
Taylor & Francis
Taylor & Francis Group

International Journal of Human–Computer Interaction

ISSN: 1044-7318 (Print) 1532-7590 (Online) Journal homepage: www.tandfonline.com/journals/hihc20

# Assessing Driver Cognitive Load from Handsfree Mobile Phone Use: Innovative Analysis Approach Based on Heart Rate, Blood Pressure and Machine Learning

**Mhd Saeed Sharif , Boniface Ndubuisi Ossai , Jijomon Chettuthara Moncy , Fahad Alblehai & Cynthia H.Y. Fu**

INTERNATIONAL JOURNAL OF HUMAN–COMPUTER INTERACTION
https://doi.org/10.1080/10447318.2025.2492804



# Assessing Driver Cognitive Load from Handsfree Mobile Phone Use: Innovative Analysis Approach Based on Heart Rate, Blood Pressure and Machine Learning

Mhd Saeed Sharif[a], Boniface Ndubuisi Ossai[a], Jijomon Chettuthara Moncy[b], Fahad Alblehai[c], and Cynthia H.Y. Fu[b,d]

[a]Intelligent Technologies Research Group, Department of Computer Science and Digital Technology, Ace, Uel, London, UK; [b]Department of Psychological Science, Uel, London, UK; [c]Computer Science Department, Community College, King Saud University, Riyadh, Saudi Arabia; [d]Centre for Affective Disorders, Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, UK

**ABSTRACT**

Although using a handheld mobile phone while driving is illegal, hands-free (HF) use remains permitted, despite causing cognitive distraction. This study investigated the cognitive impact of HF phone use on drivers using real-time physiological data—heart rate (HR) and blood pressure (BP)—and applied machine learning to classify driver cognitive load. Participants performed complex tasks while driving and reversing, both with and without HF phone use. Results showed significant increases in HR and BP during HF phone conversations. A novel feedforward neural network model achieved 97% accuracy in classifying cognitive load. The study's real-time, naturalistic approach enhances its generalisability and validity. It uniquely applies advanced ML techniques to highlight the cognitive risks of HF phone use while driving. These findings provide crucial evidence for policymakers, particularly in the UK, supporting efforts to reconsider regulations and improve road safety. The study also offers insights for traffic safety experts and behavioural researchers.

**KEYWORDS**

Machine learning; wearable sensor; driving; cognitive impact; blood pressure; heart rate

## 1. Introduction

Considering their capacity to induce disinterest and cognitive disorientation, mobile phones are prohibited while driving (Lipovac et al., 2017). As a substitute, drivers may use hands-free (HF) mobile phones (Lipovac et al., 2017; Sullman et al., 2018). Nonetheless, should the preceding vehicle slow down during a pivotal moment in the conversation, there is a likelihood of impact. This is because the upcoming driver might fail to react in time. The blood pressure and heart rate of adult drivers who experience this are higher (Mehler et al., 2008; Welburn et al., 2018; Reimer et al., 2008).

An increase in physiological signals is directly correlated with an increase in task difficulty. When three levels of task difficulty were randomly ordered during driving (Son & Park, 2011), found a near linear increase in heart rate. The findings demonstrate that heart rate is sensitive to incremental changes in cognitive workload. Moreover, Son and Park (2011) reported that basic cardiovascular measures (heart rate and blood pressure) increase with increasing cognitive workload. When task demands increased, such as entering a traffic circle, heart rate increased, and decreased as task demands decreased, such as driving on a two-lane highway. Furthermore, other studies have found that when task challenge increases, HR and BP increase, resulting in an increase in workload (Mehler et al., 2008; Welburn et al., 2018). Therefore, HR and BP were chosen as physiological measures in this study to estimate cognitive function.

Considering this, the authors propose that if HR and BP are higher during phone usage compared to during no-phone usage, the participant is loaded cognitively. Cognitive load refers to the level of details the brain can process at once, thus the driver may subsequently be disoriented (Puma et al., 2018; Reñosa et al., 2019; Sugimoto et al., 2020). Conversely, the subject is not loaded cognitively if the HR and BP during phone usage are lower than during no-phone usage. A proven approach to ascertaining the physiological effects resulting from cognitive workload is the use of the OMRON M7 Intelii IT Blood BP/HR Monitoring device. With this device, one can measure BP and HR to determine the effects of talking on an HF mobile phone. It is clinically validated and generates accurate results regard-

**CONTACT** Mhd Saeed Sharif s.sharif@uel.ac.uk Intelligent Technologies Research Group, Department of Computer Science and Digital Technology, Ace, Uel, London, UK

less of where it is placed on the upper arm. In addition, it has Bluetooth capability that transmits all measurements to physicians using mobile phone software (Kollias et al., 2020).

The present study applied two approaches. A qualitative approach was adopted based on research into cognitive load among drivers. It was the qualitative study responses that produced empirical findings about the cognitive function of the drivers, which were utilized to corroborate the ML technique results, thus validating the hypothesis. Experimental evidence indicates that engaging in mobile phones while driving contributes to unsafe driving, which is largely caused by cognitive, rather than physical distractions. Empirical data shows that hands-free phones are also linked to unsafe driving, like hand-held phones (Backer-Grøndahl & Sagberg, 2011). As in Reference (Backer-Grøndahl & Sagberg, 2011), this study employed a qualitative survey. Earlier investigations have primarily depended on average HR data analysis or on average HR and BP from one experiment to determine the study's outcomes (Mehler et al., 2008; Welburn et al., 2018; Reimer et al., 2008; Caird et al., 2018). Using physiological indicators of drivers' workload (HR and BP) and ML techniques, the present research has advanced beyond this level by classifying the cognitive function of drivers and recommending safety measures.

Previous studies have used BP and HR to measure effort in task engagement. In a simulator, Welburn et al. (2018) researched the implications of talking on a mobile phone versus not talking on a phone on BP and HR. Results showed that talking on a mobile phone while driving increased BP and HR considerably, compared to driving without a mobile phone. Talking on a mobile phone as a secondary task from Welburn et al. (2018) is like previous work from Reimer et al. (2008) and Caird et al. (2018). Researchers have investigated the prevalence of mobile phone usage according to type, in terms of placing as well as receiving calls, based on the growing number of people using hands-free mobile phones. According to the results of the investigation, handheld users and hands-free users differed significantly. Among hands-free users, there was a higher percentage of drivers that admitted placing as well as receiving calls (at least once daily), with 43 percent placing and 47 percent receiving calls, compared to 17% and 21% for handheld users. Considering this, more research is needed to comprehend hands-free mobile phone usage (Sullman et al., 2018).

The results following a much more thorough investigation into related research indicate that numerous papers such as Mehler et al. (2008), Reimer et al. (2008), Backer-Grøndahl and Sagberg (2011) and Reimer et al. (2011) have extensively utilized driving simulation trials to investigate driver behavior when using mobile phones. The present study used real-time driving (field research) to reflect participants' behavior as it would occur naturally. Therefore, a practical recommendation can be derived immediately from that situation (Viglia & Dolnicar, 2020). A critical examination of published studies indicates that previous research evaluating the generated impact from talking on mobile phones hands-free has not explored machine learning and has also not explored drivers of all age categories in a single unit (Sullman et al., 2018; Welburn et al., 2018; Backer-Grøndahl & Sagberg, 2011; Hendrick & Switzer, 2007). Considering the points mentioned in this paragraph, it remains unclear how physiological markers such as BP and HR determine the cognitive function of drivers of all ages as one unit when driving in actual time and talking on an HF mobile phone. Filling in this gap is the purpose of this study. By using all driver age groups as a complete cohort, an advanced model was created that categorized the cognitive function of drivers who talked on an HF mobile phone while driving in actual time.

This study is inspired to strive to develop an Artificial neural network (ANN) that will not only predict the consequences of talking on HF mobile phones on driving performance but additionally identify the effects of additional cognitive stressors on drivers of all age groups when driving in actual time while talking on mobile phones hands-free. Biological nervous systems are like artificial neural networks. These artificial networks can be taught to provide results since they are composed of linked networks of neurons in a comparable manner to the human neural system. 3 types of characteristics describe them: first, the interactions among the different neuron layers. Second, the training procedure that adjusts network weights. Lastly, the process that transforms the measured input of a neuron into its output. The ANN algorithm is based on the principle that inputs are fed into hidden nodes, and their accumulated sums are summed up. As a result, the output of the concealed node is subsequently swayed in a particular order (Dastres & Soori, 2021; Zakaria et al., 2024).

Based on this research design, the classifier that produces the highest performance is the focus of the analysis. Therefore, we have limited the discussion of the algorithms' mathematical operands to artificial neural networks only. The accuracy of ANNs improves over time during learning from training data. Using these algorithms, data can be classified at high speeds once they have been fine-tuned for accuracy. Equation (1) below represents the basic mathematical operation of ANN.

$$\mathrm{Y} = \mathrm{W}\mathcal{X} + \mathrm{B}$$

where $\mathcal{X}$ is the input data; Y is the output; W is the weight of the neurons; and B is the bias.

As the algorithm guesses the parameters "W" and "B", it measures the accuracy of the guess, which is sometimes called the loss. This data is used to make another guess. As this is repeated, the loss decreases progressively. Over time, the algorithm learns how to correctly match $\simeq\mathcal{X}\simeq$ to "Y". Depending on the input, the parameters "W" and "B" can be changed or tweaked to achieve the desired result (Dastres & Soori, 2021; Zakaria et al., 2024).

This contribution employed a neural network to classify drivers' cognitive function based on physiological markers of driver workload (BP and HR) in 2 classes (1 or 0) through a multilevel perceptron neural network, such that a "0" stipulates that the driver is loaded cognitively and a "1" specify that the driver is not loaded cognitively. The authors achieved this by measuring the blood pressure and heart rate of drivers while talking on a hands-free mobile phone, collecting and analyzing the data, building a network model using Python, and training, testing, and validating the data.

## 2. Related work

Speaking on a mobile phone hands-free while driving reduces drivers' alertness on the road. While driving, driving tasks and mobile phone use compete for drivers' attention (Caird et al., 2018). During driving, the task may be disrupted. Consequently, the motorist behind may not be capable of reacting in a timely manner should the preceding vehicle slow down while the conversation is in progress (Reimer et al., 2011). The crash risk ratio did not differ between handheld and hands-free mobile phones, according to Young and Schreiner (2009), indicating that the high accident risk ratio is primarily caused by conversational distraction, rather than physical distractions from holding up the phone or dialing. Mobile phone usage by type, for placing and receiving calls, has increased as more people use hands-free mobile phones. The results of the investigation showed remarkable differences between handheld and hands-free users. Drivers who made and answered calls (once a day or more) were more likely to be hands-free users, with forty-three percent of hands-free users placing calls and forty-seven percent receiving calls, as opposed to seventeen percent and twenty-one percent respectively, for handheld users. The results indicate that further investigation needs to be conducted on mobile phone use while hands-free (Sullman et al., 2018). The effects of talking on an HF mobile phone have been investigated using blood pressure and heart rate measurements (Welburn et al., 2018).

Stuiver et al. (2014) describes a short-term cardiovascular strategy to assess drivers' mental workload using data from a driving simulator study. After a short increase in task demands (forty seconds), heart rate and blood pressure were measured as indicators of mental effort. The driving simulator study involved 15 drivers participating in six 1.5-hour sessions. To increase workload demands, short segments of fog (40 s) were used to induce two traffic density levels (7.5-minute segments). Systolic blood pressure increased as traffic density increased, and blood pressure variability decreased. When driving in fog under low traffic conditions, heart rate variability and blood pressure variability measures decreased, indicating an increased effort. The described short-term measures can be used to indicate cardiovascular reactivity as a function of workload.

In the past three decades, machine learning research and applications have grown rapidly. In connection with accelerated technological developments, sophisticated ML algorithms, as well as the emergence of immense amounts of data, ML has advanced enormously. In this study, we investigate an effective, reliable, and feasible way to measure physiological results of HF mobile phone usage, based on HR and BP signals. Much work has been done on BP and HR evaluation, feature retrieval, and classification. BP and HR data classification continues to be largely reliant on feature retrieval. In recent years, machine learning methods for heart rate estimation and blood pressure prediction have been growing in popularity (Maqsood et al., 2021).

Maqsood et al. (2021) carried out an extensive assessment of characteristic retrieval strategies in BP prediction that utilized photoplethysmography (PPG) indices. The characteristic retrieval strategies have been subsequently split between 3 separate categories to evaluate the relevance of every category. Category A consists of time-based characteristics; Category B offers statistical characteristic retrieval, and Category C offers frequency domain-based characteristics. The evaluation incorporated a few ML algorithms and measured each one's effectiveness based on various viewpoints. The research findings from 2 openly accessible datasets indicate that the features corresponding to category A were better dependable compared with other strategies for BP estimation. The study determined that deep learning models delivered more effective outcomes than all conventional machine learning techniques. Based on the findings of this case study, experts will be able to select the most suitable and effective approach to characteristic extraction and machine learning algorithms.

A machine learning technique named classification tree was used to predict increased blood pressure based on body mass index (BMI), waist circumference (WC), hip circumference (HC), and waist-hip ratio (WHR). Among 400 college students aged 16-63 (56.3% women), 400 data were collected. In the training group, 15 trees were calculated for each sex, using different numbers and combinations of predictors. It was found that BMI, WC, and WHR are the combinations that produce the best prediction for women, with the lowest deviation (87.42) and misclassification (.19). In the training set, the model's specificity was 81.22 percent, and sensitivity was 80.86 percent while in the test sample it was 65.15 percent and 45.65 percent, respectively. With the lowest deviation (57.25) and misclassification (.16), BMI, WC, HC, and WHC were the best predictors for men. In the training set, this model had a sensitivity of 72% and a specificity of 86.25%, whereas in the test set, it had a sensitivity of 58.38% and a specificity of 69.70%. In terms of predictive power, the classification tree analysis outperformed the traditional logistic regression method (Golino et al., 2014).

According to a new study, correlated variables (body mass index, age, exercise, alcohol, smoke level, etc.) were used to predict systolic blood pressure using machine learning techniques, specifically artificial neural networks. Data was split into two parts, eighty percent for training the machine and twenty percent for testing its performance. The prediction system was constructed and validated using back-propagation neural networks and radial basis function networks. A backpropagation neural network is used to predict the absolute difference between the measured and predicted value of systolic blood pressure under 10 mm Hg based on a data set with 498 participants. The probability value for men is 51.9% and for women, it is 52.5%. Based on the same input variables and network status, these values are 51.8% and 49.9% for men and women. As a result of this novel method of predicting systolic blood pressure, young and middle-aged people who don't measure their blood pressure regularly may receive early warnings of problems. In addition, due to daily fluctuations, isolated blood pressure measurements may not be very accurate. Medical staff can use this predictor as another reference value. According to the experimental results, artificial neural networks are suitable for modeling and predicting systolic blood pressure (Wu et al., 2014).

The importance of monitoring blood pressure continuously cannot be overstated; nonetheless, the traditional cuff BP monitoring methods are cumbersome for users. With ML algorithms, a cuff-less, non-invasive, and continuous system for measuring blood pressure was proposed using a photoplethysmography (PPG) signal and demographic features. The feature extraction process was performed on 219 PPG signals. The time, frequency, and time-frequency properties of PPG signals were analyzed. For diastolic blood pressure (DBP) and systolic blood pressure (SBP), each regression model was selected. The Relief feature selection algorithm and Gaussian process regression (GPR) outperform other algorithms in determining DBP and SBP, respectively. The ML model can be implemented in hardware systems to continuously monitor blood pressure and avoid any critical health conditions caused by sudden changes (Chowdhury et al., 2020).

To monitor and predict HR based on the wearable sensor (accelerometer)-generated data, it is essential to analyze data analytics and machine learning. Therefore, this study explored various robust data-driven models, such as linear regression, support vector regression, autoregressive integrated moving average (ARIMA), k-nearest neighbor (KNN) regression, random forest regression, decision tree regression, and long short-term memory recurrent neural network algorithm. The accelerometer's univariant heart rate time-series data from healthy people can be used to make future HR predictions using a recurrent neural network algorithm. Under different durations, the models were evaluated. Based on a very recently collected data set, the results demonstrate that an ARIMA model with linear regression and walk-forward validation is effective for predicting heart rate for all durations and other models for durations longer than one minute. According to the results of this study, accelerometer data analytics can be used to predict future HR more accurately (Oyeleye et al., 2022).

To determine which machine learning technique is most suitable in classifying fatal heart rate signals, a study focused on the most adopted and effective machine learning techniques, including artificial neural networks, support vector machines, extreme learning machines, radial basis function networks, and random forests. By applying the above-mentioned machine learning approaches, fatal heart rate measurements were classified as normal or hypoxic. To evaluate the success of the classifiers, confusion matrix, and performance metrics were employed. Despite all machine learning approaches delivering good results, artificial neural networks yielded the best results with 99.73% sensitivity and 97.94% specificity. According to the study results, artificial neural networks outperform other algorithms (Cömerta & Kocamaz, 2017).

A further literature survey of research in machine learning by this study has unfolded some noteworthy algorithms that are typically utilized for evaluation and prediction. The notable algorithms include Artificial Neural Networks, random forest, Bayesian modeling, K-means clustering, KNN, and SVM (Mahesh, 2020; Singh et al., 2016; Kotsiantis et al., 2006; Chen et al., 2014). For regression and classification analysis, Support Vector Machines (SVM) are very reliable and efficient. These models are supervised learning models. SVMs can also perform non-linear classification by employing the kernel trick, which involves projecting their inputs into high-dimensional feature spaces. Essentially, it draws a line separating classes. The margins are drawn in a way that diminishes the space separating the classes and the margins, thus reducing ambiguity in categorization. This approach is commonly adopted given that it has a high level of accuracy and uses minimal processing power (Mahesh, 2020).

The decision tree (DT) is typically utilized as a nonlinear classifier. The algorithm is also quick and easy to use when it comes to classifying and training large amounts of data. In the form of a tree, a decision tree illustrates choices and their outcomes. Graph nodes indicate choices or events, and graph edges indicate conditions or decision rules. There are nodes and branches in every tree. An attribute is depicted by a node, and a value is depicted by a branch (Mahesh, 2020).

ML algorithms such as K-nearest neighbors (KNN) are often extensively implemented in supervised learning for addressing classification and regression tasks. Apart from functioning on relatively small amounts of data at a reasonable speed, the algorithm is also simple to implement and comprehend. In the KNN approach, an item is categorized according to the collective vote of its closest K-neighbors. Based on this technique, items are classified in relation to their closest K-neighbors, ranking them in the top popular group. As a training dataset, an entire dataset is utilized (Singh et al., 2016). In a study, the mental workload of each subject was classified based on the HR variability (HRV) metric. A K-nearest neighbor method achieved an average classification accuracy of 98.77%. The highest average classification accuracy (80.56%) was achieved using HRV signals from 5 subjects for training and one subject for testing. The results of this study may improve operators' safety and wellbeing by analyzing HRV signals that are indicators of mental workload in various subjects (Shao et al., 2020).

As opposed to SVMs or Decision Trees, Logistic Regression incorporates probability, and it can be adjusted online to incorporate new data easily (via gradient descent). As it returns probabilities, classification thresholds can be easily adjusted. It is possible to substitute the logistic model for discriminant analysis. There is no assumption about the structure of the independent variables, and there is no linear relationship between the predictors and target variables.

Nonlinear effects can be handled by it. Nevertheless, reliable results require a large sample size (Singh et al., 2016).

ANN are one of the most efficient tools for data exploration and evaluation (Kotsiantis et al., 2006). Feedforward neural networks are robust and massive deep learning models. Three layers of neurons are fused together; the input layer, the hidden layer, and the output layer. Using convolutional neural networks (CNNs) (Li et al., 2019), separated heart sound signals into abnormal and normal without ECGs. According to the experimental findings, the developed CNN model has greater classification precision, better classification capability, and an elevated F-score, than the backpropagation neural network blood pressure model. A 99.01% classification precision rate is also achieved by the enhanced CNN.

The present investigation explores heart rate and blood pressure as physiological markers of cognitive load. More task burden increases heart rate and blood pressure readings, making heart rate and blood pressure considered as one of the widely researched cognitive load indices (Reimer et al., 2011; Scheepers and Ellemers, 2005). The heart rate and blood pressure of the subjects in phone mode will be contrasted with the heart rate and blood pressure of the subjects in no-phone mode to measure the subjects' cognitive function. Employing data from heart rate and BP, the present research centered on the classification of drivers' cognitive load using ML.

It is expected that the subjects' HR and BP will be higher for phone conditions than for no-phone condition (Zokaei et al., 2020). Therefore, the following hypothesis was proposed: "When a participant's HR and BP are higher in phone mode than in no-phone mode, the participant is viewed as cognitively loaded, culminating in poor performance. In contrast, the participant is not cognitively loaded.

## 3. Methodology

ANN, SVM, Logistic Regression, Decision tree, KNN, and Random Forest ML techniques were applied to the data from blood pressure and heart rate signals.

### 3.1. Subject selection

Healthy drivers in the age groups of young (17–39), older (40–69), and elderly (70 and over) made up the participants. Sixteen subjects took part. Nonetheless, the five subjects' data were excluded due to technical problems encountered during the testing procedures. A sum of 214 simulated data points was produced as well as employed in this research (Gifford et al., 2022). This total also contains data from eleven subjects: 6 females and five males. The subjects' ages varied from 18 to 89 years, with a standard deviation of 16.8 and a mean age of 42.9. The contributors supplied their free and informed consent. Each participant completed two tasks, one simple and one difficult. Participants also completed questionnaires outlining their individualized perspectives on workload. The questionnaire asked about the age, gender, and driving experience level of the driver such as elderly, experienced, or novice.

Each item on the survey form reflects a weighted percentage value, with 0% being the lowest and 100% being the greatest. These boxes represent participants' self-reported cognitive load. Following the experiment, each participant answered the questionnaire by checking the box that, in their opinion, best captures the effect of cognitive load based on their perception throughout the research (where "0" denotes no load and "100" denotes maximum load) (Toroyan et al., 2011; Tornros & Bolling, 2005).

### 3.2. Experiment protocol

Using the Omron blood pressure monitor, the authors took data during control tasks, simple tasks, and difficult tasks. The subjects' BP and HR were recorded with this device. Bay parking in reverse with no phone use is the control task. Before starting the control task, participants' baseline HR and BP were taken. On completion of the control task and within the experimental time frame, new measurements of BP and HR were taken (Welburn et al., 2018; Scheepers and Ellemers, 2005). Reversing into the bay while on the phone is the phone task. The phone task is divided into 2 segments: a simple task (task 1) and a difficult task (task 2). One trial per task for everyone. The following is a simple task procedure: The investigator turns on the phone's power knob. The subject initiates an audio call with the word "Experiment". The phone acknowledges as well as rings that number connected to "Experiment". A previously taped message plays, thus: "Count from 50 up to 200". The subject answers the message while he or she drives to the bay and parks there. In the case of difficult tasks, a similar procedure is used, except using this instruction: "Count backwards from 100, taking away 3 each time".

The hypothesis for this study was formulated by concentrating on the progression of task complexity from simple to difficult tasks throughout the examination, thus allowing a thorough investigation of how drivers' performance changed with an increase in task difficulty. We therefore did not use task randomisation. A standardized procedure was used during which the research protocol was kept constant (Spyropoulou & Linardou, 2019; Desmet & Diependaele, 2019). To ensure reliable results, all aspects of the protocol were kept the same. The authors ensured all subjects had no previous information about the tasks before the testing began to reduce any likelihood of order effects. To avoid anticipation, each experimental session had only one participant. Figures 1 and 2 show the test site entrance and test site car park, while Figures 3 and 4 show driving in traffic and reverse parking in the test site car park.

### 3.3. Data collection and data description

For this study, data collection entails measuring and analyzing variables of interest, such as blood pressure and heart rate, in a systematic way that allows testing hypotheses and evaluating results. To get HR and BP data, an OMRON M7

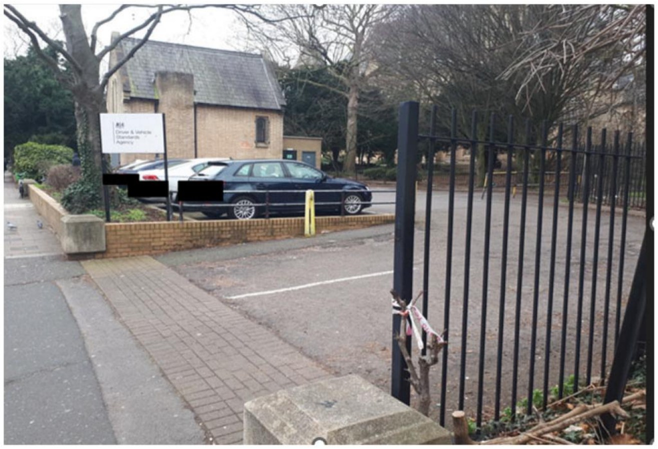

**Figure 1.** Test site entrance.

**Figure 2.** Test site car park.


**Figure 3.** Driving in traffic.

Intelii IT BP Sensor was utilized. The capability of Bluetooth is another advantage. Like mobile phones, it requires pairing for them to connect. "OMRON Connect" is the designation of the application. With the aid of phone holders, two mobile phones were mounted on the dashboard of the research car. One was used to video the experimental route (Samsung Galaxy A12) while the other was used for talking hands-free (Samsung Galaxy A52s 5 G). The Omron blood pressure/heart rate measurer was connected to the experimental research phone via Bluetooth thus: The Bluetooth of the phone was turned on via settings and the Omron blood pressure/heart rate measurer was selected from the list of devices that appeared on the research phone, thereby allowing Bluetooth connection between the measuring device and the research phone. The measuring device was not mounted but rather was utilized when measurement was required. Measurement duration (45 s) is the same for every BP/HR measurement session. The measuring device measures both BP and HR at the same time and shows both BP and HR readings on the screen at the same time, and measurement was taken within the experiment time frame. The user manual states that OMRON digital blood pressure monitors don't need to be calibrated on a regular basis. When the device is powered on, it usually does a calibration self-check. If there is an issue, it will show an error message or other on-screen notification. Before every experiment, its functionality was verified. The cuff was examined for potential damage, such as air leaks, and for general wear and tear that might lead to device malfunctions.

The subject's upper arm was bound with the cuff so that it was in line with the chest. Furthermore, the tubing was positioned over the middle of the subject's front arm. Stretching the cuff's edge ensures that the sensor is firmly fastened and uniformly tight. After pressing the "ON" button on the device, the cuff inflates. The measurements are obtained when the cuff has reached full inflation and stopped inflating and the readings on the screen are constant. Throughout the experiment, every measurement was sent instantly through Bluetooth to the researcher's mobile phone. The monitor measures numerical values that respectively indicate the participants' heart rate and blood pressure. The first column of an Excel spreadsheet's main data file displays the date, whilst the next column shows the measurement time. The subsequent columns show the participant's Systolic blood pressure in mmHg, Diastolic blood pressure in mmHg, and heart rate in bpm in the sequence from left to right.

### 3.4. Feature extraction and data processing

The experiment for this study was designed so that BP and HR measurements were taken during the experimentation window (Welburn et al., 2018; Scheepers & Ellemers, 2005) to guarantee precise and trustworthy readings. Here is a brief explanation of the procedures used to take BP and HR readings: Participants specified bay parking technique was described and then practiced for fifteen minutes (Hettiarachchi et al., 2018). The subjects had a five-minute break (Hettiarachchi et al., 2018). The authors obtained baseline HR and BP values. The subjects drove from the car park entry point and carried out bay parking with no phone use. The subjects' BP and HR were taken again. The subjects took a break for five minutes (Hettiarachchi et al., 2018). The simple task began. The subjects proceeded from the car park entrance and carried out bay parking whilst talking on

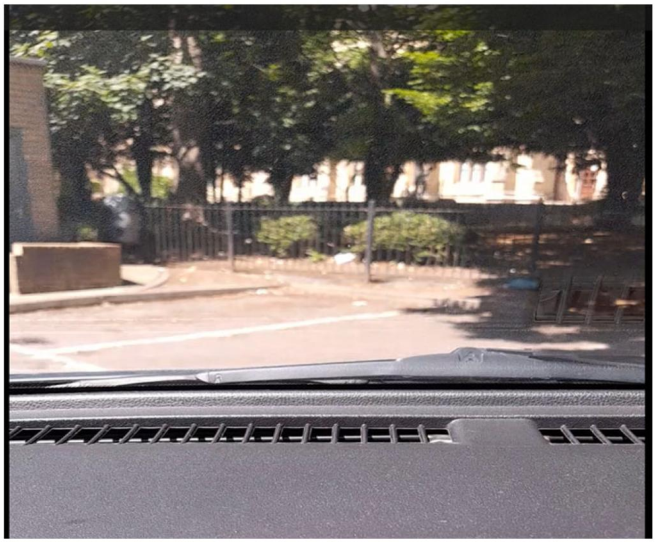

Figure 4. Reverse parking.

the phone. The subjects' HR and BP were measured. After resting for ten minutes (Hettiarachchi et al., 2018), the experimenter repeated the simple task method using the difficult task. The driver's mean HR for simple and difficult tasks (phone mode), mean BP for simple and difficult tasks (phone mode), age, and gender are the input values used to assess the participant's cognitive function and for the ML categorization. A binary class that indicates whether the driver is cognitively loaded is produced by the classifier. Class 1 indicates "not cognitively loaded", while Class 0 indicates "cognitively loaded". HR as well as BP are expected to increase in correlation with self-reported cognitive load. HR and BP have been selected as the sole physiological measures for this study because increased physiological signals such as HR and BP are directly correlated with increased task difficulty (Mehler et al., 2008). According to Son and Park (2011), a near linear increase in heart rate was observed when three levels of task difficulty were randomly ordered during driving. According to these findings, heart rate can distinguish incremental changes in cognitive workload with high sensitivity.

### 3.5. Measurement process flow

Figure 5 illustrates the project's key components in a block diagram. In the first block, BP and HR are recorded using a non-invasive sensor as detailed in Section "C" above. Data processing includes finding the mean of the driver's HR for a simple task and difficult task (phone mode) and finding the mean of the driver's BP for a simple task and difficult task (phone mode). Extrapolated values comprise the driver's mean HR for a simple task and difficult task (phone mode), the driver's mean BP for a simple task and difficult task (phone mode), and the driver's gender and age. Blocks 3–5 illustrate the steps followed in the ML procedure in which modeling, as well as classification tasks, have been performed, which resulted in ANN reaching the optimum level of performance out of the 6 algorithms applied in this research. Section "F" below illustrates the step-by-step sequence of the research process.

### 3.6. Step-by-step Sequence of Investigation

1. Researcher meets the subject at the testing site's car park.
2. Researcher briefs the subject at testing site's car park (procedure).
3. Check subject's driving license, issue date & another form of ID.
4. Document age, gender, and driver category (elderly, experienced or novice).
5. Document driving licence number. Subject signs declaration.
6. Check subject's eyesight (read a car registration number twenty meters away).
7. The subject's driving ability is tested in the experimental car.
8. Failure? Yes, the subject is withdrawn but continues otherwise.
9. Do a 15-minute illustration and practice session of the specific bay parking technique.
10. Subject rest for 5 min.
11. Measure baseline BP & HR using Omron M7 Intelli IT BP/HR monitor for 45 s.
12. The subject drives from the site entrance and parks in a bay without using a phone.
13. Measure subject's BP & HR.
14. 5 min rest.
15. Subject drives towards the site entrance.
16. Researcher switches the phone power button on. The phone type is Samsung Galaxy A52s 5G.
17. Subject says "experiment".
18. Subject listens and confirms voice recognition.
19. The phone says, "Calling experiment".
20. The phone rings and switches to the pre-recorded message.
21. Message to subject: "Count from 50 up to 200".
22. Subject begins the task as he or she drives from the entrance towards the bays.
23. Subject finishes bay parking.
24. Subject's BP & HR are recorded.
25. Subject rest for 10 min.
26. Subject drives towards the cite entrance.
27. Researcher switches the phone power button on.
28. Subject says "Jump".
29. Subject listens and confirm voice recognition.
30. The phone says, "Calling Jump".
31. The phone rings and switches on to a pre-recorded message.
32. Message to subject: "Count backwards from 100 taking away 3 each time".
33. Subject begins the task as he or she drives from the entrance towards the bays.
34. Subject finishes the bay parking.
35. Subject's BP & HR are recorded.

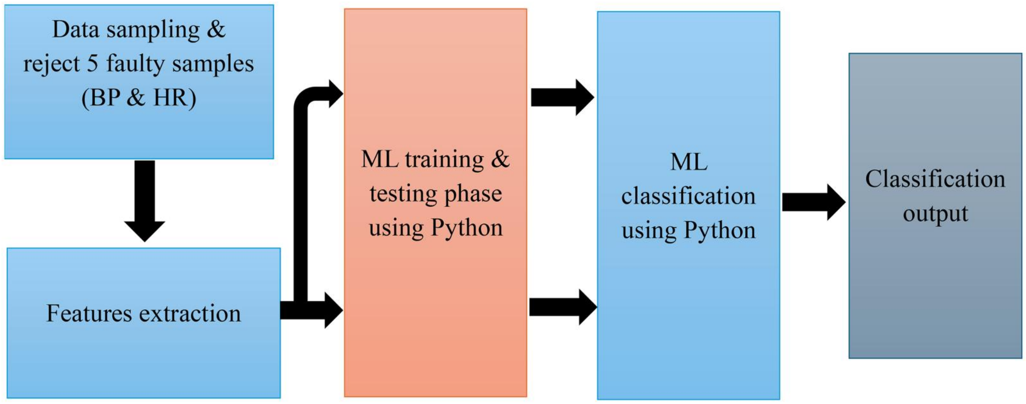


Figure 5. Research plan block diagram.

36. Subject fills out a survey form regarding perceived workload.
37. Subject is debriefed.
38. Researcher elucidates the study's significance to science & public safety.
39. Researcher answers questions from subjects.
40. Subject is given a voucher for participating before departing.
41. Data processing.
42. 5 faulty samples were rejected.
43. Find the average of the drivers' HR for simple tasks and difficult tasks (phone mode).
44. Find the average of the drivers' BP for simple tasks and difficult tasks (phone mode).
45. Extract drivers' average HR for simple tasks and difficult tasks (phone mode).
46. Extract drivers' average BP for simple tasks and difficult tasks (phone mode).
47. Extract drivers' gender and age.
48. Organize dataset in eleven columns and 214 rows (214 data points) using Excell spreadsheet.
49. Import data from the csv file "finaldataset.csv" using Python's read () function.
50. Read data into a dataframe using Pandas read_csv ().
51. Encode categorised variables and separate training from testing data.
52. Print data information to ensure there are no null values.
53. Train and test the ML models.
54. Classify the cognitive function of drivers using ML classification algorithms.
55. Visualise the classified outputs.

## 4. Modelling and implementation

SVM, KNN, Logistic Regression, Decision Trees, Random Forests, and ANN were selected as part of the recommended procedure. At first, the data were analyzed for every ML algorithm to train, test, and validate the hypothesis. Eighty percent of the data consisted of training data, while twenty percent consisted of testing data. Data were cleaned, processed, and scaled to ensure consistency across the dataset. Data was tested to ensure true generalization ability. However, the algorithm that gave the highest accuracy is the focus in terms of analysis and illustration in the present study's design. ANN gave the highest accuracy. Therefore, the training and testing details for this algorithm have been described below in this section.

The training algorithm for ANN in the present study is 'backpropagation'. Artificial Neural Networks (ANNs) are trained using the supervised learning technique known as backpropagation, which iteratively modifies the network's weights in response to the discrepancy between the intended output and the actual goal. The network may learn from its failures and gradually improve its predictions by altering the connections between neurons and efficiently propagating this error information backward down the network layers to adjust the weights and minimize the overall loss function (Al-Sammarraie et al., 2018).

The learning rate, which regulates how much the weights and biases are changed in each iteration during backpropagation, the weights connecting neurons between layers, and the biases associated with each neuron (which indicate the strength of connections) are the main parameters in backpropagation. Here's how backpropagation makes use of these parameters: The network processes input data, calculating each neuron's output according to its weights and biases. The "error" is the difference between the actual target value and the expected output. The network propagates the fault backward. To reduce the overall error and raise the prediction accuracy of the model, the method computes the gradient of the error function regarding these weights and biases (Al-Sammarraie et al., 2018).

Two datasets were created utilizing related parameters that would affect the individuals' performance to validate the theory. The first dataset includes the driver's mean heart rate for simple and difficult tasks (phone mode), as well as the driver's mean blood pressure for simple and difficult tasks. The second dataset includes information collected from participants regarding their age and gender. As targets, binary data points were used. 0 denotes "cognitively loaded" and 1 denotes "not cognitively loaded".

Six classifiers were used to compare classification accuracy. A model's accuracy, and how precisely the predictions align with the data, is an important factor in its performance (Sharif et al., 2022). It is therefore crucial to examine the

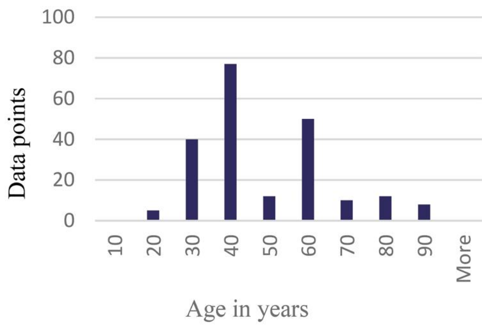


**Figure 6.** Participant's age distribution.

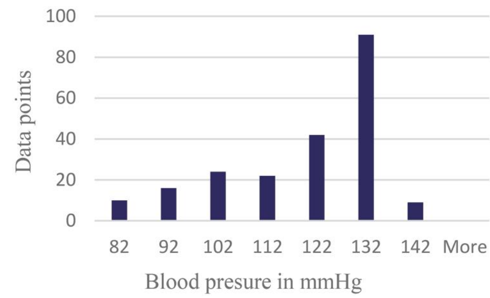


**Figure 7.** Data points versus BP without the phone.

precision of each model. To determine the model's accuracy, the authors collected a small portion of the dataset for validation. The six techniques were all implemented in the Phyton programming language. Research has used this language to analyze data, develop algorithms, and model ML (Chun, 2000). A description of the computing procedure is given in the following paragraphs.

Some library tools for data processing, ML, and data display were imported using Python scripts. Additionally, the script can load and pre-process or prepare data, train, assess classifiers, as well as display model results. The employed dataset consists of two hundred and fourteen data points organized in eleven columns using an Excel spreadsheet. Physiological data such as HR and BP, as well as individual data such as gender and age are included in the columns. Inputs to the classifier include heart rate, blood pressure, gender, and age. The dataset contains both categorical and numerical data, and there are no missing values.

Using codes, the computation process commenced with loading and reading relevant data. Python's read () function was used to import and read the relevant CSV file. Data was imported from the csv file "finaldataset.csv" as well as read into a dataframe using Pandas read_csv (), then processed and cleaned, which included dealing with invalid or missing values, encoding categorized variables, and separating training from testing data. With respect to the overall number of rows, the dataset has two hundred and fourteen observations, whereas the total number of columns suggests 11 variables. Data information was printed to make sure that there were no null values. The cognitive function of the drivers was then classified using ML classification algorithms. Reliability and precision evaluations were conducted on the machine-learning models. With a 97% accuracy rate, the ANN model produced the best results.

Weights and biases for all layers of the neural network are trainable parameters in the ANN. The number of parameters in each layer is calculated by multiplying the number of inputs by the number of neurons plus the number of biases. The size of the output may differ based on the amount of input batches used for training. To normalize the data-frame, we applied normalization techniques and divided the dataset into training and testing. 650 is the epoch while the batch size is = 8. There was an increase in model accuracy over time as training, as well as validation epoch numbers, increased. According to the findings, the ANN model can make predictions based on well-fitting data. This research was designed to optimize the accuracy of all the classifiers during the classification process as can be seen in the epoch count for the ANN's model accuracy graph in Figure 14 below. In addition, analyses focus on the classifier which gives the best performance. Hence, the authors have scoped or limited the mathematical expression of the algorithms to ANN only.

## 5. Result and analysis

This study's 214 simulated data points depict the participants to be analyzed. Histograms were used in the results and analysis, as shown below. Histograms are graphs that display the distribution of continuous data. They show how frequently values fall into different groups. The height of each bar indicates how many objects in the dataset fit into a particular category. The values on the x-axis from Figures 7 and 8 show blood pressure groups, while Figures 9 and 10 represent heart rate groups. The height of each bar indicates the proportion of data points (subjects) that fit in each blood pressure bracket and each heart rate bracket respectively. For instance, in Figure 7, ten data points fit in the BP bracket (0-82) while 16 data points fit in the BP bracket (83-92) and so on. Similarly, the values on the x-axis in Figure 11 represent the group values of participants' percentage self-reported cognitive load.

As shown in Figure 6 above, 40-69 years are the biggest group (65%), whereas 70 years and above and 17-39 are roughly 16% and 19% respectively. From statistical measures without a phone: Mean BP = 114.52, Max BP = 142, Min BP = 80 and SD = 15.35. The statistics of BP with the phone are: Mean BP = 121.30, Max = 151, Min = 89 and SD = 13.19. Results for heart rate without phone: Mean HR = 71.07, Max HR = 110, Min = 55 and SD = 12.64. For HR with phone: Mean HR = 77.43, Max HR = 115, Min HR = 55 and SD = 13.20. From the results, the mean blood pressure and mean heart rate during phone use are greater than the mean blood pressure and mean heart rate during no-phone use. Thus, subjects' mean BP with phone tasks is 6.78 mmHg more than the mean BP without a phone. Similarly, participants' mean HR with phone tasks is 6.36 beats per minute more than the mean HR without a phone. This is due to the cognitive requirement as a consequence of the extra cognitive burdens (simple task and difficult task).

Since the reader is more likely to understand if the BP increment with and without a phone is emphasized, as

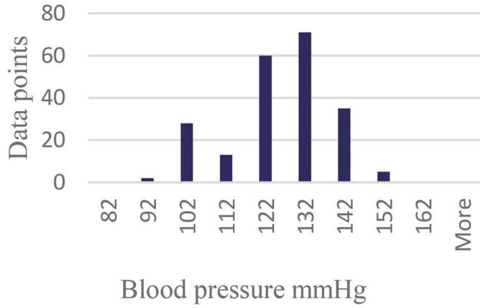


**Figure 8.** Data points versus BP with the phone.

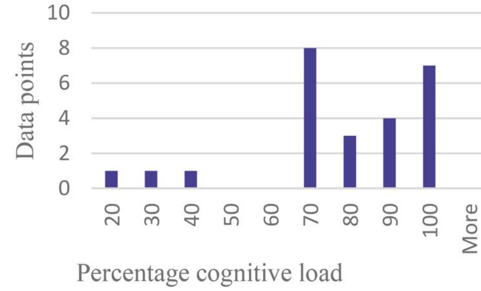


**Figure 11.** Participants self-reported cognitive load.

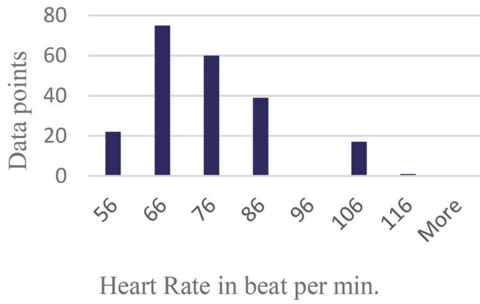


**Figure 9.** Data points versus HR without the phone.

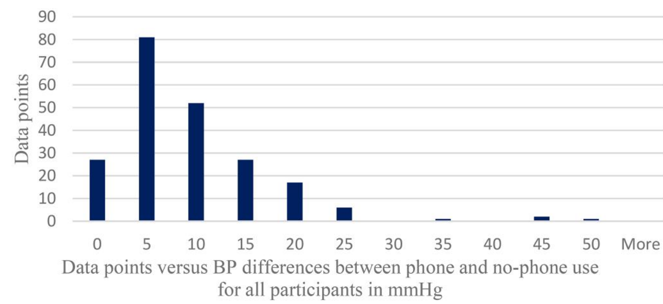


**Figure 12.** Data points versus BP differences between phone and no-phone use for all participants.

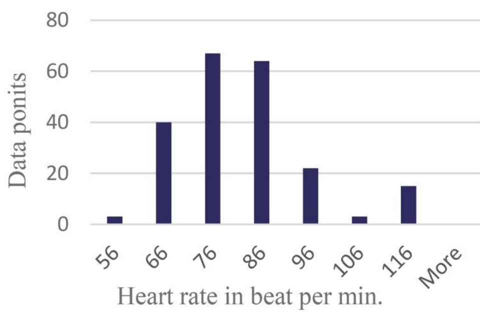


**Figure 10.** Data points versus HR with the phone.

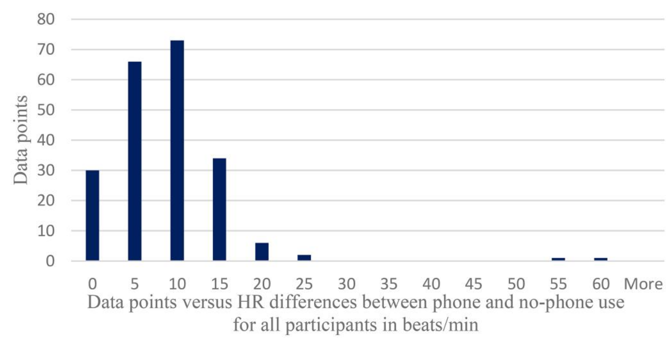


**Figure 13.** Data points versus HR differences between phone and no-phone use for all participants.

normal blood pressure (blood pressure when not driving) varies from subject to subject and some may be hypertensive, the illustration focuses on BP increments from no-phone to phone usage and HR increments from no-phone to phone usage. A graph showing data points versus blood pressure differences between phone and no-phone use and a graph showing data points versus heart rate differences between phone and no-phone use have been plotted as shown in Figures 12 and 13 above. These graphs have been used to illustrate the relationships between the participants and BP differences and the relationship between the participants and HR differences. The figures show that blood pressure and heart rate differences between phone and non-phone use are distributed across most of the subjects.

The proportion of subjective cognitive load in percentage on account of the extra cognitive burdens (simple task and difficult task) is depicted in Figure 11 above by randomly selecting approximately 25 data samples from the total data samples. 22 participants (data points) reported having a 70% or higher response rate. Participants reporting less than 70% in total = 3. The hypothesis for this study has been corroborated by the results as illustrated above and the self-reported cognitive load by the participants. From the results, this study's contribution has demonstrated that based on drivers of all age categories as a single unit in a real driving setting, the average BP and HR of the drivers increased whilst talking on mobile phones hands-free and exceeded those under no-phone conditions. Table 1 above presents the accuracy results comparing each of the classifiers employed in this research.

**Table 1.** Accuracy results comparison.

| Model | Accuracy (%) |
|---|---|
| ANN | 97 |
| SVM | 80 |
| Logistic regression | 82 |
| K-Nearest Neighbor | 71 |
| Decision tree | 72 |
| Random forest | 78 |

Figure 14 shows the model precision graph for ANN. The authors ran the codes and achieved 97% training accuracy as best while the validation accuracy was 91% as shown on the graph. Numerous studies in the literature have

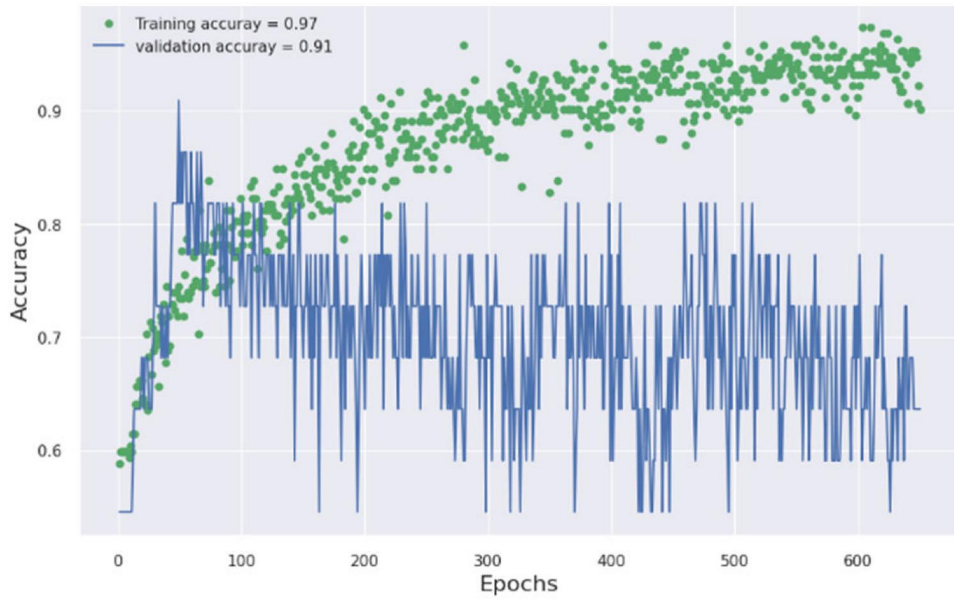


**Figure 14.** Model precision versus epoch for ANN.

reported that ANN yields the best outcomes, even though several other ML techniques give satisfactory results (Wu et al., 2014; Cömerta & Kocamaz, 2017). However, overfitting is a common problem with neural networks (Jabbar & Khan, 2015). Therefore, the performance analysis of the ANN model for this study leans toward the training and validation accuracies due to this significant flaw. To determine whether an ANN model is overfitting or underfitting, it is essential to use both training and validation accuracies when assessing its performance. This allows for better model optimization by pinpointing areas that require improvement, especially when it comes to adjusting model complexity and fine-tuning hyperparameters to improve generalization on unseen data. Overfitting occurs when the training accuracy is much greater than the validation accuracy, indicating that the model has learned the data (including noise) too well and may not perform well on new data. The model may not be sophisticated enough to recognize the patterns in the data, indicating underfitting, if the accuracy of both training and validation is insufficient (Jabbar & Khan, 2015).

A trained neural network's performance is frequently assessed using validation data, which is also used to choose the network that is thought to be best suited for the given task (Foody, 2017). The authors have tracked changes in training and validation accuracy with various hyperparameter settings (e.g., number of hidden layers and learning rate). The ideal parameters that maximize generalization and reduce overfitting were found. To make sure the model isn't assessed on data it has previously seen during training, the authors divided the data into separate training and testing sets. Since validation accuracy is based on data that hasn't been used to train the model, it is a better predictor of how well the model will perform when it is exposed to fresh data than training accuracy. For a model to function well, the validation accuracy must be equal to or marginally lower than the training accuracy (Foody, 2017). This claim is supported by the current study's 91% validation accuracy as shown on the model precision graph in Figure 14.

## 6. Discussion

Driving while on a mobile phone has been banned owing to the inattention and cognitive dysfunction it causes. HF mobile phones are permitted during driving as an alternative. In adult drivers 18-66 years of age, however, HF mobile phone use during driving increases HR and BP (Mehler et al., 2008; Welburn et al., 2018; Reimer et al., 2008). Even so, the neurophysiological impact of additional cognitive demands (dual task) in young novice drivers (18-19 years old) and older drivers (65+ years old) is unclear. It is unclear how neurophysiological markers such as HR and BP measure cognitive functions during driving and talking on an HF mobile phone. It is also uncertain if previous studies have employed real-time driving (field research) to reflect participants' behavior as it would occur naturally.

With advanced ML strategies for drivers' biological markers, such as HR and BP, this paper explored the cognitive implications of talking HF on mobile phones in actual time. To prove the hypothesis, quantitative as well as qualitative procedures were applied. The qualitative method relied on a survey, grounded in research on drivers' cognitive load. The participants' qualitative responses dispensed the empirical proof about their cognitive function that was employed to confirm the findings of the ML approaches; therefore, the hypothesis was proved.

Only a few studies have investigated the physiological implications of talking HF on a mobile phone during driving (Mehler et al., 2008; Welburn et al., 2018). In the present investigation, talking on an HF mobile phone while driving resulted in a substantial increase in BP and HR, much higher than when driving with no phone as was reported in related papers (Mehler et al., 2008; Welburn et al., 2018; Reimer et al., 2008). Findings show that participants' mean BP with phone tasks is 6.78 mmHg more than the mean BP without a phone. Similarly, participants' mean HR with phone tasks is 6.36 beats per minute more than the mean HR without a phone. This is due to the cognitive requirement that arose from the secondary cognitive demand (task due to phone use) as supported by the qualitative results. Feedforward neural networks were used to classify participants' cognitive performance, and 97% accuracy was achieved. The papers (Puma et al., 2018; Reñosa et al., 2019) support the present study's findings confirming that as task difficulty increases, HR and BP increase, causing an increase in workload.

The findings provide supplementary corroboration that talking on an HF mobile phone causes cognitive distraction. This is rather concerning because several studies such as Lipovac et al. (2017) and Welburn et al. (2018) have provided similar evidence. It can be argued that the increasing number of traffic events on urban roads is adding to the cognitive load on drivers as they are talking on HF mobile phones (Di Flumeri et al., 2018). The additional mental impact that an increase in traffic events may have on drivers is beyond the scope of this research,

With respect to self-reported cognitive load, by randomly selecting approximately 25 data samples from the total data samples. The subjective cognitive load across the subjects, as a consequence of the extra cognitive burdens (simple and difficult tasks, respectively), indicate that twenty-two participants reported having a 70% or higher reaction rate. Participants reporting less than 70% in total = 3. It can be argued that the sample used in this investigation is too low to draw conclusions. Nevertheless, the total findings from the 214 data points in this investigation (quantitative), correlate with the findings based on the 25 data samples collected.

The phone task employed in this study is a representation of talking on an HF mobile phone as maintained by the study design. It can be argued that people on phone calls always talk about what they are familiar with, which may not increase the cognitive load significantly. However, like the present research, the study by Yan et al. (2018) also employed numerical tasks with differing levels of intricacies as secondary tasks while talking on an HF mobile phone while driving which instigated various levels of workload in the distracted drivers, and a significant result was achieved. Like Yan et al. (2018), the present study provides additional insight. The study by Haque and Washington (2014) also demonstrated that subjects in a phone conversation had talked about unfamiliar content. In Haque and Washington (2014), the individuals had to listen to a full question, do a basic math problem, or solve a verbal challenge before they could respond appropriately to the phone conversation. "If three wine bottles cost ninety-three dollars, what is the cost of one wine bottle?" was an example of a numerical question used in Haque and Washington (2014). These types of queries involve simultaneous storage and processing of information and therefore distract drivers by boosting their cognitive burdens. Moreso, talking during a conversation on the phone has been extensively studied (Caird et al., 2018; Yan et al., 2018; Haque & Washington, 2014; Carsten, 2020).

The classification algorithms used in this study include ANN, SVM, Logistic Regression, KNN, Random Forest, and Decision Tree. As shown in previous literature, these algorithms are proven to be simple to use when identifying the category of new observations based on training data (Golino et al., 2014; Wu et al., 2014; Cömerta & Kocamaz, 2017). For instance, Golino et al. (2014), highlighted the strength of the classification Tree while, Wu et al. (2014) had applied ANN and asserted that ANN are suitable for model prediction. Cömerta and Kocamaz (2017) have applied ANN, support vector machines, and random forests. According to Cömerta and Kocamaz (2017), despite all machine learning techniques producing satisfactory results, ANN produced the best results with 99.73% sensitivity and 97.94% specificity. For large datasets, KNN is slow, while SVM needs a long training time (Mahesh, 2020; Singh et al., 2016). A major disadvantage of decision trees is that they are prone to overfitting the training data, while neural networks have a greater computational burden, are prone to overfitting, and are empirical in nature (Golino et al., 2014). Despite the drawbacks, related studies show remarkable strength and benefits from the use of these classifiers for data analysis and prediction, which led the authors to select these classifiers for the present study.

There are alternative algorithms to the algorithms used in this study, including linear regression, Naïve Bayes, XGBoost, and LSTM. The linearity of a linear regression model is one of its primary advantages: Besides being relatively simple to apply and performing well with linear data, it also has the drawbacks of being prone to underfitting and assuming the data is independent (Nakayama et al., 2022). Naïve Bayes is a linear classifier that is faster when applied to big data, but it is sensitive to data quality, which is one of its main drawbacks. It can produce inaccurate or biased results if the data is noisy, incomplete, or imbalanced (Ting et al., 2011). As with decision trees, XGBoost combines multiple decision trees to make predictions, but it has the disadvantage of requiring significant computational resources, especially when using large datasets or many iterations (Liew et al., 2021). Long-short-term memory (LSTM) requires more memory to train and is easy to overfit (Oyeleye et al., 2022). However, machine learning has some potential limitations when it comes to analyzing physiological data. The quality of the data provided to ML determines how smart and effective it can be. For accurate modeling, substantial data is often required (Golino et al., 2014; Wu et al., 2014; Cömerta & Kocamaz, 2017). By using machine learning algorithms, bias and discrimination may be maintained due to overfitting and underfitting of models. The use of ML may also reduce the critical thinking and judgment of the researcher or analyst if it is overused in the analysis.

This study has some notable limitations and strengths. As a first point, despite the study's large sample size, it was limited to a convenience sample of drivers in London ranging in age from 18 to 89. For this study, we used a sample dataset from a small population rather than a population dataset. It is also noted by the authors that some ML studies have used limited or small sample sizes. For instance Rahman et al. (2015), collected EEG data from 10 subjects in its experiment, from which only data from five subjects were analyzed, data from the other five subjects were ignored because of the higher degree of noise and artifacts. Ganglberger et al. (2017), selected 12 participants for all recordings in its experiment while Tjolleng et al. (2017) recruited 15 male participants only in a study. The present study has overcome limitations due to the small sample size through 214 simulated data from 11 subjects (the 11 subjects inclusive). Several significant studies such as Xanthis et al. (2020) and Shepperd and Kadoda (2001), have also employed data simulation to address shortcomings due to a small sample size. Xanthis et al. (2020) has overcome limitations due to small sample sizes through advanced simulations on a realistic computer model of human anatomy without using a real MRI scanner and without scanning patients. Shepperd and Kadoda (2001). used simulation to generate a large validation dataset for comparing software prediction techniques.

Strengths of the study include equipment's Bluetooth capability. Throughout the experiment, every measurement was sent instantly through Bluetooth to the researcher's mobile phone. Secondly, to optimize the generalisability of the findings from the field experiment for this study, this research has conducted experiments in a car whilst the participants were driving in real-time. The experimental testing site is Wood Green Driving Test Center, London. During the study, all the subjects drove the same car, which was authorized by the research ethics committee of the University of East London, specifically for the investigation. Identical research methods were followed for the entire subjects but driving situations for each driver were anticipated to vary due to external influences such as weather and the effects of other vehicles. Nevertheless, the simple and hard tasks were not introduced to the subjects earlier than the testing site entrance. One car at a time can be parked in the bay. Hence extraneous variables and other vehicles have a relatively minimal effect.

The present study examined the cognitive ramifications of talking on an HF mobile phone during driving. The distracted driving induced by commencing or terminating a call, trying to find a number to call, or mistakes such as the phone accidentally dropping, are various effects of mobile phone usage during driving that deserve further examination. With these modes of mobile phone usage, the risk of interference with driving may well increase further. A detailed study of this issue is also necessary. It is believed, however, that the use of hands-free phones has some notable advantages because there are drivers who depend on them for work, for example, delivery drivers, who need to find out about their next job, or taxi drivers who require accurate navigation applications, not to mention paramedics and police who must constantly be on the radio.

## 7. Conclusions

The physiological impact on drivers due to talking HF on mobile phones was examined in real-time, using BP as well as HR signals. To predict the effect on the participants, a model was created. HR and BP of participants increased during the phone condition and exceeded those during the no-phone condition. A survey was conducted to gather subjective data from each subject. In reliance on the responses that participants submitted to the qualitative survey, empirical proof was obtained pertaining to their cognitive function. After examination, the most suitable algorithms for the dataset were chosen. By employing the questionnaire responses, ML methods were verified. Therefore, the consequences of driving and talking on mobile phones hands-free (which differ individually among subjects) were validated. The Feedforward network reached ninety-seven percent accuracy.

Based on statistical measures for BP without phone: Mean BP = 114.52, Max BP = 142, Min BP = 80, and SD = 15.35. The statistics of BP with the phone are: Mean BP = 121.30, Max = 151, Min = 89, and SD = 13.19. Heart rate without phone: Mean HR = 71.07, Max HR = 110, Min = 55, and SD = 12.64. For HR with phone: Mean HR = 77.43, Max HR = 115, Min HR = 55, and SD = 13.20. The findings from statistical measures indicate as follows: Mean BP along with mean HR during phone mode are greater than mean BP as well as mean HR during no-phone mode. These findings from the quantitative study illustrate that while the HR and BP in the phone mode are greater than those during no-phone mode, the participant is loaded cognitively, causing poor task execution. A participant's performance is regarded as good if the values are lower. Similarly, the qualitative questionnaire form shows that participants' cognitive load elevated considerably when they performed the telephone tasks. The outcomes of this investigation validate the hypothesis.

As part of the discussion and analysis, earlier related research with similar findings was acknowledged and cited by the authors. The authors discussed the benefits and drawbacks of the classifiers applied in this research, and why these classifiers were selected for this study. The potential of using ML for classifying physiological data and how data simulation was used to overcome limitations due to small sample sizes by several studies was discussed. The authors have used both quantitative and qualitative techniques to address the study's research question and have achieved significant results as detailed above and have also improved the research by using machine learning approaches. By creating public understanding pertaining to the shortcomings of talking on mobile phones HF, this research contributes to the United Kingdom's Department of Transport and Public Safety. Consequently, the government can consider the existing findings regarding the precariousness of talking on a mobile phone HF while driving and will be able to measure and revise their road safety advancement strategies.

Despite their numerous drawbacks, hands-free mobile phones can also offer some noteworthy advantages, such as making correspondence more convenient, particularly in emergencies. To reduce the risks inherent with distracted driving, this study generally encourages drivers to minimize discussion duration, indulge in only necessary conversations while driving, and possibly explore the use of voice command mobile phones. Our forthcoming research will concentrate on how the developed approach might be applied generally throughout the United Kingdom.

## Acknowledgment

The authors gratefully acknowledge the support and funding from the school of ACE, and the Researchers Supporting Project number (RSPD2025R564), King Saud University, Riyadh, Saudi Arabia.

## Author contributions

CRediT: **Mhd Saeed Sharif**: Conceptualization, Funding acquisition, Project administration, Resources, Supervision, Visualization, Writing – original draft, Writing – review & editing; **Boniface Ndubuisi Ossai**: Data curation, Formal analysis, Investigation, Methodology, Project administration, Writing – original draft, Writing – review & editing; **Jijomon Chettuthara Moncy**: Formal analysis, Investigation, Methodology, Validation, Writing – original draft; **Fahad Alblehai**: Resources, Software, Validation, Visualization, Writing – original draft, Writing – review & editing; **Cynthia H.Y. Fu**: Conceptualization, Data curation, Resources, Validation, Writing – original draft, Writing – review & editing.

## Disclosure statement

The authors report there is no conflicting interest to declare.

## Data availability statement

The data used to support the findings of this study are currently under restriction while the research findings are commercialised. Requests for data after publication of this article will be considered by the authors.

## References

Al-Sammarraie, N., Al-Mayali, Y., & El-Ebiary, Y. (2018). *Classification and diagnosis using back propagation Artificial Neural Networks (ANN)* [Paper presentation]. International Conference on Smart Computing and Electronic Enterprise (ICSCEE). IEEE. https://doi.org/10.1109/ICSCEE.2018.8538383

Backer-Grøndahl, A., & Sagberg, F. (2011). Driving and telephoning: Relative accident risk when using hand-held and hands-free mobile phones. *Safety Science*, *49*(2), 324–330. https://doi.org/10.1016/j.ssci.2010.09.009

Caird, J. K., Simmons, S. M., Wiley, K., Johnston, K. A., & Horrey, W. J. (2018). Does talking on a cell phone, with a passenger, or dialing affect driving performance? An updated systematic review and meta-analysis of experimental studies. *Human Factors*, *60*(1), 101–133. https://doi.org/10.1177/0018720817748145

Carsten, O. (2020). *Distracted Driving: An Overview Oliver*. Institute for Transport Studies University of Leeds.

Chen, W., Wang, Y., Cao, G., Chen, G., & Gu, Q. (2014). *A random forest model-based classification scheme for neonatal amplitude - integrated EEG* [Paper presentation]. IEEE International Conference on Bioinformatics and Biomedicine.

Chowdhury, M. H., Shuzan, N. I., Chowdhury, M. E. H., Mahbub, Z. B., Uddin, M. M., Khandakar, A., & Reaz, M. B. I. (2020). Estimating blood pressure from the photoplethysmogram signal and demographic features using machine learning techniques. *Sensors (Basel, Switzerland)*, *20*(11), 3127. https://doi.org/10.3390/s20113127

Chun, W. (2000). *Core Python Programming* (First Edition); Prentice Hall PTR.

Cömerta, Z., & Kocamaz, A. F. (2017). *Comparison of Machine Learning Techniques for Fetal Heart Rate Classification* [Paper presentation]. 3rd International Conference on Computational and Experimental Science and Engineering. https://doi.org/10.12693/APhysPolA.132.451

Dastres, R., & Soori, M. (2021). Artificial neural network systems. *International Journal of Imaging and Robotics (IJIR)*, *21*(2), 13–25.

Desmet, C., & Diependaele, K. (2019). An eye-tracking study on the road examining the effects of handsfree phoning on visual attention. *Transportation Research Part F: Traffic Psychology and Behaviour*, *60*, 549–559. https://doi.org/10.1016/j.trf.2018.11.013

Di Flumeri, G., Borghini, G., Aricò, P., Sciaraffa, N., Lanzi, P., Pozzi, S., Vignali, V., Lantieri, C., Bichicchi, A., Simone, A., & Babiloni, F.,. (2018). EEG—Based mental workload neurometric to evaluate the impact of different traffic and road conditions in real driving settings. *Frontiers in Human Neuroscience*, *12*, Article 509. https://doi.org/10.3389/fnhum.2018.00509

Foody, G. (2017). Impacts of sample design for validation data on the accuracy of feedforward neural network classification. *Applied Sciences*, *7*(9), 888. https://doi.org/10.3390/app7090888

Ganglberger, W., Gritsch, G., Hartmann, M., Fürbass, F., Perko, H., Skupch, A., & Kluge, T. A. (2017). Comparison of Rule-Based and Machine Learning Methods for Classification of Spikes in EEG. *Journal of Communications*, *12*(10), 589–595. https://doi.org/10.12720/jcm.12.10.589-595

Gifford, A. T., Dwivedi, K., Roig, G., & Cichy, R. M. (2022). A large and rich EEG dataset for modeling human visual object recognition. *NeuroImage*, *264*, 119754. https://doi.org/10.1016/j.neuroimage.2022.119754

Golino, H. F., Amaral, L. S. D., Duarte, S. F. P., Gomes, C. M. A., Soares, T. D. J., Reis, L. A. D., & Santos, J. (2014). Predicting increased blood pressure using machine learning. *Journal of Obesity*, *2014*, 637612–637635. https://doi.org/10.1155/2014/637635

Haque, M., & Washington, S. (2014). A parametric duration model of the reaction times of drivers distracted by mobile phone conversations. *Accident; Analysis and Prevention*, *62*, 42–53. https://doi.org/10.1016/j.aap.2013.09.010

Hendrick, J., & Switzer, J. (2007). Hands-free versus Hand-held cell phone conversations on a breaking response by young drivers. *Perceptual and Motor Skills*, *105*(2), 514–522. https://doi.org/10.2466/pms.105.2.514-522

Hettiarachchi, I., Hanoun, S., Nahavandi, D., Iskander, J., Hossny, M., & Nahavandi, S. (2018). *Towards More Accessible Physiological Data for Assessment of Cognitive Load - A validation Study* [Paper presentation]. International Conference on Systems, Man, and Cybernetics. IEEE. https://doi.org/10.1109/SMC.2018.00517

Jabbar, H. K., Khan, R. Z. (2015). Methods to avoid over-fitting and under-fitting in supervised machine learning (comparative study). Computer Science, Communication & Instrumentation Devices. ISBN:978-981-09-5247-1.

Kollias, A., Papadatos, S. S., Dominiczak, A. F., Parati, G., & Stergiou, G. S. (2020). Automated office blood pressure measurements in primary care are misleading in more than one third of treated hypertensives: The VALENTINE-Greece Home Blood Pressure Monitoring study. *Hellenic Journal of Cardiology: HJC = Hellenike Kardiologike Epitheorese*, *61*(3), 174–177. https://doi.org/10.1016/j.hjc.2019.01.005

Kotsiantis, S. B., Zaharakis, I. D., & Pintelas, P. E. (2006). Machine learning: A review of classification and combining techniques. *Artificial Intelligence Review*, *26*(3), 159–190. https://doi.org/10.1007/s1046200790523

Li, F., Liu, M., Zhao, Y., Kong, L., Dong, L., Liu, X., & Hui, M. (2019). Feature extraction and classification of heart sound using 1D convolutional neural networks. *EURASIP Journal on Advances in Signal Processing*, *2019*(1). https://doi.org/10.1186/s13634-019-0651-3

Liew, X., Hameed, N., & Clos, J. (2021). An investigation of XGBoost-based algorithm for breast cancer classification. *Machine Learning with Applications*, *6*, 100154. https://doi.org/10.1016/j.mlwa.2021.100154

Lipovac, K., Đerić, M., Tešić, M., Andrić, Z., & Marić, B. (2017). Mobile phone use while driving-literary review. *Transportation Research Part F: Traffic Psychology and Behaviour*, *47*, 132–142. https://doi.org/10.1016/j.trf.2017.04.015

Mahesh, B. (2020). Machine learning algorithms—A review. *International Journal of Science and Research (IJSR)*, *9*(1), 381–386.), https://doi.org/10.21275/ART20203995

Maqsood, S., Xu, S., Springer, M., & Mohawesh, R. (2021). *A Benchmark Study of Machine Learning for Analysis of Signal Feature Extraction Techniques for Blood Pressure Estimation Using Photoplethysmography (PPG)*. School of Information and Communications Technology, University of Tasmania. https://doi.org/10.1109/ACCESS.2021.3117969

Mehler, B., Reimer, B., Pohlmeyer, A. E., & Coughlin, J. F. (2008). The association between heart rate reactivity and driving performance under dual task demands in late middle-aged drivers. *Advances in Transportation Studies an International Journal*, 53–70.

Nakayama, M., Sun, Q., & Xia, J. (2022). *Estimation of Older Driver's Cognitive Performance and Workload Using Features of Eye Movement and Pupil Response on Test Routes* [Paper presentation]. 26th International Conference Information Visualization (IV), Vienna, Austria.

Oyeleye, M., Chen, T., Titarenko, S., & Antoniou, G. (2022). A predictive analysis of heart rates using machine learning techniques. *International Journal of Environmental Research and Public Health*, *19*(4), 2417. https://doi.org/10.3390/ijerph19042417

Puma, S., Matton, N., Paubel, P. V., Raufaste, É., & El-Yagoubi, R. (2018). Using theta and alpha band power to assess cognitive workload in multitasking environments. *International Journal of Psychophysiology: Official Journal of the International Organization of Psychophysiology*, *123*, 111–120. https://doi.org/10.1016/j.ijpsycho.2017.10.004

Rahman, T., Ghosh, A. K., Shuvo, M. H., & Rahman, M. M. (2015). *Mental Stress Recognition using K-Nearest Neighbor (KNN) Classifier on EEG Signals* [Paper presentation]. Proceedings of the International Conference on Materials, Electronics & Information Engineering, ICMEIE-2015, In, Rajshahi, Bangladesh, 5–6 June.

Reimer, B., Mehler, B., Coughlin, J. F., Roy, N., & Dusek, J. A. (2011). The impact of a naturalistic hands-free cellular phone task on heart rate and simulated driving performance in two age groups. *Transportation Research Part F: Traffic Psychology and Behaviour*, *14*(1), 13–25. https://doi.org/10.1016/j.trf.2010.09.002

Reimer, B., Mehler, B., Coughlin, J. F., Wang, Y., D'Ambrosio, L. A., Roy, N., & Dusek, J. A. A. (2008). *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* [Paper presentation]. International Conference on Cyberworlds. In Proceedings of the 2008 International Conference on Cyberworlds, Hanzhou, China, 22–24; New York, NY, USA: IEEE.

Reñosa, C., Bandala, A., & Vicerra, R. (2019). *Classification of Confusion Level Using EEG Data and Artificial Neural Networks*. De La Salle University Manila.

Scheepers, D., & Ellemers, N. (2005). When the pressure is up: The assessment of social identity threat in low and high-status groups. *Journal of Experimental Social Psychology*, *41*(2), 192–200. https://doi.org/10.1016/j.jesp.2004.06.002

Shao, S., Wang, T., Wang, Y., Su, Y., Song, C., & Yao, C. (2020). Research of HRV as a measure of mental workload in human and dual-arm robot interaction. *Electronics*, *9*(12), 2174. https://doi.org/10.3390/electronics9122174

Sharif, S., Tamang, M., & Fu, C. (2022). *Predicting the Health Impacts of Commuting Using EEG Signal Based on Intelligent Approach*. School of Architecture.

Shepperd, M., & Kadoda, G. (2001). *Comparing software prediction techniques using simulation*. Empirical Software Engineering Research Group, School of Design, Engineering, and Computing.

Singh, A., Thakur, N., & Sharma, A. (2016). A review of supervised machine learning algorithms. *3rd International Conference on Computing for Sustainable Global Development. IEEE*, 1310–1315.

Son, J., & Park, M. (2011). *Estimating Cognitive Load Complexity Using Performance and Physiological Data in a Driving Simulator*. Daegu Gyeongbuk Institute of Science & Technology Human LAB Deagu.

Spyropoulou, I., & Linardou, M. (2019). Modelling the effect of mobile phone use on driving behaviour considering different use modes. *Journal of Advanced Transportation*, *2019*, 1–10. https://doi.org/10.1155/2019/2196431

Stuiver, A., Brookhuis, K. A., Waard, D., & Mulder, B. (2014). Short-term cardiovascular measures for driver support: Increasing sensitivity for detecting changes in mental workload. *International Journal of Psychophysiology: Official Journal of the International Organization of Psychophysiology*, *92*(1), 35–41. https://doi.org/10.1016/j.ijpsycho.2014.01.010

Sugimoto, F., Kimura, M., Takeda, Y., Akamatsu, M., Kitazaki, S., Yajima, K., & Miki, Y. (2020). Effects of one-pedal automobile operation on the driver's emotional state and cognitive workload. *Applied Ergonomics*, *88*, 103179. https://doi.org/10.1016/j.apergo.2020.103179

Sullman, M. J., Przepiorka, A. M., Prat, F., & Blachnio, A. P. (2018). The role of beliefs in the use of hands-free and handheld mobile phones while driving. *Journal of Transport & Health*, *9*, 187–194. https://doi.org/10.1016/j.jth.2018.04.001

Ting, S., Ip, W., & Tsang, A. (2011). Is naïve bayes a good classifier for document classification? *International Journal of Software Engineering and Its Applications*, *5*(3), 37–46.

Tjolleng, A., Jung, K., Hong, W., Lee, W., Lee, B., You, H., Son, J., & Park, S. (2017). Classification of a Driver's cognitive workload levels using artificial neural network on ECG signals. *Applied Ergonomics*, *59*(Pt A), 326–332. https://doi.org/10.1016/j.apergo.2016.09.013

Tornros, E. B., & Bolling, A. (2005). Mobile phone use - Effects of handheld and handsfree phones on driving performance. *Accident Analysis and Prevention*, *37*(5), 902–909. https://doi.org/10.1016/j.aap.2005.04.007

Toroyan, T., Harvey, A., Khayesi, M., Krug, E., Lanvers-Casasola, P., Rodrigues, E., Sethi, D., Shakespeare, T., Sminkey, L., Vesic, J., & Zambon, F. (2011). Mobile phone use: A growing problem of driver distraction. *Health Information and Knowledge*., World Health Organization.

Viglia, G., & Dolnicar, S. (2020). *A review of experiments in tourism and hospitality*. Annals of Tourism Research.

Welburn, S. C., Amin, A., & Stavrinos, D. (2018). Effect of electronic device use while driving on cardiovascular reactivity. *Transportation Research. Part F, Traffic Psychology and Behaviour*, *54*, 188–195. https://doi.org/10.1016/j.trf.2018.01.001

Wu, T., Pang, G., & Kwong, G. (2014). *Predicting Systolic Blood Pressure Using Machine Learning* [Paper presentation]. The 7th International Conference on Information and Automation for Sustainability, Colombo, Sri Lanka, 22–24.

Xanthis, C. G., Filos, D., Haris, K., & Aletras, A. H. (2020). *Simulator-generated training datasets as an alternative to using patient data for machine learning: An example in myocardial segmentation with MRI*. Laboratory of Computing, Medical Informatics and Biomedical-Imaging Technologies, School of Medicine, Faculty of Health Sciences, Aristotle University of Thessaloniki.

Yan, W., Xiang, W., Wong, S. C., Yan, X., Li, Y. C., & Hao, W. (2018). *Effects of Hands-free Cellular Phone Conversational Cognitive Tasks on Driving Stability Based on Driving Simulation Experiment*. Department of Civil Engineering, The University of Hong Kong, Pokfulam Road, Hong Kong.

Young, R., & Schreiner, C. (2009). Real-world personal conversations using a hands-free embedded wireless device while driving: Effect on airbag-deployment crash rates. *Risk Analysis: An Official Publication of the Society for Risk Analysis*, *29*(2), 187–204.). https://doi.org/10.1111/j.1539-6924.2008.01146.x

Zakaria, M., Al-Shebany, M., & Sarhan, S. (2024). Artificial neural network: A brief overview. *International Journal of Engineering Research and Applications*, ISSN: 2248-9622, *4*(2)Version 1), 7–12.

Zokaei, M., Jafari, M. J., Khosrowabadi, R., Nahvi, A., Khodakarim, S., & Pouyakian, M. (2020). Tracing the physiological response and behavioural performance of drivers at different levels of mental workload using driving simulators. *Journal of Safety Research*, *72*, 213–223. https://doi.org/10.1016/j.jsr.2019.12.022

## About the authors

**Mhd Saeed Sharif** leads the Intelligent Technologies Research Group and MSc Computer Science courses. His research focuses on AI, telehealth, medical technology, and big data. Collaborating globally with clinicians and policymakers, he advances healthcare systems. He's widely published, awarded, and active in international conferences, editorial roles, and professional societies.

**Boniface Ndubuisi Ossai** is a doctorate researcher with a wide interest in AI and road safety. He is an active member of the Motor School Association of Great Britain, he has acquired a wealth of information and experience regarding matters around road safety, and he is passionate about research into driver behaviour.

**Jijomon Chettuthara Moncy** is a Research Fellow at the UEL. He earned his PhD (2022) in Electrical Engineering from IIT Palakkad, with coursework at IIT Madras, focusing on EEG-based biometric identification. He holds an M.Tech in signal processing, B.Tech in Electronics and Communication Engineering.

**Fahad Alblehai** is currently an associate professor at the computer science department, community college, King Saud University. He received a bachelor's degree in computer science from King Saud University and a master's and a PhD degree in computing from Malaysia. He acts as the head of the natural sciences department.

**Cynthia H.Y. Fu** is Professor of Affective Neuroscience, UEL, and Professor of Affective Neuroscience and Psychotherapy, Centre for Affective Disorders, Institute of Psychiatry, Psychology and Neuroscience, King's College. Her research includes predictive modelling of treatment outcomes and behavior, integrating clinical, neurobiological, and computational data to enhance precision in mental health interventions.