, and visualization of the outputs.

For each machine-learning algorithm, the authors trained and tested the model, evaluated the model's accuracy, and validated the hypothesis. Training dataset = 80%, testing dataset = 20%. Target data = binary data points (1 or 0), where 1 = not cognitively loaded, and 0 = cognitive loaded. Results were obtained from the classifiers, respectively, and the classification accuracies for the classifiers were compared. A classifier with the highest accuracy is said to have the best performance. This is because a classifier's performance is measured by its classification accuracy. In this project, the ANN achieved the highest accuracy out of the four algorithms used.

The block diagram in Figure 4 is an illustration of the project representing the significant parts. The first block represents EEG, BP, and HR acquisition from the subjects using non-invasive devices as described in Section 3.3 above. As described in Section 3.4, this experiment used an EEG device that already provides the EEG Power Spectrum dataset, the analysis and processing, therefore, tilted towards creating a training dataset by combining the identified values. Therefore, the averaged power spectra of theta waves and lower alpha waves were computed. Further data processing includes averaging driver's heart rate for Task 1 and Task 2 (phone condition), and averaging driver's blood pressure for Task 1 and Task 2 (phone condition). Features extraction was done by selecting the most critical variables and eliminating the redundant and irrelevant ones. The variables are as follows: theta band for easy task, theta band for hard task, lower alpha band for easy task, and lower alpha band for hard task. Additionally, extracted values include the driver's average heart rate for Task 1 and Task 2 (phone condition), driver's average blood pressure for Task 1 and Task 2 (phone condition), driver's age, and gender. The fourth and fifth blocks represent the action performed in the ML process where modeling and classification tasks were done, resulting in ANN achieving the highest accuracy amongst the four algorithms employed in this project.
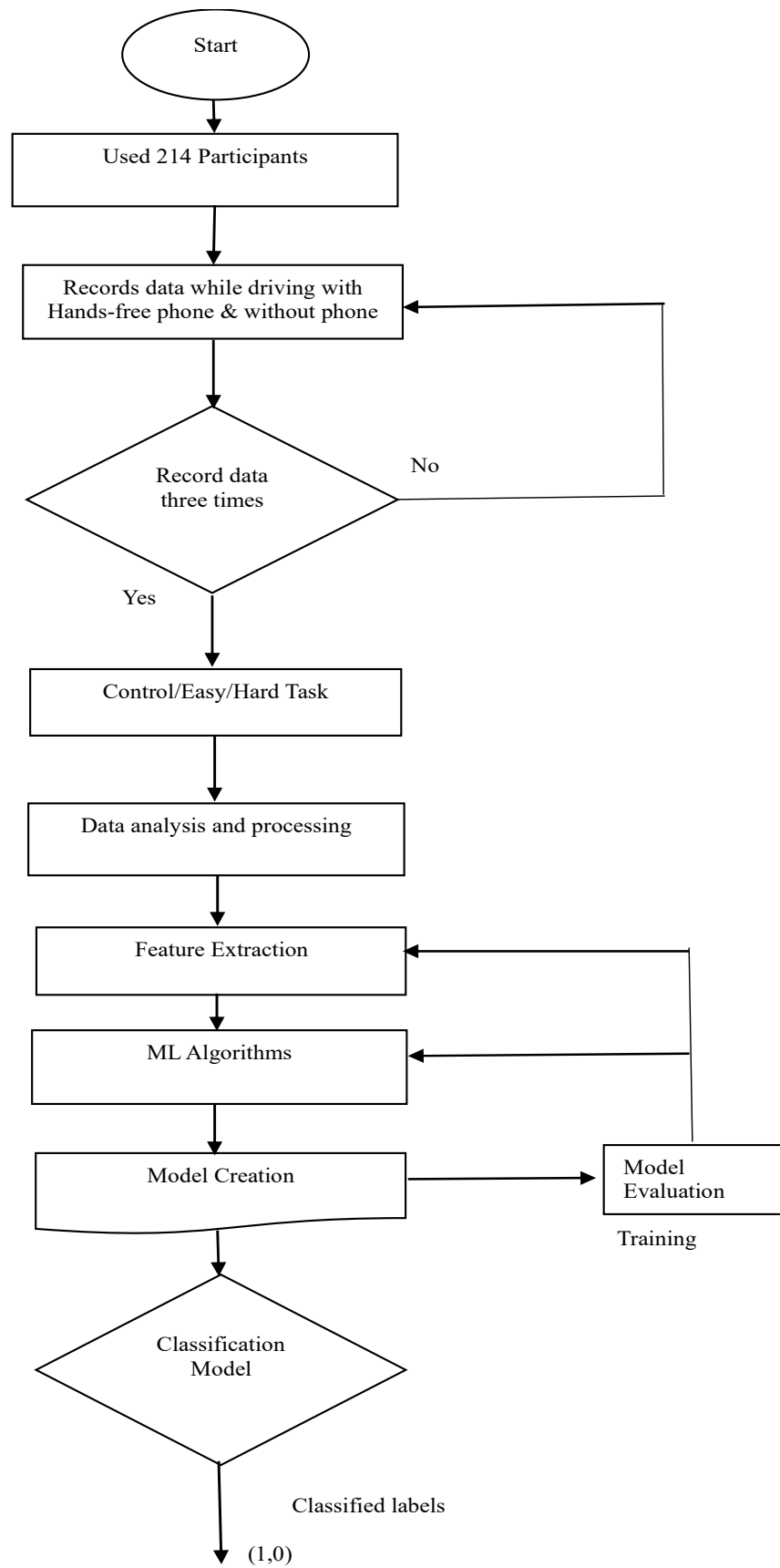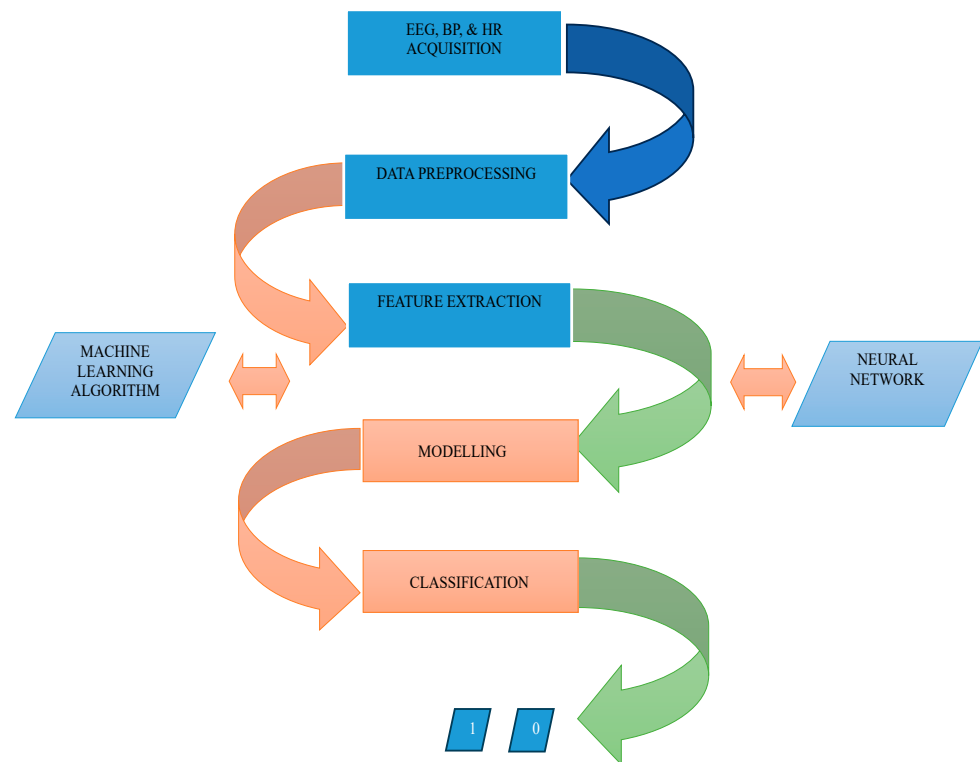
Figure 3. Flowchart of driver's cognitive performance.

Figure 4. Block diagram of the project.

4. Modelling and Implementation

A combination of SVM, KNN, decision tree, and ANN were selected for the proposed approach. Initially, the data were processed for each machine-learning algorithm. Features selection was done by selecting the most critical variables and eliminating the redundant and irrelevant ones before being employed to train and test the model and validate the proposed hypothesis. Training data accounted for 80% of the data, while testing accounted for 20%. To validate the hypothesis, two datasets were created using associated parameters that would show a clear impact on the subjects' performance. The first dataset consists of the main parameters from EEG signals, such as theta and lower alpha bands, diver's average HR for Task 1 and Task 2 (phone condition), and driver's average BP for Task 1 and Task 2 (phone condition), whereas the second dataset consists of data collected from subjects such as their age and gender. Data points with binary values were used as targets. A target data value of 0 indicates "cognitively loaded" and a value of 1 indicates "not cognitively loaded".

To compare the classification accuracy, four classifiers were used. An important factor in the performance of a model is its accuracy, or how well its predictions match the data [9,10,35]. Hence, each model's accuracy must be properly assessed. The authors collected a small part of the dataset for validation to evaluate the model's accuracy. All four techniques were implemented in Phyton. Python is a high-level, general-purpose programming language. There are many researchers who use this programming language for analyzing data, developing algorithms, and machine-learning modelling [36]. The following paragraphs provide an overview of the computing method.

Firstly, Python scripts were used to import several library functions for data analysis, machine learning, and data visualization. In addition, the script can load and preprocess data, train and evaluate various classifiers, and visualize model performance. The dataset used in this project contains 214 entries with 11 columns on an excel spreadsheet. The columns include physiological data such as heart rate, blood pressure, and EEG frequency bands, as well as demographic data such as age and gender. EEG, blood pressure, heart

rate, age, and gender are inputs to the classifier. There are no missing values in the dataset, and both numerical and categorical data are included.

The computing method began with importing and reading pertinent data using codes. A CSV file containing the relevant data was imported and read using the read () function from Python. The code provided, imports the dataset from the csv file "finaldataset.csv" and uses pandas read_csv () to read it into a dataframe. After the data were imported and read into the dataframe, it was processed and cleaned, including handling missing or invalid values, encoding categorical variables, and splitting the data into training and testing sets. The total number of rows indicated that the dataset has 214 records or observations, while the total number of columns suggested that the dataset has 11 variables or features. To ensure that there were no null values, data information was printed. After this, machine-learning classification algorithms were applied to create models to classify the cognitive performance of the drivers. The machine-learning models were evaluated for reliability and precision. Results show that the ANN model produced the highest accuracy, with an accuracy rate of 97%.

For the ANN, the trainable parameter in the model includes weights and biases of all the layers. The number of parameters for each layer was computed as the product of the number of inputs and the number of neurons in the layer, plus the number of biases. The size of the output may vary depending on the size of the input batch during training. Epoch or iterations = 650. Batch size = 8. The training and validation epoch counts showed the model accuracy increasing over time. The number of epochs normally decides how many times the weight of the network is changed. For example, as the number of epochs increases, weights are changed in the neural network the same number of times, and the boundary goes from underfitting to optimal. If the number of epochs is too small, the model may not learn the underlying patterns in the data resulting in underfitting. On the other hand, if too large, it results in overfitting [37]. Overall, the findings imply that the neural network model can be utilised to make predictions because it fits the available data well.

5. Result and Analysis

For data analysis in this research, the simulated 214 data points represent the participants. The performance of the algorithms was compared. Results and analysis also included the use of histograms as shown below from Figures 5–10. By reviewing the ROC curve in Figure 11 and testing accuracy graphs in Figure 12, the models' effectiveness was evaluated. Histograms are graphs that show how continuous data are distributed. They display the frequency with which values fall into certain categories. The number of items in the dataset that fall into a certain category is shown by the height of each bar.

Figure 5 below shows the age distribution of the participants, older drivers 40–69 years are the majority (65%), while young (18–39) and elderly (70 years and above) are approximately 19% and 16%, respectively. From Table 2 below, participants average blood pressure without phone = 114.52, while average blood pressure with phone is 121.30. Participants average HR without phone = 71.07. Average HR with phone = 77.43. Theta easy average = 2,430,000.00, SD = 1,362,906.00, whereas theta hard average = 2,890,000.00, SD = 1,435,002.00. Average lower alpha easy = 3,080,000.00, SD = 3,842,181.00, whereas average lower alpha hard = 3,540,000.00, SD = 5,323,524.00. The results from Table 2 below shows that for all the participants, average theta and average lower alpha readings are higher for a hard task than for an easy task. Also, the average BP and average HR under phone condition are higher than the average BP and average HR under no phone condition. The EEG graphs from Figures 6–9 have been scaled to show a clear statistic distribution. The figures have been represented in standard forms. For example: $499,000.00 = 4.99 \times 10^5$, which has been represented as 4.99E+5 on the graph [10]. Standard form is a way of writing very small or very large numbers by comparing the powers of 10 [38]. It is also known as scientific notation.

The values on x axis from Figures 6–9 represent EEG groups. The height of each bar represents the number of data points (participants) that fall within each EEG group as shown on the graphs. For example, in Figure 7, two data points fall within the EEG group (0–4.99E+5), while 12 data points fall within the EEG group (4.99E+5–8.99E+5), and so on. Similarly, the values on the *x*-axis in Figure 10 represent participants' self-reported cognitive load. Table 2 below shows a summary of the statistical analysis of the results from BP, HR, theta, and lower alpha EEG signals in this study. The results from BP, HR, theta, and lower alpha EEG signals as shown in the table corroborate the outcome of our qualitative survey and endorse our hypothesis as stated above.
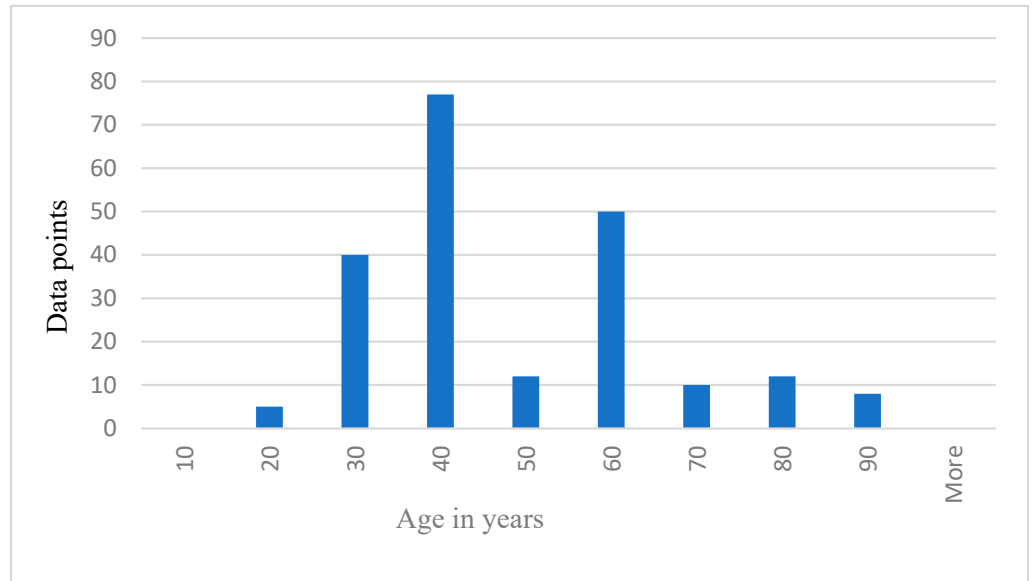
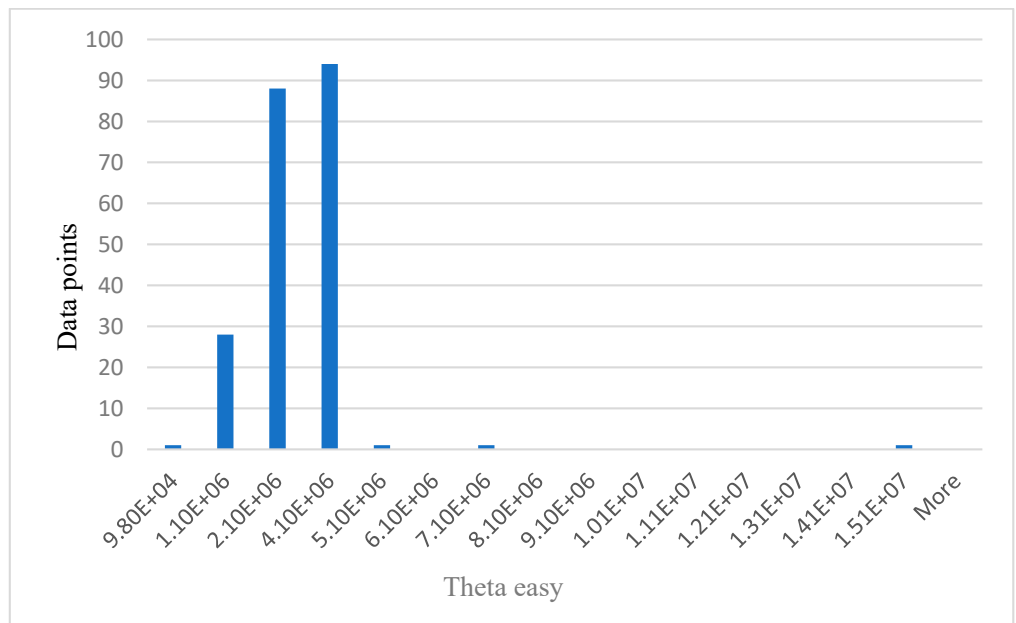**Figure 5.** Data points representing age distribution of the participants.

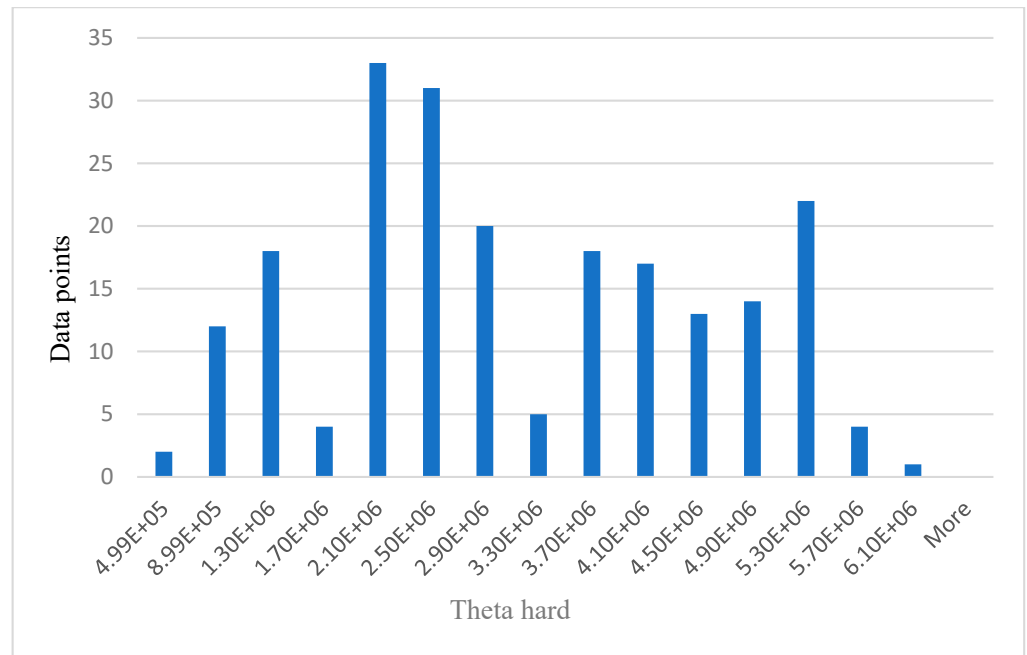**Figure 6.** Relationship between data points and theta easy.

Figure 7. Relationship between data points and theta hard.
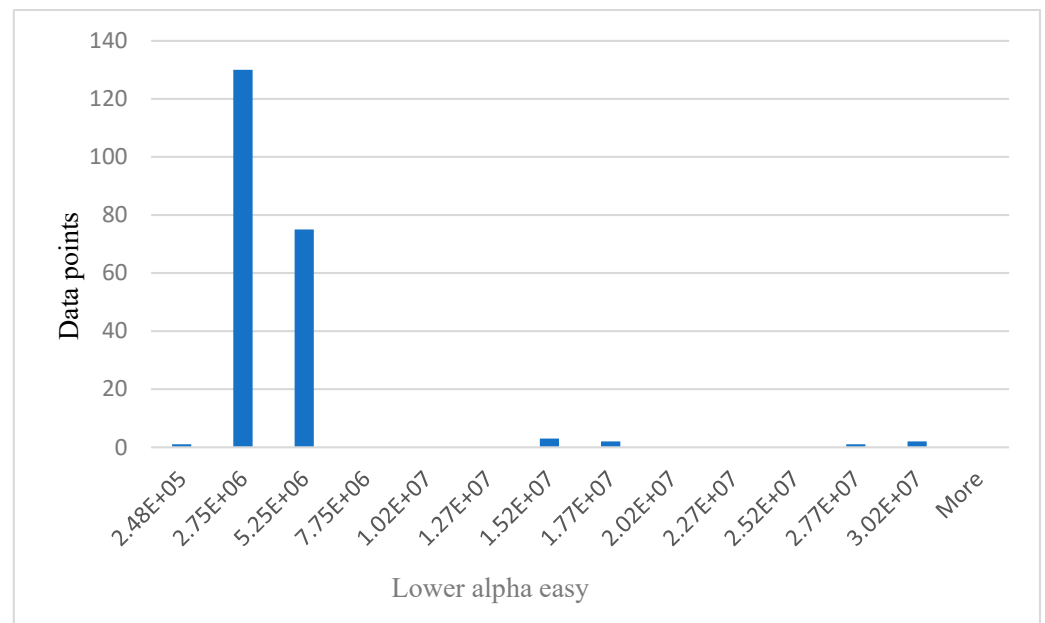

Figure 8. Relationship between data points and lower alpha easy.

About 25 data points have been randomly taken from the total data samples and plotted to show the distribution of self-reported cognitive load in percentage because of the additional cognitive stressor (easy task and hard task) as shown in Figure 10 below. The number of data points representing 70% or over = 22. Number of data points representing less than 70% report = 3. The self-reported cognitive load supports and validated our hypothesis.

As was noted previously, the effectiveness of the models was assessed by visualizing the ROC and model-accuracy graphs, the algorithms' results were compared, and it showed that the feedforward neural network attained a 97% accuracy rate. Figures 11 and 12 below show the ROC and model-accuracy graphs for the feedforward neural network, respectively. While this study's dataset was sizable, it is nonetheless not a large dataset such as population dataset. A large dataset could lead to improved ROC performance [39]. The

study from [39–41] also noted that the size of a dataset plays a key role in a model's performance. Also, [39] added that for the application of deep learning, such as in geosciences, many papers currently published use very small datasets of less than 5000 samples, which has a direct implication. The study from [42] has also demonstrated that a small sample size leads to large inaccuracies in the estimated validation parameters associated with ROC analysis. Table 3 below shows the accuracy result comparison of the classifiers used in this study.
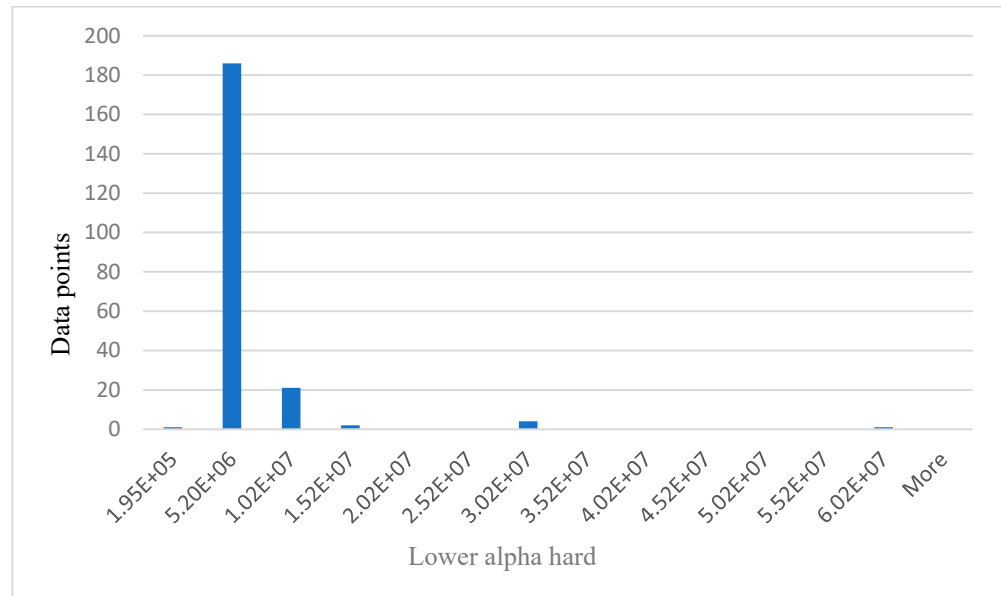

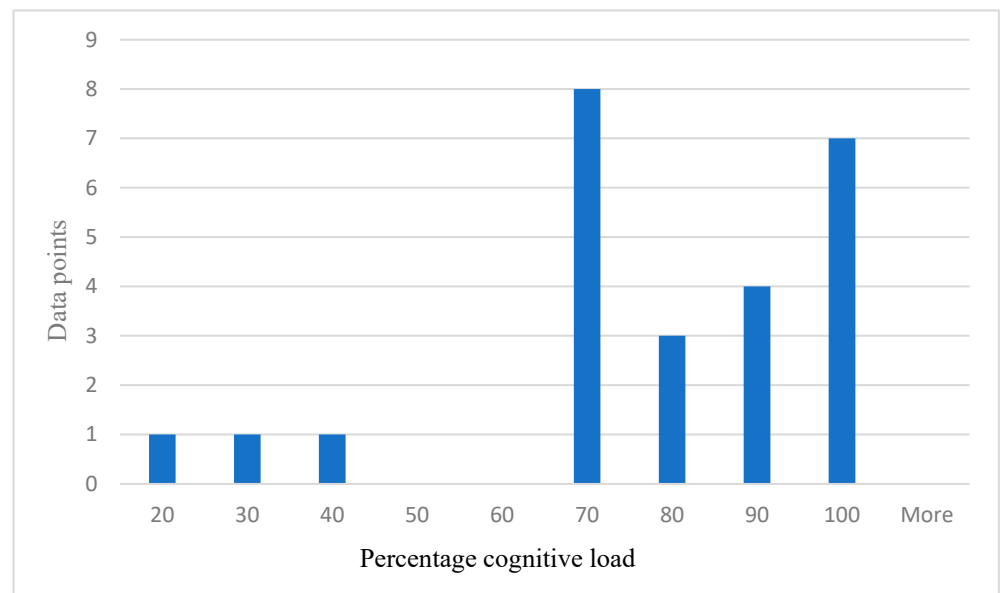**Figure 9.** Relationship between data points and lower alpha hard.

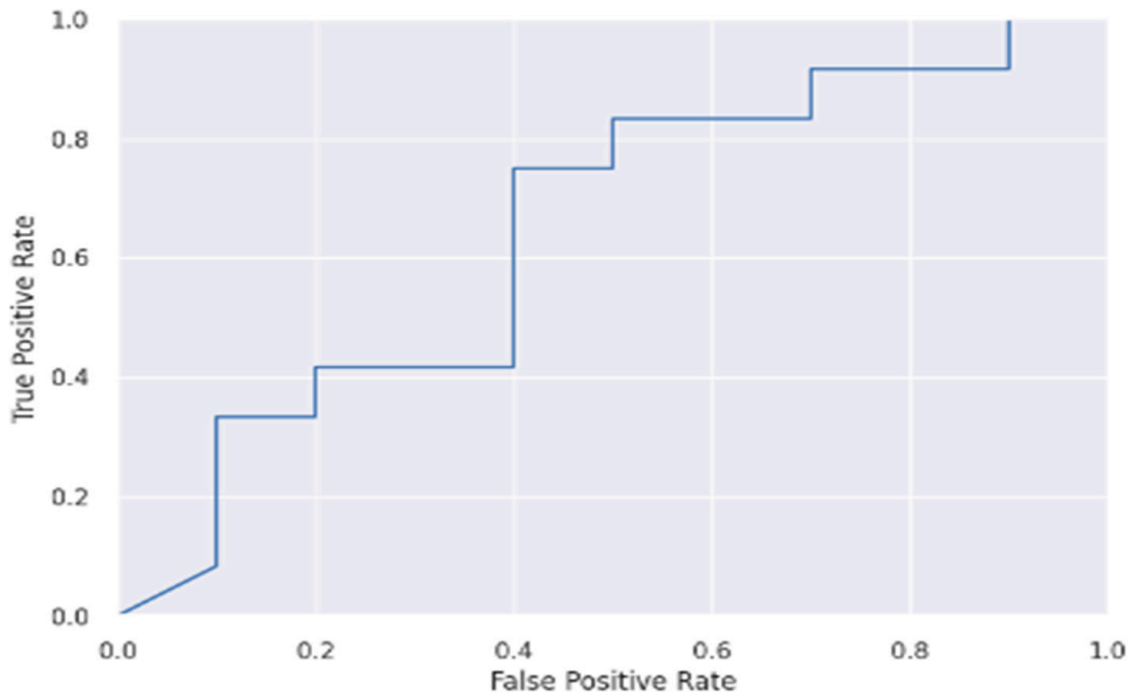**Figure 10.** Distribution of self-reported cognitive load.
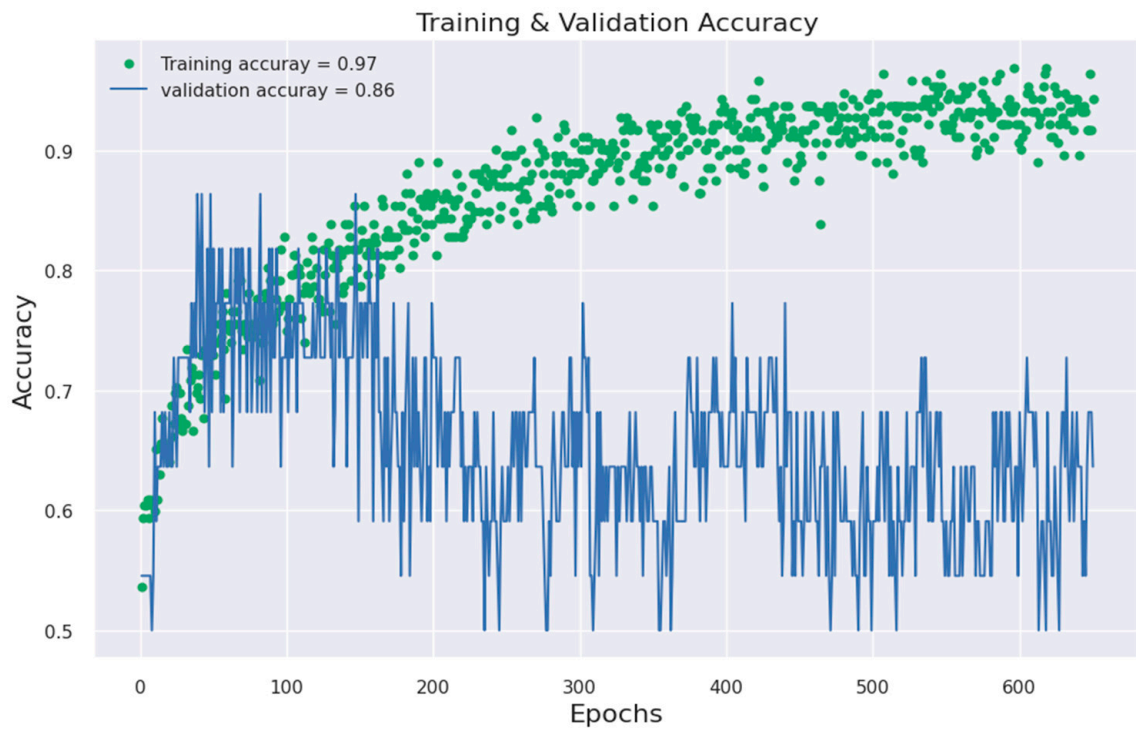
Figure 11. ROC curve of ANN.

Figure 12. ANN model accuracy against epoch.

Table 2. Results from BP, HR, theta, and lower alpha signals.

Statistical Measures of EEG, HR, and BP	Average	Min	Max	SD
BP without phone	114.52	80.00	142.00	15.35
BP with phone	121.30	89.00	151.00	13.19
HR without phone	71.07	55.00	110.00	12.64
HR with phone	77.43	55.00	115.00	13.20
Theta easy	2,430,000.00	4.92	14,500,000.00	1,362,906.00
Theta hard	2,890,000.00	2.28	5,940,000.00	1,435,002.00
Lower alpha easy	3,080,000.00	1.10	29,500,000.00	3,842,181.00
Lower alpha hard	3,540,000.00	2.55	55,500,000.00	5,323,524.00

Table 3. Accuracy results comparison.

Model	Accuracy
ANN	97%
K-Nearest Neighbour	72%
Support Vector Machine	75%
Decision Tree	65%

6. Comparison of Classification Results with Other Literature

The analysis of the highest-performing classifier is based on this research design. As a result, Table 4 below compares experimental results with other literature. A list of related literature is shown in the first column. This list of literature has been gathered from the reference list of this paper, showing their respective numbers. The second column shows the experimental results using ANN classifiers. With an accuracy of 97%, Table 4 shows that the present research performed remarkably in terms of results. By creating more awareness of the drawbacks of talking on a hands-free mobile phone, this study benefits public safety and the UK Department of Transport. Therefore, policy makers can examine the existing evidence about the risks of talking on a hands-free mobile phone while driving. As a result of this research, the government and policy makers will be able to evaluate and redefine their safety promotion strategy on the roads.

Table 4. Comparison of classification results with other literature.

List of Literatures	Classification Results Using ANN
12	95%
17	90.19%
18	79.98%
19	92%
Present study	97%

7. Discussion

Driving while using a handheld mobile phone is prohibited due to the distraction and cognitive impairment it creates. As an alternative, hands-free mobile phones are allowed while driving. But using a hands-free phone while driving raises EEG, BP, and HR in adults [43–45]. However, an overview of the literature unfolded that the effect of added cognitive demands (phone tasks) while talking on a hands-free mobile phone in real-time driving using drivers of all age groups as a complete cohort has not been investigated.

The present study has filled this gap. This study used machine-learning techniques to analyze the electroencephalogram (EEG), heart rate (HR), and blood pressure (BP) signals of drivers and classified the cognitive performance of drivers using EEG, HR, and BP measured while talking on hands-free mobile phones in real time. The authors employed both qualitative and quantitative methods. The qualitative approach for this study focused on the questionnaire, based on research on the cognitive load on drivers. The answers provided by the drivers to the qualitative questionnaire were used to find the ground truth regarding their cognitive performance. After testing several machine-learning algorithms, these answers were used to verify the machine-learning techniques (quantitative method), and the hypothesis was, therefore, validated. Using EEG, very little research has looked at the physiological impacts of talking on a hands-free mobile phone while driving [6]. In this research, talking on a hands-free mobile phone while driving caused a significant difference between drivers' average amplitude of theta EEG frequency band for easy tasks and drivers' average amplitude of theta EEG frequency band for hard tasks.

A similar notable difference was observed between drivers' average amplitude of lower alpha EEG frequency band for easy tasks and drivers' average amplitude of the lower alpha EEG frequency band for hard tasks. Although not including real-time driving, studies such as [6] have made similar findings to ours. From Table 2 above, results from theta and lower alpha EEG signals show that participants' average theta band signal for easy tasks is 2,430,000.00, whereas the average theta band signal for hard tasks is 2,890,000.00. Similarly, the participants' average lower alpha band signal for easy tasks is 3,080,000.00, while the average lower alpha band signal for hard tasks is 3,540,000.00. This contrast is attributive to the cognitive demand because of the additional cognitive stressor (phone task). This result is consistent with the outcome of our qualitative results as expressed below.

Unsurprisingly, participants' amplitude of theta and lower alpha EEG bands increased when performing the hard tasks, exceeding the amplitude of theta and lower alpha EEG bands during easy tasks. Also, participants' BP and HR under phone condition were higher than the BP and HR under no phone condition. The quantitative findings indicate that the tasks had a considerable cognitive impact on the participants. Ninety-seven percent accuracy was reached when classifying the participants' cognitive performance using feedforward neural network. Our findings have made prevalent reference to the assertion from [10,30], which pointed out that as task difficulty increases, the theta and lower alpha EEG frequency band increases, causing an increase in cognitive workload. Similarly, a reference can be made to the study from [10,30] where the relationship between theta and lower alpha bands and a person's cognitive workload was also highlighted.

Furthermore, the research from [8–10,35] makes it clear that the EEG device used in this experiment already provides an EEG Power Spectrum dataset. These studies have applied similar EEG devices in their experiments. To this end, to determine the subject's level of workload, and as a major part of the signal processing for this study, the input values to our ML model are the averaged power spectra of theta waves and lower alpha waves. There is, nonetheless, an alternative method of signal processing such as Fast Fourier Transform (FFT). The Fast Fourier Transform algorithm, as its name suggests, determines the Discrete Fourier Transform of an input much more quickly than computing it directly [30]. However, this is not needed for our study for the aforementioned reason.

The distribution of data samples representing drivers' self-reported cognitive load in percentage because of the added cognitive stressor (easy task and hard task) was such that 22 data points represent a 70%-plus response rate, while three data points represent drivers reporting less than 70% in total. One could counter that this sample is too small to generalise the findings. However, the overall results from the 214 data samples in this study are consistent with the results from the 25 data samples that were randomly chosen.

The authors acknowledge the research from [46], where an estimation of older drivers' cognitive performance and workload using features of eye movement and pupil response on test routes was done. Since the test was conducted on actual roads and the results were reflective of actual driving behaviour, this study and ours have similar high-ecological validity [27,46]. The authors note that drivers travelled 10 kilometres on various types of roads with three right turns, two left turns, seven crossroads, and four straightaways, but it is not known at what speed they drove and if they were meant to drive at the same speed or not. It is unclear what effect that may have had on drivers' eye movement and cognitive load.

In this research, participants drove in a built-up area in the Wood Green driving test area in London, comprising approximately 500 yards and consisting of three turns. Participants drove the same car, which was approved only for the experiment by the University of East London's research ethics committee. Although the same driving route was assigned to all participants, driving conditions for each driver were expected to be different because they drove on public roads where factors such as weather and the influence of other vehicles varied beyond the control of the experiment. However, the phone tasks were not applied to the participants until they had entered the driving test center car park. The car park bay parking routine comprises one car per time. Therefore, the influence of other vehicles and external factors is expected to be extremely minimal.

According to [46], participants drove their own car, which can be a good reference in terms of generated results because they will be familiar with the controls. However, it is not known what kind of gear transmission was used and if the gear transmission was the same amongst drivers because a manual gear transmission will certainly engage the drivers more than an automatic gear transmission. This is bound to increase drivers' cognitive load as they drive. Our experiment has used one car for all the participants (automatic gear transmission).

Furthermore, in [46], regression analysis was introduced to create relationships between cognitive performance and extracted features. Several metrics of conventional eye movement were extracted from the recorded data. The extracted variables include horizontal eye-movement deviation, vertical eye-movement deviation, mean pupil size, etc. A linear regression model's primary benefit is its linearity: The other benefits include being easy to implement and working best with linear data, whereas some drawbacks include assuming that the data is independent and being prone to underfitting [47]. The following papers constituted the foundation of the work presented in this article [1,2,5,7,10,13,25,26,28,32,33].

There are several clear drawbacks and advantages affiliated to this work. Firstly, without undermining the scope of the present study (human drivers' cognitive distraction), if the human driver is replaced by an artificial intelligence (AI) driver, cognitive distractions related to talking on the phone while driving becomes irrelevant because the robot driver is the driver in this context [48]. However, it is imperative to examine whether voice conversations between the robotic driver and the passenger could distract the robotic driver. In that case, the question becomes "How can it be monitored or controlled"? There is a potential challenge to the present research in this regard. The safety of autonomous cars (driverless cars) raises important concerns about their widespread acceptance and the shift from human drivers to autonomous cars. In general, automated car safety and human interaction are well-established fields. Among autonomous vehicles, safety is divided into cyber security and road safety. Road safety involves the physical safety of the vehicle while driving and interacting with other autonomous and human-driven vehicles. Cyber security refers to the vulnerability of onboard computers to attacks on their functionality and sensitive information, as well as unauthorized access to the private information of passenger [48].

According to [48], autonomous vehicles contribute to active passenger safety by preventing collisions. Active safety depends on the system's ability to detect threats and respond appropriately in those situations. It is possible to measure responsiveness by taking the product of the vehicle's speed and response time in relation to its look-ahead distance. Equation (1) provides the response ratio, where v is the vehicle speed, T is the response time, and d is the sensor look-ahead distance.

$$\text{Response ratio} = \frac{vT}{d} \tag{1}$$

In urban driving scenarios, the sensor look-ahead distance could be replaced with the vehicle gap to compute responsiveness. For navigation safety at high speeds of about 100 km per hour, with small gaps of approximately 10 m, the response time of the system must be on the order of milliseconds. For the vehicle to be more responsive, the system must be able to identify dangerous scenarios in real time. Also, [49] discusses active safety systems for autonomous vehicles in unstructured areas, emphasizing the importance of perception reliability. As of now, the reliability of autonomous cars' mechanical and electronic components is well studied and addressed [50]. As a result, software reliability is the key challenge regarding the safety of autonomous cars. Passenger safety is largely determined by the way vehicles respond and recover from unexpected circumstances, such as sensor failures, data loss, and communication breakdowns.

Furthermore, ref. [51] explored conversations between an autonomous vehicle and its users to identify elements of trust in the dialogue. Through their interaction with the vehicle, ref. [51] examined the extent to which people engaged in "actual conversation" with the vehicle, as well as their attempt to establish emotional and interpersonal connections. In general, even when the autonomous vehicle offered conversational interaction, people tended to engage in command-based interactions, and only occasionally did they embrace a more conversational approach despite the vehicle's apparent capacity to engage in conversation. This behaviour corroborates major concerns expressed by the public regarding a willingness to adopt AI driver technology, particularly relating to issues of trust.

Further limitations to the present study tilt towards sample size. While the study sample was sizable, it only comprised a convenient sample of London drivers aged 18 to 89. However, as it is not a population sample, a higher sample size could lead to improved machine-learning model performance. Secondly, there is strong evidence that driving exposure affects driving performance significantly [52]. This should be further examined as it may also result in notable changes in EEG readings and disparities in cognitive demand between participant groups based on driving experiences. The study's equipment's Bluetooth capabilities is one of its strengths. Every measurement taken during the experiment was immediately transmitted over Bluetooth to the experimenter's research phone. Lastly, this study's field experiment was done in an automobile, with participants operating the vehicle to maximize the generalizability of the results.

8. Conclusions

Using EEG, HR, and BP signals, the authors examined the neurophysiological effects of talking on a hands-free mobile phone in real time on drivers. A model was developed to classify the cognitive performance of drivers. In multiple algorithm analyses, EEG theta and lower alpha signals increased during the hard tasks and exceeded those during the easy tasks. The BP and HR under phone condition were higher than the BP and HR under no-phone condition.

To gather subjective data from each participant, the authors used a questionnaire. The answers provided by the drivers to the qualitative questionnaire produced empirical evidence regarding their cognitive performance. For the dataset, the authors researched and selected the most appropriate algorithms. The answers to the survey were used to verify the machine-learning techniques and the effects of driving and talking on hands-free

mobile phones (which vary subjectively among participants) were, therefore, validated. An accuracy level of 97% was achieved by using the feedforward network.

From Table 2 above, participants average blood pressure without phone = 114.52, while average blood pressure with phone is 121.30. Participants' average HR without phone = 71.07. Average HR with phone = 77.43. Further statistical results shows that participants' theta easy average = 2,430,000.00, whereas theta hard average = 2,890,000.00. Average lower alpha easy = 3,080,000.00, whereas average lower alpha hard = 3,540,000.00.

According to these results from the quantitative approach, while the EEG theta and lower alpha bands are higher for the hard task than for the easy task, and while the BP and HR under the phone condition are higher than the BP and HR under the no-phone condition, the subject is cognitively loaded, evoking poor performance. If they are shown to be lower, the subject's performance is deemed to be good. In a similar vein, the results of the qualitative survey form indicate that cognitive workload on participants increased significantly while performing the phone tasks. The results of this study have validated our hypothesis.

The authors have comprehensively evaluated the potential challenge to this study by considering the feasibility of replacing the human driver with an AI driver and monitoring and controlling cognitive distraction due to conversation between the AI driver and a passenger. In summary, autonomy contributes to active passenger safety by preventing collisions. The ability of the system to respond to incoming threats and plan proper responses in those situations is vital to active safety. However, software reliability is the biggest challenge when it comes to autonomous cars' safety. For passenger safety, how vehicles react and recover from unforeseen circumstances, such as sensor failure, data loss, and communication failures, is crucial. However, the public still has major safety concerns regarding the adoption of AI driver technology despite the emerging technology.

The authors acknowledged and cited earlier related research that had findings that were comparable to this research in the discussion and analysis section. Despite the overwhelming drawbacks, talking on hands-free mobile phones has some notable benefits, such as making communication generally more accessible, especially emergency calls. Nonetheless, the overall result from this study generally advises members of the public to limit conversation times, do essential talking whilst driving, and perhaps use voice-activated mobile phones to minimize the risks due to distraction. The generalization of the created model across the United Kingdom will be the focus of our upcoming study.

Appendix A

Questionnaire Form

Thank you for taking part in our research, your feedback will be very much appreciated. Please evaluate your experienced workload and assess the impact of mental demand by filling in this questionnaire based on the scale provided below.

About You:

Age [ ]

Gender [ ]

Driver Group: Novice [ ]   Experienced [ ]   Elderly [ ]

Please tick the box which you think best represents the impact of mental demand from your experience during the experiment. 0 represents no impact while 100 represents a maximum impact.

0 [ ]	41-45 [ ]	[ ] 86-90
1-5 [ ]	46-50 [ ]	[ ] 91-95
6-10 [ ]	51-55 [ ]	[ ] 96-100
11-15 [ ]	56-60 [ ]	
16-20 [ ]	61-65 [ ]	
21-25 [ ]	66-70 [ ]	
26-30 [ ]	71-75 [ ]	
31-35 [ ]	76-80 [ ]	
36-40 [ ]	81-85 [ ]	

References

1. Oviedo-Trespalacios, O.; King, M.; Haque, M.M.; Washington, S. Risk factors of mobile phone use while driving in Queensland: Prevalence, attitudes, crash risk perception, and task-management strategies. Australia Queensland University of Technology (QUT), Centre for Accident Research and Road Safety—Queensland. *PLoS ONE* **2017**, *12*, e0183361. [CrossRef] [PubMed]
2. Spyropoulou, I.; Linardou, M. Modelling the Effect of Mobile Phone Use on Driving Behaviour Considering Different Use Modes. *J. Adv. Transp.* **2019**, *2019*, 2196431. [CrossRef]
3. Desmet, C.; Diependaele, K. An eye-tracking study on the road examining the effects of handsfree phoning on visual attention. *Transp. Res. Part F* **2019**, *60*, 549–559. [CrossRef]
4. Horsman, G.; Conniss, L. Investigating evidence of mobile phone usage by drivers in road traffic accidents. *Digit. Investig.* **2015**, *12*, S30–S37. [CrossRef]
5. Reimer, B.; Mehler, B.; Coughlin, J.F.; Roy, N.; Dusek, J.A. The impact of a naturalistic hands-free cellular phone task on heart rate and simulated driving performance in two age group. *Transp. Res. Part F Traffic Psychol. Behav.* **2011**, *14*, 13–25. [CrossRef]
6. Yoshitsugu, Y. Brainwave Signal Measurement and Data Processing Technique for Daily Life Applications. *Adv. Technol. Res.* **2009**, *28*, 145–150.
7. Welburn, C.; Amin, A.; Stavrinos, D. Effect of electronic device use while driving on cardiovascular reactivity. *Transp. Res. Part F Traffic Psychol. Behav.* **2018**, *54*, 188–195. [CrossRef]
8. Sharif, S.; Tamang, M.; Fu, C. *Predicting the Health Impacts of Commuting Using EEG Signal Based on Intelligent Approach*; School of Architecture, Computing and Engineering, UEL: London, UK, 2022.
9. Sharif, S.; Tamang, M.; Fu, C. *Evaluating the Stressful Commutes Using Physiological Signals and Machine Learning Techniques*; School of Architecture, Computing and Engineering, UEL: London, UK, 2022.
10. Reñosa, C.; Bandala, A.; Vicerra, R. *Classification of Confusion Level Using EEG Data and Artificial Neural Networks*; De La Salle University Manila: Metro Manila, Philippines, 2019.
11. Sweller, J.; Ayres, P.; Kalyuga, S. *Cognitive Load Theory*; School of Education, University of New South Wales Sydney: Sydney, NSW, Australia, 2011; p. 2052.
12. Tjolleng, A.; Jung, K.; Hong, W.; Lee, W.; Lee, B.; You, H.; Son, J.; Park, S. Classification of a Driver's cognitive workload levels using artificial neural network on ECG signals. *Appl. Ergon.* **2017**, *59*, 326–332. [CrossRef] [PubMed]
13. Sullman, M.J.; Przepiorka, A.M.; Prat, F.; Blachnio, A.P. The role of beliefs in the use of hands-free and handheld mobile phones while driving. *J. Transp. Health* **2018**, *9*, 187–194. [CrossRef]
14. Carsten, O. *Distracted Driving: An Overview Oliver*; Institute for Transport Studies University of Leeds: Leeds, UK, 2020.
15. Pagano, T.P.; Santos, V.R.; Bonfim, Y.D.S.; Paranhos, J.V.D.; Ortega, L.L.; Sa, P.H.M.; Nascimento, L.F.S.; Winkler, I.; Nascimento, E.G.S. Machine Learning Models and Videos of Facial Regions for Estimating Heart Rate: A Review on Patents, Datasets, and Literature. *Electronics* **2022**, *11*, 1473. [CrossRef]
16. Chen, S.; Ji, Z.; Wu, H.; Xu, Y. A Non-Invasive Continuous Blood Pressure Estimation Approach Based on Machine Learning. *Sensors* **2019**, *19*, 2585. [CrossRef]
17. Moshawrab, M.; Adda, M.; Bouzouane, A.; Ibrahim, H.; Raad, A. Cardiovascular Events Prediction using Artificial Intelligence Models and Heart Rate Variability. *Procedia Comput. Sci.* **2022**, *203*, 231–238. [CrossRef]
18. Seo, J.; Laine, T.; Sohn, K. An Exploration of Machine Learning Methods for Robust Boredom Classification Using EEG and GSR Data. *Sensors* **2019**, *19*, 4561. [CrossRef]
19. Vishwanath, M.; Jafarlou, S.; Shin, I.; Lim, M.M.; Dutt, N.; Rahmani, A.M.; Cao, H. Investigation of Machine Learning Approaches for Traumatic Brain Injury Classification via EEG Assessment in Mice. *Sensors* **2020**, *20*, 2027. [CrossRef] [PubMed]
20. Rahman, T.; Ghosh, A.K.; Shuvo, M.H.; Rahman, M.M. Mental Stress Recognition using K-Nearest Neighbor (KNN) Classifier on EEG Signals. In Proceedings of the International Conference on Materials, Electronics & Information Engineering, ICMEIE-2015, Rajshahi, Bangladesh, 5–6 June 2015.
21. Duan, L.; Ge, H.; Ma, W.; Miao, J. EEG feature selection method based on decision tree. *Bio-Med. Mater. Eng.* **2015**, *26*, S1019–S1025. [CrossRef] [PubMed]
22. Islam, M.R.; Barua, S.; Ahmed, M.U.; Begum, S.; Di Flumeri, G. Deep Learning for Automatic EEG Feature Extraction: An Application in Drivers' Mental Workload Classification. In *Human Mental Workload: Models and Applications, Proceedings of the Third International Symposium, H-WORKLOAD 2019, Rome, Italy, 14–15 November 2019*; Springer International Publishing: Berlin/Heidelberg, Germany, 2019.
23. Ganglberger, W.; Gritsch, G.; Hartmann, M.; Fürbass, F.; Perko, H.; Skupch, A.; Kluge, T. A Comparison of Rule-Based and Machine Learning Methods for Classification of Spikes in EEG. *J. Commun.* **2017**, *12*, 589–595. [CrossRef]
24. Guo, L.; Rivero, D.; Dorado, J.; Rabuñal, J.; Pazos, A. Automatic epileptic seizure detection in EEGs based online length feature and artificial neural networks. *J. Neurosci. Methods* **2010**, *191*, 101–109. [CrossRef]

25. Reimer, B.; Mehler, B.; Coughlin, J.F.; Wang, Y.; D'Ambrosio, L.A.; Roy, N.; Dusek, J.A. A Comparison of the Effect of a Low to Moderately Demanding Cognitive Task on Simulated Driving Performance and Heart Rate in Middle Aged and Young Adult Drivers International Conference on Cyberworlds. In Proceedings of the 2008 International Conference on Cyberworlds, Hanzhou, China, 22–24 September 2008; IEEE: New York, NY, USA, 2008.

26. Scheepers, D.; Ellemers, N. When the pressure is up: The assessment of social identity threat in low and high-status groups. *J. Exp. Soc. Psychol.* **2005**, *41*, 192–200. [CrossRef]

27. Di Flumeri, G.; Borghini, G.; Aricò, P.; Sciaraffa, N.; Lanzi, P.; Pozzi, S.; Vignali, V.; Lantieri, C.; Bichicchi, A.; Babiloni, F.; et al. EEG—Based Mental Workload Neurometric to Evaluate the Impact of Different Traffic and Road Conditions in Real Driving Settings. *Front. Hum. Neurosci.* **2018**, *12*, 509. [CrossRef]

28. Mehler, B.; Reimer, B.; Coughlin, J.F.; Dusek, J.A. Impact of Incremental Increases in Cognitive Workload on Physiological Arousal and Performance in Young Adult Drivers. *Transp. Res. Rec.* **2009**, *2138*, 6–12. [CrossRef]

29. Zokaei, M.; Jafari, M.J.; Khosrowabadi, R.; Nahvi, A.; Khodakarim, S.; Pouyakian, M. Tracing the physiological response and behavioural performance of drivers at different levels of mental workload using driving simulators. *J. Saf. Res.* **2020**, *72*, 213–223. [CrossRef]

30. Puma, S.; Matton, N.; Paubel, P.-V.; Raufaste, É.; El-Yagoubi, R. Using theta and alpha band power to assess cognitive workload in multitasking environments. *Int. J. Psychophysiol.* **2018**, *123*, 111–120. [CrossRef]

31. Gifford, A.T.; Dwivedi, K.; Roig, G.; Cichy, R.M. A large and rich EEG dataset for modeling human visual object recognition. *NeuroImage* **2022**, *264*, 119754. [CrossRef] [PubMed]

32. Tornros, E.B.; Bolling, A. Mobile phone use—Effects of handheld and handsfree phones on driving performance. *Accid. Anal. Prev.* **2005**, *37*, 902–909. [CrossRef] [PubMed]

33. EEG Band Power Values: Units, Amplitudes, and Meaning. (n.d.). Available online: http://support.neurosky.com/kb/development-2/eeg-band-power-values-units-amplitudes-and-meaning (accessed on 18 August 2019).

34. Hjortskov, N.; Rissén, D.; Blangsted, A.K.; Fallentin, N.; Lundberg, U.; Søgaard, K. The effect of mental stress on heart rate variability and blood pressure during computer work. *Eur. J. Appl. Physiol.* **2004**, *92*, 84–89. [CrossRef]

35. Sulaiman, N.; Ying, B.S.; Mustafa, M.; Jadin, M.S. Offline LabVIEW-based EEG Signals Analysis for Human Stress Monitoring. In Proceedings of the 2018 9th IEEE Control and System Graduate Research Colloquium (ICSGRC), Shah Alam, Malaysia, 3–4 August 2018; IEEE: New York, NY, USA, 2018.

36. Chun, W. *Core Python Programming*, 1st ed.; Prentice Hall PTR: Hoboken, NJ, USA, 2000.

37. Siddique, A.B.; Khan, M.M.R.; Arif, R.B.; Ashrafi, Z. Study and Observation of the Variations of Accuracies for Handwritten Digits Recognition with Various Hidden Layers and Epochs using Neural Network Algorithm. In Proceedings of the 2018 4th International Conference on Electrical Engineering and Information & Communication Technology (iCEEiCT), Dhaka, Bangladesh, 13–15 September 2018; IEEE: New York, NY, USA, 2018.

38. Neill, H.; Johnson, T. *Mathematics: A Complete Introduction*; John Murray: London, UK, 2018.

39. Dawson, H.; Dubrule, O.; John, C. Impact of dataset size and convolutional neural network architecture on transfer learning for carbonate rock classification. *Comput. Geosci.* **2022**, *171*, 105284. [CrossRef]

40. Alwosheel, A.; Cranenburgh, S.; Chorus, C. Is your dataset big enough? Sample size requirements when using artificial neural networks for discrete choice analysis. *J. Choice Model.* **2018**, *28*, 167–182. [CrossRef]

41. Althnian, A.; AlSaeed, D.; Al-Baity, H.; Samha, A.; Bin Dris, A.; Alzakari, N.; Elwafa, A.A.; Kurdi, H. Impact of Dataset Size on Classification Performance: An Empirical Evaluation in the Medical Domain. *Appl. Sci.* **2021**, *11*, 796. [CrossRef]

42. Hanczar, B.; Hua, J.; Sima, C.; Weinstein, J.; Bittner, M.; Dougherty, E.R. Small-sample precision of ROC-related estimates. *Bioinformatics* **2010**, *26*, 822–830. [CrossRef]

43. Backer-Grøndahl, A.; Sagberg, F. Driving and telephoning: Relative accident risk when using hand-held and hands-free mobile phones. Institute of Transport Economics. *Saf. Sci.* **2011**, *49*, 324–330. [CrossRef]

44. Bunce, D.; Sisa, L. Age differences in perceived workload across a short vigil. *Ergonomics* **2002**, *45*, 949–960. [CrossRef] [PubMed]

45. Caird, J.K.; Simmons, S.M.; Wiley, K.; Johnston, K.A.; Horrey, W.J. Does Talking on a Cell Phone, with a Passenger, or Dialing Affect Driving Performance? An Updated Systematic Review and Meta-Analysis of Experimental Studies. *Hum. Factors J. Hum. Factors Ergon. Soc.* **2017**, *60*, 101–133. [CrossRef]

46. Nakayama, M.; Sun, Q.; Xia, J. Estimation of Older Driver's Cognitive Performance and Workload using Features of Eye movement and Pupil Response on Test Routes. In Proceedings of the 26th International Conference Information Visualisation (IV), Vienna, Austria, 19–22 July 2022; IEEE: New York, NY, USA, 2022.

47. Darlington, R.; Hayes, A. *Regression Analysis and Linear Models: Concepts, Applications and Implementation*; Guilford Publications, Inc.: New York, NY, USA, 2017.

48. Elbanhawi, M.; Simic, M.; Jazar, R. In the Passenger Seat: Investigating Ride Comfort Measures in Autonomous Cars. *IEEE Intell. Transp. Syst. Mag.* **2015**, *7*, 4–17. [CrossRef]

49. Ozguner, U.; Stiller, C.; Redmill, K. Systems for Safety and Autonomous Behavior in Cars: The DARPA Grand Challenge Experience. *Proc. IEEE* **2007**, *95*, 397–412. [CrossRef]

50. Kumpfmüller, H. The impact of electronic components on the reliability of cars. *Qual. Reliab. Eng. Int.* **1993**, *9*, 251–255. [CrossRef]

51. Large, D.R.; Clark, L.; Burnett, G.; Harrington, K.; Luton, J.; Thomas, P.; Bennett, P. It's Small Talk, Jim, But Not as We Know It. Engendering Trust through Human-Agent Conversation in an Autonomous, Self-Driving Car. In Proceedings of the 1st International Conference on Conversational User Interfaces, Dublin, Ireland, 22–23 August 2019; Human Factors Research Group University of Nottingham: Nottingham, UK, 2019.

52. Crundall, D. Hazard prediction discriminates between novice and experienced drivers. *J. Accid. Anal. Prev.* **2016**, *86*, 47–58. [CrossRef]