big data and
cognitive computing

Article

Effects of Neuro-Cognitive Load on Learning Transfer Using a Virtual Reality-Based Driving System

Usman Alhaji Abdurrahman [1,*], Shih-Ching Yeh [2], Yunying Wong [3] and Liang Wei [1]

[1] School of Information Science and Technology, Fudan University, Shanghai 200433, China; weiliang@fudan.edu.cn
[2] Department of Computer Science and Information Engineering, National Central University, Taoyuan City 32001, Taiwan; shihching.yeh@g.ncu.edu.tw
[3] School of Psychology, Fudan University, Shanghai 200433, China; cindy0_0@126.com
* Correspondence: usmanalhaji79@gmail.com or aausman18@fudan.edu.cn

Abstract: Understanding the ways different people perceive and apply acquired knowledge, especially when driving, is an important area of study. This study introduced a novel virtual reality (VR)-based driving system to determine the effects of neuro-cognitive load on learning transfer. In the experiment, easy and difficult routes were introduced to the participants, and the VR system is capable of recording eye-gaze, pupil dilation, heart rate, as well as driving performance data. So, the main purpose here is to apply multimodal data fusion, several machine learning algorithms, and strategic analytic methods to measure neurocognitive load for user classification. A total of ninety-eight (98) university students participated in the experiment, in which forty-nine (49) were male participants and forty-nine (49) were female participants. The results showed that data fusion methods achieved higher accuracy compared to other classification methods. These findings highlight the importance of physiological monitoring to measure mental workload during the process of learning transfer.

Keywords: cognitive load; learning transfer; multimodal fusion; physiological measures; virtual reality; driving simulator

Citation: Abdurrahman, U.A.; Yeh, S.-C.; Wong, Y.; Wei, L. Effects of Neuro-Cognitive Load on Learning Transfer Using a Virtual Reality-Based Driving System. *Big Data Cogn. Comput.* **2021**, *5*, 54. https://doi.org/10.3390/bdcc5040054

Academic Editors: Achim Ebert, Peter Dannenmann and Gerrit van der Veer

Received: 13 August 2021
Accepted: 7 October 2021
Published: 13 October 2021

1. Introduction

Psychologically, learning is the change in behavior resulting from individual experience. A subject is said to have learned when it perceives and changes its behavior [1]. The processes of learning and mainly remembering depend on relative changes in the nervous system. The effects of learning are first retained in the brain, after which a more permanent neural change takes place.

However, the piece of information to be learned can be either simple or complex. Educational research describes that simple information can be learned and understood easily. Likewise, a simple task can be solved within the shortest period. This is not true for complex information and complex tasks. Learning transfer is not likely to happen when subjects are overwhelmed with complex learning materials [2]. According to [3], learning is always hampered whenever the learning task requires too much cognitive workload. This working memory is considered volatile or short-term and limited, whereas it is infinite if it is long-term. So, the idea here is that knowledge should be moved to long-term memory so that when the subject is presented with new material, they can retrieve it from memory [4]. However, if the subject cannot recover it from memory, the working memory becomes overloaded, leading to memory failures [4]. Students find it difficult to recall previous pieces of information, especially complex ones, without prior knowledge [5]. This is because the working memory is insufficient; it can only deal with limited information at a time [6]: only about seven meaningful units of information can be stored in it at any given time [7].

The study of cognitive mechanisms has excellent potential for giving modified content from a mental point of view [8]. Cognitive load (CL) is a multidimensional construct representing the load that performing a particular task imposes on the learner's cognitive system when solving a given problem [9,10]. Cognitive load theory consists of a design of instructional methods that efficiently use people's limited mental processing capacity to apply acquired knowledge and skills to new situations [10]. As mentioned in [10,11], the total cognitive load of an individual is an amalgamation of at least two of the following factors: an intrinsic cognitive load, which is a natural difficulty of the material itself and over which instructors have no control over, an extraneous cognitive load which is produced as a result of the method used in introducing the material, and a germane cognitive load, which is the load related to processes that contribute to the construction and automation of schemas. If these factors combined exceed the participant's working memory capacity, it would result in a cognitive memory overload that hampers their learning. At the same time, when a participant's knowledge surpasses the task's difficulty, time and energy would be wasted in solving the task, and in the end, nothing would be learned. Thus, the task's difficulty should be equal to the subject's proficiency to facilitate effective learning [12].

Cognitive load can be determined by either subjective or objective measures [13]. Subjective measures include self-reported mental effort, perceived difficulty, or stress level, while objective measures include physiological, performance-based, and brain activity measures [13]. Even though general limitations exist in measuring cognitive load, the specific method is currently regarded as predominant in measuring the cognitive load [14,15]. In this study, estimating real-time cognitive workload has been explored using psychophysiological metrics such as eye gaze, pupil dilation, heart rate, and task performance modality.

VR technology is a platform that helps incorporate cognitive and functional approaches to learning [16,17]. The function of VR is to provide immersive, sensorimotor, engaging content, and at the same time simulate an array of real or imagined tasks and environments [16].

Driving is one of the most ubiquitous, cognitively demanding, and dangerous activities of our daily lives [17]. Thus, safe driving requires continuous synchronization of processes such as reaction time, attention, visuospatial skills, planning, and execution function. VR provides rehabilitation and safety assessment of driving-related skills at the true limits of the individual's capabilities [17,18].

We plan to develop a VR-based driving system that would help us analyze the effects of neuro-cognitive load on learning transfer. In this paper, multimodal data fusion methods, several machine learning classification algorithms, and strategic analytic methods were explored for cognitive load measurement.

Related Research

Our study explored measuring real-time cognitive workload using psychophysiological metrics such as eye gaze, pupil dilation, heart rate, and performance measures. Each of these metrics has been deliberated concerning cognitive workload measurement. Pupil dilation has been used in Human–computer Interaction for measuring cognitive load [19]. Pupil dilation is well known for responding quickly to changes to the brightness in the visual field and a subject's cognitive load while performing an assigned task [20]. Eye trackers are used in generating eye-gaze coordinates, which can be used and evaluate pupillary data in real-time. Such metrics can be used in assessing the cognitive load, which can be utilized to assist users when a high cognitive load is detected [19].

Moreover, several types of research of numerous cognitive tasks have shown that increased heart rate is connected with increased cognitive load [21–23]. According to Jorna and Peter GAM [24], there was a significant increase in heart rate when participants were subjected to more challenging multi-task conditions than single-task conditions. Heart rate and blood pressure have been shown to increase with increasing cognitive demand or workload in a range of environments [25–27]; the most comprehensive work has been carried out in aviation [28,29]. Brookhuis et al. explained that an increase in task demand increases heart rate, such as entering a traffic circle, and it dropped when task demands decreased, for example, driving on a two-lane highway. Lenneman et al. [26] reported that a number of cardiovascular parameters were measured during four different conditions: baseline, during single-task driving, while subjects drove and engaged in a mildly challenging working memory task, and while engaged in a more complex version of the same active memory task. The authors discovered that heart rate increased continuously as the conditions became cognitively demanding.

Task driving performance metrics are another common method to measure cognitive load [10]. In a VR-based driving system, driving behavior such as how well a subject used the brake and accelerator during driving, and performance metrics such as how the number of times a subject failed the assigned task and the number of scores obtained during the driving task, are associated with cognitive load [30].

As far as we know, there is no single article in the literature that used the multimodal data fusion method to measure the cognitive load on learning transfer using the virtual reality-based driving system; however, the data fusion method to measure cognitive load has been used in different applications [31–34]. Zhang X et al. [34] used 46 distinct photoplethysmogram features to enhance the cognitive workload's measurement accuracy. Barua S et al. [31] employed the multimodal fusion method by applying machine learning (ML) in detecting and classifying various driver's cognitive states, such as sleepiness, stress, and cognitive load, based on physiological data. Putze F et al. [33] reported a case where a simple majority voting fusion method was applied in combining skin conductance, respiration, EEG, and pulse to categorize cognitive load in visual and cognitive tasks.

2. Materials and Methods

2.1. Hypotheses

This study was designed to determine the effects of neuro-cognitive load on learning transfer from a novel VR-based driving system. Psychophysiological metrics were used to assess responses to different levels of difficulty experienced by the participants and how those responses affected the transfer of learning. There are easy and difficult routes, and thus, we have the following hypotheses:

a. The addition of several turns, intersections, and landmarks on the difficult routes would elicit increased psychophysiological activation, such as increased heart rate, eye gaze, and pupil dilation.
b. Due to an increase in psychophysiological activation, participants would make more mistakes when driving on difficult routes.
c. An increase in cognitive load combined with the more cognitively demanding route difficulty would increase response level.

2.2. Participants

A total of 98 university undergraduates of Fudan University participated in the experiment. Out of this number, 49 were male participants and 49 were female participants. All the participants had no real-life driving experience. There were two categories in the study: easy and difficult routes. There were two sessions with the same difficulty level of each category. Thus, we had 196 data points (98 subjects* 2 sessions). However, some of the data were missing, which was the result of the unexpected movements of a participant's head above the sensors' detection range. This resulted in the removal of 11 subjects' data (7 females and 4 males) at the preprocessing stage. Hence, we were left with 185 data entries. The participants were recruited through an in-house advertisement. All the volunteers were briefed about the nature of the experiment and its protocols. The experiment was conducted after the participants filled out the informed consent form and were approved by Fudan University. All the experiments were conducted under relevant guidelines and regulations.

2.3. System Design

In this work, a virtual reality-based driving system was designed to measure and determine the effects of neuro-cognitive load on learning transfer. The two main components of the system were: a VR-driving system and a data capture system, as depicted in Figure 1.

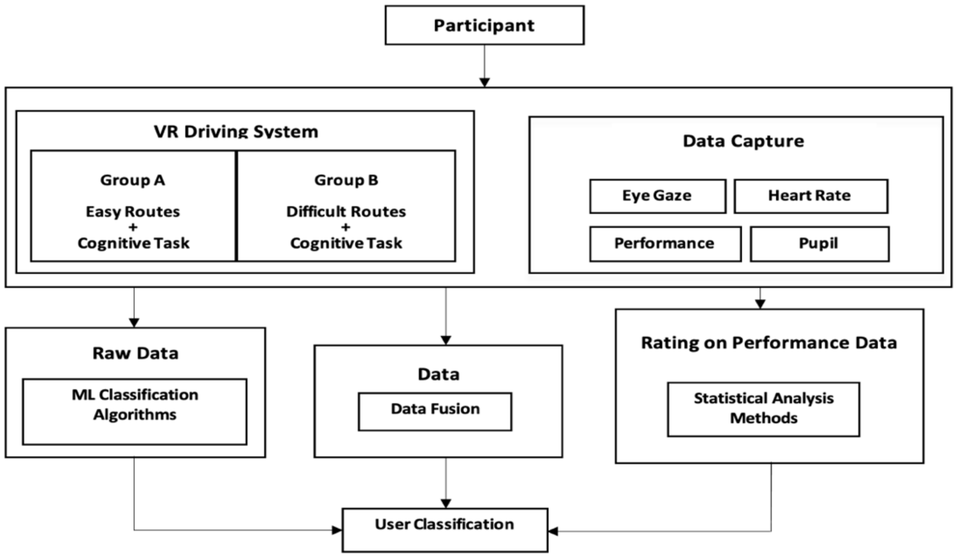

Figure 1. The structure of the VR driving system.

Vive Pro Eye was used for tracking eye gaze data. The VR driving system used in the experiment is given in Figure 2. A Logitech G27 steering-wheel controller was used to control the virtual agent vehicle in the driving environment. Important landmarks in the driving environments such as traffic lights, roadways, intersections, vehicles, and buildings were designed with the Autodesk Maya [35] and ESRI CityEngine [36], while a Unity3D [37] was used for developing the game platform.

Figure 2. The VR driving system.

The simulation schema was developed to reduce the effect of contextual changes on physiological arousal. The driving route consisted of a city roadway that consists of long straightaways, several turns, and intersections, and landmarks (composed of a Basketball Court, McDonald's, Convenience Store, Gas Station, Post Office, Church, and Walmart) to provide minimal stimulation and reduce monotony.

The data capture system keeps track of the different psychophysiological metrics while the participant was engaged in driving. All psychophysiological measures were recorded concurrently throughout the experiment. For tracking eye gaze, a Vive Pro Eye tracker [38] recorded the eye gaze data at 50 Hz. Gaze origin and pupil dilation were among the recorded data. A heartbeat recording device designed in our lab was used to record the user's heart rate at 500 Hz while driving the virtual environment. The participant's task performance was measured in the virtual environment.

The difficulty levels were easy and difficult routes. We had two (4) different levels of difficulty. At each level, there are two different tasks developed for the VR-based driving system.

2.4. Experimental Setup

There were two groups in the experiment: easy (group A) and difficult (group B) routes. There were two sessions (A1, A2 and B1, B2) with the same difficulty level for each group, as shown in Figure 3. The easy routes comprised three (3) turns and intersections, 7 landmarks, and 3 traffic lights, while the difficult route had five (5) turns and intersections, 7 landmarks, and 5 traffic lights. A participant was either assigned to group A (easy route) or group B (difficult route): this is carried out arbitrarily. Additionally, each session lasted approximately 20 min (including a tutorial).

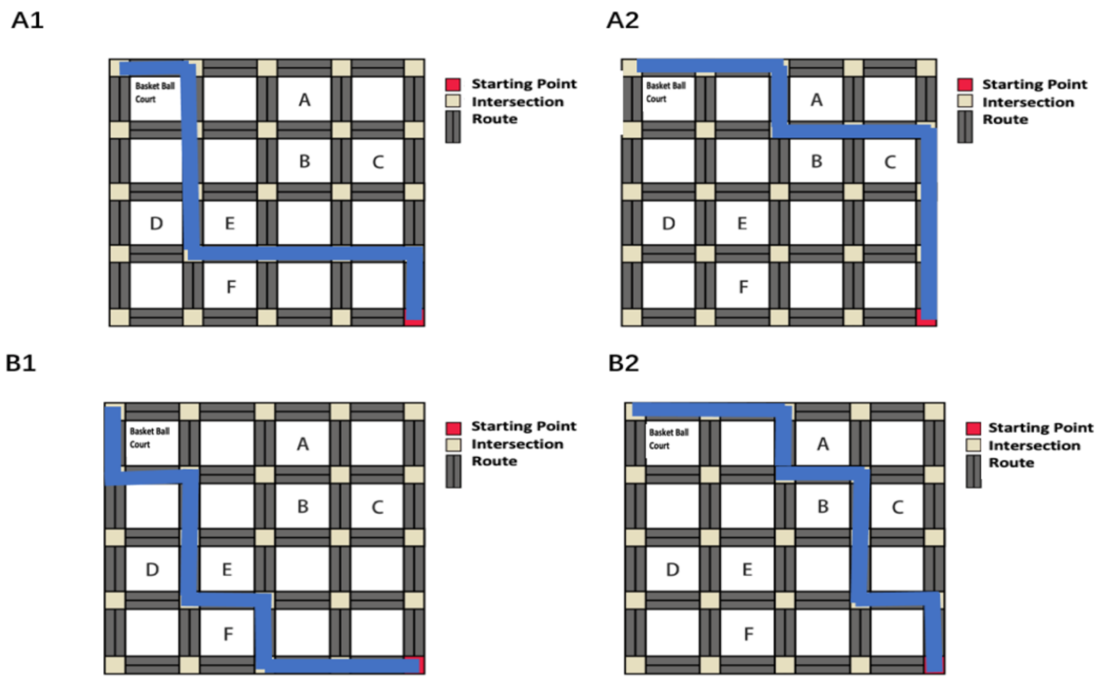

Figure 3. Upper views of easy and difficult routes. (**A1**,**A2**): easy routes with the same difficulty level, (**B1**,**B2**): difficult routes with the same difficulty level.

Before the commencement of the first session of each group, an informed consent form was obtained. Then, the baseline data were collected for the psychophysiological and performance metrics in a quiet environment. A participant was asked to practice driving. This is where they would be asked to move forward, make a reverse, apply a brake, and turn left and right, in accordance with the requirement of the main study. At the same time, a pre-recorded route video tutorial concerning the experiment was then given to the participant before the commencement of the exercise. The tutorial would introduce the participant of each group (A or B) to the routes they would follow, and turns, intersections, landmarks, and traffic lights they would see along the routes that would help them complete the task. This entire video was only played once. The participant was asked to try memorizing the routes in the video. The researchers set up the heartbeat and eye-tracking sensors on the participant's body. Before recording, all the sensors were checked to make sure all signals were in good shape and working perfectly. Afterward, the participant was then asked to carry out two pre-selected driving assignments (one at a time) from the starting point to the destination according to the route they just watched.

Moreover, there were several landmarks that a participant would use to locate the right turn to take to reach the destination. So, whenever a participant makes a wrong turn, they would be dragged to the starting point, re-watch the pre-recorded video guide and start driving all over again. Throughout this exercise, the researchers examined the heartbeat and eye-tracking data to ascertain its quality.

The practice and tutorial sessions were included to assess the system and improve a subject's driving performance. However, we did not consider the performance of these sessions in analyzing the data of this article.

3. Data Acquisition

All physiological features (such as heart rate, eye gaze, pupil dilation) and driving performance data were extracted from each experiment session. Before the start of each session, the baseline data of the sensors were recorded for about 20 min both for the psychophysiological metrics and for the driving performance, in a quiet environment. In total, 11 of the 196 subjects' data were dropped from the analysis at the preprocessing stage because of the motion artifacts that made it impossible to obtain an accurate quantification of the selected analysis.

The eye gaze data were preprocessed, and the noise level was reduced using the median value method [39]. The preprocessed eye-tracking data extracted the pupil dilation, fixation rate, saccade rate, and blink rate. The eye closure duration range used for the blink rate was from 75 to 400 ms, according to [40].

For the eye gaze and pupil dilation data, after preprocessing, we extracted 10 features [30,41]: fixation rate, blink rate, Mean (M), and Standard Deviation (SD) of blink duration, M and SD of fixation duration, M and SD of pupil dilation, and M and SD of saccade duration. Likewise, we extracted M and SD for the heart rate.

Different data were measured for the participant's task performance based on the ability to transfer knowledge and task performance. The ability to transfer knowledge included the number of times a participant made a right turn during one driving assignment. The task performance determined the efficiency of the participant in completing the task; this includes the time (in seconds) a participant completed the driving, the number of collisions on the edge of the road. The driving performance measures and their meanings are shown in Table 1. We extracted 10 features here: Mean (M) and Standard Deviation (SD) of Gametimer, M and SD of WrongCount, M and SD of BarricadeCollider, M and SD of IntersectionCount, and M and SD of RedLightCount.

Table 1. Performance Features and their Descriptions.

Driving Measures	Meaning
Gametimer	Task completion time (seconds).
WrongCount	The total number of wrong turns.
BarricadeCount	The total number of collisions on the edge of the road.
IntersectionCount	The total number of red lights at intersections.
RedLightCount	The total number of red lights running.

A set of analyses were invoked to determine the differences in response to completing the Easy Category and the Difficult Category and differences between Male Category and Female Category related to the VR-based driving system. An Analysis of Variance (ANOVA) was applied to each dependent psychophysiological variable to determine whether the participant was experiencing a challenging working memory. Significant results were reported when $p < 0.05$.

After acquiring all the data for both the psychophysiological and performance metrics, the results of the data analysis are presented in the next section.

4. Results

The primary target of the research is to determine the effects of neuro-cognitive load on learning transfer. This can be achieved by evaluating the differences in psychophysiological response patterns associated with driving on easy routes versus going on difficult routes. As anticipated, the difficult routes elicited an increase in psychophysiological activation (i.e., increase in pupil dilation and heart rate). The results show that the increase in cognitive load was associated with the conditions of the more cognitively demanding routes that led to the rise in response level.

4.1. Analytic Strategy

An Analysis of Variance (ANOVA) was employed to determine the differences in response to completing both the easy and difficult routes. The following sub-sections discussed the application of this analytic strategy on both the psychophysiological and performance measures.

4.1.1. Physiological Measures

An ANOVA method was applied to each dependent psychophysiological variable and a significant effect appears across all the measures: pupil dilation ($F(1, 176) = 321.81$, $p < 0.05$), heart rate ($F(1, 178) = 34.91$, $p < 0.05$). Table 2 shows the average physiological response of the two measures. Pupil dilation increased significantly by 0.51 mm and then by an additional 1.41 mm while driving on easy and difficult routes. Generally, pupil dilation increased by 1.92 mm from the baseline through the difficult routes as depicted in Figure 4.

Table 2. Average Physiological Response Measures.

	Pupil Dilation (mm)	Heart Rate (bpm)
Baseline	3.61 (0.12)	70.6 (10.8)
Easy Routes	4.12 (0.66)	72.6 (7.6)
Difficult Routes	5.53 (0.29)	75.0 (12.7)

Note: The entries are means and the standard deviations in parentheses.

The incremental increase in heart rate was also significant from the baseline to the easy routes and from easy routes to the difficult routes. In these two transitions, heart rate increased by 2.0 bpm and 2.4 bpm, respectively. Thus, heart rate increased by an average of 4.4 bpm from the baseline through the difficult routes (Figure 4).

Additionally, as predicted in the hypotheses, participants that drove the difficult routes committed more mistakes. The main effect of difficulty level was shown with the ANOVA, resulting in higher Gametimer, WrongCount, IntersectionCount, and BarricadeCollider, and a significant interaction was obtained between the aforementioned performance features ($F(3, 540) = 38.14$, $p < 0.05$).

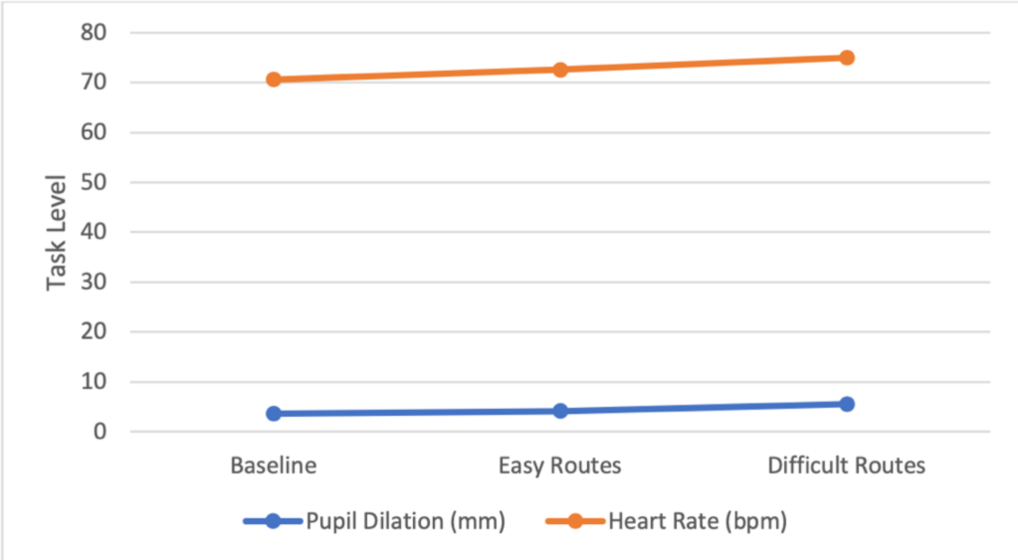

Figure 4. Mean physiological measures in relation to the task level.

4.1.2. Driving Performance

Table 3 gives a summary of the driving performance features. Just like all the physiological measures with considerable changes, the performance measures associated with the task level were also significant. Generally, a statistically significant difference emerges across all the three levels of the task (F(5, 1080) = 38.11, $p < 0.05$), and the interaction between these terms was also significant.

Table 3. Summary of Driving Performance Features.

Driving Performance	Baseline	Easy Route	Difficult Route
Gametimer	138.76 (30.65)	148.98 (32.94)	253.55 (155.23)
WrongCount	0.13 (0.34)	0.18 (0.46)	0.82 (1.50)
BarricadeCollider	0.13 (0.61)	0.19 (0.67)	0.45 (1.14)
IntersectionCount	6.67 (0.98)	7.52 (1.42)	13.12 (7.80)
RedLightCount	3.21 (1.57)	3.95 (1.68)	6.88 (4.20)

Note: The entries are means and the standard deviations in parentheses.

As shown in Table 3, there was a remarkable increase in all the driving performance features across the three levels of the task. The Gametimer increased by a modest 10.22 s and then by a significant 104.57 s from the baseline to easy and easy routes to difficult routes, respectively. WrongCount increased by a modest 0.05 and by a significant 0.64 from the baseline to the easy routes and easy routes to difficult routes. Likewise, incremental increases in BarricadeCollider, IntersectionCount, and RedLightCount were also substantial from the baseline to the easy route and from the easy route to the difficult route. During these periods, BarricadeCollider increased by 0.06 and 0.26, respectively, and IntersectionCount increased by 0.85 and 5.6, respectively, while RedLightCount increased by 0.74 and 2.93, respectively. Across all the performance features, all the increases from the baseline to the easy route were minimal and not statistically significant. The statistically significant results obtained during the easy route and difficult route tasks may indicate the number of efforts made by the participants in the combined demands of driving as well as an increase in psychophysiological activation which led to an increase in cognitive load.

4.1.3. Female Participants versus Male Participants

It is also observed that female participants paid more attention to landmarks (F(3, 720) = 5.64, $p < 0.05$) and drove slower than their male counterparts as depicted in Figure 5. This made them spend much time to complete the exercise (F(1, 180) = 39.79, $p < 0.05$).

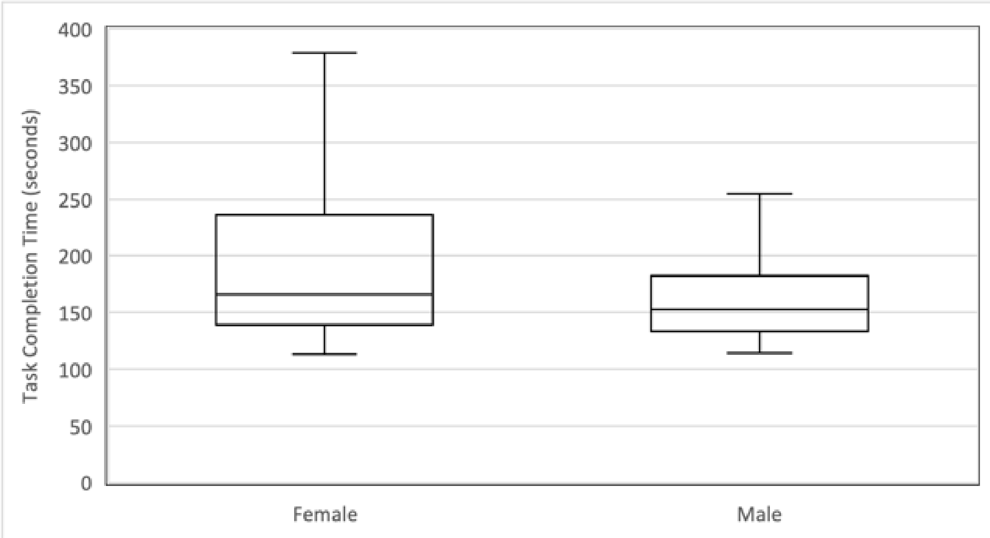

Figure 5. Box plot of task completion time.

4.2. Data Fusion Methods

As part of cognitive load evaluation, different measures from the recorded data were supplied into the classifiers. Five well-known ML algorithms were used in determining the cognitive workload classification. These algorithms include Discriminant Analysis, K-Nearest Neighbors (KNN), Decision Trees, Artificial Neural networks (ANN), and Support Vector Machine (SVM).

Much information can be combined in three approaches [42]: feature-level fusion, decision-level fusion, and hybrid-level fusion. The structures of the three-level fusion approaches are given in Figure 6.

Figure 6a depicts the structure of a feature-level fusion. The features obtained from different modalities composed of pupil dilation, eye gaze, heart rate, and user performance features were input into the "preprocess" module. In this module, all the features would be normalized and have their dimensions reduced using principal component analysis (PCA). The input to the classifier is the preprocessed vector, while the output is the level of CL.

The classifier takes the preprocessed vector as input and outputs the level of cognitive load (CL).

Figure 6b depicts the structure of decision-level fusion. In this case, all the features from each modality mentioned in feature-level fusion were preprocessed independently and then passed to the different classifiers. Each of these classifiers produces a CL as a sub-decision (S).

The final decision (CL_{final}) is calculated from the weighted average of the sub-decision vectors in the data fusion module, as shown in (1).

$$CL_{final} = w_1 S_1 + w_2 S_2 + w_3 S_3 + w_4 S_4. \tag{1}$$

For the hybrid-level fusion, all the processes of feature-level and decision-level fusions are combined. There are different ways of making this combination; however, Figure 6c shows the hybrid-level fusion adopted for this work. The feature-level fusion part takes features (pupil dilation and heart rate in this case) as input and produces a level of CL as a sub-decision, while the other sub-decisions would be computed bypassing the feature vector into a classifier as discussed under decision-level fusion. The overall decision is the weighted average of all sub-decisions.

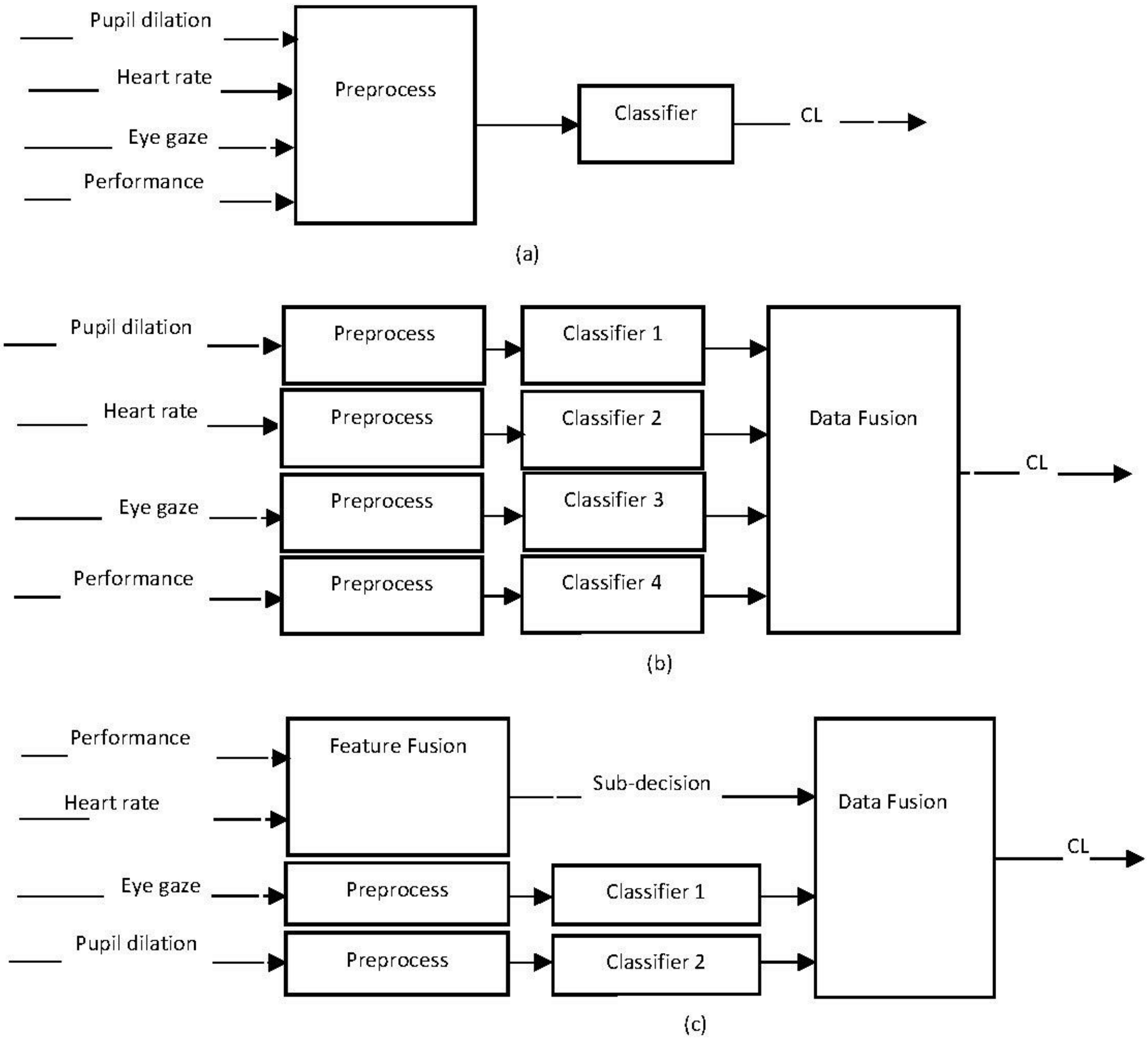

Figure 6. The structures of the three-level fusion approaches. (**a**): feature-level fusion, (**b**): decision-level fusion, and (**c**): hybrid-level fusion.

4.2.1. Feature-Level Fusion versus Single Classification Algorithms

All the classifiers listed in Table 4 were applied in feature-level fusion and single classification algorithms. For comparison, their accuracies were also presented in Table 5. As shown in the table, the best accuracy of each of the algorithms is written in bold and their average accuracies were also shown in the table. The best accuracy of feature-level fusion, 96.56%, is bigger than the best accuracy of each single classification algorithm. The best accuracies of the two features, performance features and pupil dilation, were from the SVM algorithm, while for heart rate and eye gaze, their best accuracies were obtained from the K-Nearest Neighbor. The best result achieved from the single classification algorithm, 94.72%, for the feature-level fusion was obtained from the SVM algorithm. This shows that the feature-level fusion outperformed all the single classification algorithms. These findings also suggest that the data fusion method can perform better than single classification algorithms by producing higher accuracy in CL measurement.

Table 4. Machine Learning Classifiers.

Classifier Index	Algorithm	Parameters
1	Decision Tree	Complex tree
2		Medium tree
3		Simple tree
4	SVM	Linear SVM
5		Quadratic SVM
6		Cubic SVM
7		Sigmoid SVM
8		Gaussian SVM
9		Polynomial SVM
10	Discriminant Analysis	Linear Discriminant Analysis
11		Quadratic Discriminant Analysis
12	KNN	Fine KNN
13		Medium KNN
14		Coarse KNN
15		Cosine KNN
16		Cubic KNN
17		Weighted KNN
18	ANN	Levenberg–Marquardt algorithm with 10 hidden neurons
19		Conjugate Gradient Backpropagation and with 10 hidden neurons
20		RPROP algorithm and with 10 hidden neurons
21		Gradient Descent with momentum and with 10 hidden neurons
22		Gradient Descent and with 10 hidden neurons

Table 5. Feature-Level Fusion and Accuracies of Single Classifiers (%).

Classifier Index	Pupil Dilation	Heart Rate	Eye Gaze	Performance Features	Feature Fusion
1	74.30	90.61	87.32	92.94	94.92
2	94.71	87.22	78.56	79.43	95.32
3	85.32	91.72	83.34	90.73	92.78
4	93.71	81.71	73.43	**94.60**	90.23
5	76.10	91.10	84.32	90.73	94.34
6	74.73	87.81	83.72	89.62	87.89
7	47.12	77.81	80.65	68.92	79.04
8	**94.72**	86.11	67.43	88.53	90.43
9	92.00	87.8	78.76	87.42	91.03
10	93.00	84.42	86.51	68.32	91.56
11	90.61	90.00	87.97	67.21	87.65
12	83.21	**92.20**	76.43	87.43	85.43
13	93.1	88.31	81.00	86.9	89.67
14	92.71	70.00	79.31	65.62	88.98
15	84.12	81.70	**90.45**	86.93	87.96
16	94.70	87.81	89.00	85.23	86.89
17	91.90	90.64	73.65	88.00	90.87
18	73.98	89.31	84.76	94.00	94.87
19	93.42	87.91	67.78	82.91	95.76
20	89.40	90.80	84.34	56.01	**96.56**
21	91.61	71.24	78.84	61.34	94.00
22	82.81	62.23	90.43	55.23	93.43
Average	85.79	84.93	81.27	80.37	90.89

Note: Bold font means best accuracies of each algorithm.

4.2.2. Decision-Level Fusion versus Single Classification Algorithms

The resultant decision-level fusion is the weighted average of the four (4) sub-decisions indicated in Figure 6b. For every sub-decision, various weights have been tested to determine the best accuracy for a particular algorithm. All the classifiers in Table 4 were applied for each of the sub-decisions. The best decision-level fusion accuracy was 94.67%, which was comparable to the best accuracy of the feature-level fusion.

4.2.3. Hybrid-Level Fusion versus Single Classification Algorithms

Hybrid-level fusion performed better than the feature-level and decision-level fusions with the highest accuracy of 97.14%. The best accuracy was realized when pupil dilation and performance were combined for sub-decision one with the SVM algorithm, heart rate for sub-decision two with the KNN algorithm, and eye gaze for sub-decision three with KNN.

5. Discussions of Results

The primary target of the research is to determine the effects of neurocognitive load on learning transfer from a novel VR-based driving system. As predicted, the addition of several turns, intersections, and landmarks on the difficult routes elicited an increase in psychophysiological activation, such as an increase in pupil dilation, heart rate, and eye gaze. Thus, our discussions would be as follows.

5.1. Psychophysiological Response Patterns Associated with Cognitive Load

These findings of an increase in heart rate with the increase in cognitive demand are supported by several studies. Task difficulty elicits an increase in psychophysiological activation, such as heart rate [21,43,44]. Heart rate increases while the overall Heart Rate Variability decreases when mental effort increases [45]. As Verway et al. [46] reported, in a case of participants subjected to cognitive tasks while driving compared to those in control in which no cognitive task was performed, the results showed that participants indicated increased heart rate and reduced HRV when performing the cognitive task. Moreover, Mohanavelu et al. [47] presented a cognitive workload analysis of fighter pilots in a high-fidelity flight simulator environment during different flying workload conditions. The results showed that HRV features were significant in all flying segments across all workload conditions.

Our findings related to pupil dilation and the cognitive load were also supported by Pomplun et al. [20]. In this study, they came up with a gaze-controlled human–computer interaction (HCI) task that ran at three different speeds with three different levels of task difficulty. Each of these levels of task difficulty was combined with two levels of background brightness, making six different trial types. Each type was shown to each of the participants four times. Before the commencement of the experiment, participants were asked not to let any blue circle reach its full size. The results showed that the pupil diameter was significantly affected by the task difficulty. In another study, Palinko et al. [48] evaluated the driver's CL associated with pupil diameter measurements from a remote eye tracker. They compared the CL estimates based on the physiological pupillometric data and participant's performance data. The results obtained show that the performance and physiological data largely agree with the task difficulty.

The use of performance features is a fundamental assessment of cognitive load [49]. Important features, such as intersection [50], wrong count, and speed [51], are considered to be performance indicators for a cognitive load. Speed has been shown to decrease as workload increases [51]. According to Engström J et al., entering into uncertain situations such as a complex non-signalized intersection increases a cognitive load [50]. All the aforementioned results are in agreement with our findings.

5.2. Multimodal Data Fusion

As shown in Table 5, the feature-level fusion outperformed all the single classification algorithms in CL measurement. This can be seen as their best accuracy, and the average accuracy is shown in the table. Several types of research that use data fusion are in existence in the literature [31,34]. Barua S et al. [31] employ ML's data fusion method to detect and classify different driver states based on physiological data. They used several ML algorithms to determine the accuracy of sleepiness, cognitive load, and stress classification. The results show that combining features from several data sources improved performance by 10–20% compared to using features from a single classification algorithm. In another development, X Zhang et al. [34] proposed an ML method using 46 kinds of photoplethysmogram (PPG) features to improve the cognitive load's measurement accuracy. They tested the method on 16 different participants through the classical n-back tasks (0-back, 1-back, and 2-back). The accuracy of the machine learning method in differentiating different levels of cognitive loads induced by task difficulties can reach 100% in 0-back vs. 2-back tasks, which outperformed the traditional HRV-based and single-PPG-feature-based methods by 12–55%. Even though these studies were not designed to evaluate the effects of neurocognitive load on learning transfer, the results obtained in our study are in agreement with what is available in the existing results in measuring cognitive load using the data fusion method.

Putze F et al. [33] applied a simple majority voting fusion in combining skin conductance, EEG, respiration, and pulse to categorize CL in visual and cognitive tasks. The results revealed that the decision-level fusion outperformed the single modality method in one task, while it was surpassed in other tasks.

In another study by Hussain S et al. [32], they combined the features GSR, ECG, Eye, and RESP from physiological sensors into a classification model, and participant's task performance features were applied to different classification models; sub-decisions were then combined using majority voting. This hybrid-level fusion approach improved the classification accuracy by 6% compared to single classification methods.

6. Conclusions and Future Work

Learning transfer is of paramount concern for training researchers and practitioners. However, whenever the learning task requires too much cognitive workload, it makes it difficult for the transfer of learning to occur. The main contribution of this paper is to systematically present the cognitive workload measurements of individuals based on their heart rate, eye gaze, pupil dilation, and performance features obtained when they used the VR-based driving system. Data fusion methods were used to accurately measure the cognitive load of these users. Easy routes and difficult routes were used to induce different cognitive loads. Five (5) well-known ML algorithms were considered in classifying individual modality features and multimodal fusion. The best accuracies of the two features performance features and pupil dilation were obtained from the SVM algorithm, while for the heart rate and eye gaze, their best accuracies were obtained from the KNN method. The multimodal fusion approaches outperformed single-feature-based methods in cognitive load measurement.

Moreover, all the hypotheses set aside in this paper have been achieved. One of the goals of the experiment was that the addition of several turns, intersections, and landmarks on the difficult routes would elicit increased psychophysiological activation, such as increased heart rate, eye gaze, and pupil dilation. In line with the previous studies, the VR platform was able to show that the increase in difficulty level found in tasks such as the difficult routes directly induce autonomic changes in psychophysiological arousal. We also showed that due to an increase in psychophysiological activation, participants committed more mistakes when driving on difficult routes. Making more mistakes was found to be associated with difficult routes. Moreover, increased difficulty was shown to cause an increase in psychophysiological responses.

To this end, it might be beneficial to consider the data fusion approach when dealing with physiological data such as the ones used in this article. The methods used in this writeup could provide significant benefits in developing a knowledge-based or decision support system that could provide a reasonable means for physiological sensor signal-based applications.

Our recommendations for future work are based on the results and limitations of the work described in this article:

a. Among the limitations of this work is the use of the data collected from a driving simulator. Even though the data collected from driving simulators are controllable and reproducible, and it is also possible to encounter dangerous driving conditions without the risk of physical injury, there are challenges attached, such as motion sickness, driving a simulator can be boring, it can be more demanding to stay alert, and participants can be biased towards a false sense of safety. Thus, there is a need to use the data collected in real-world driving situations in future work and evaluate the proposed approaches.
b. The correlation among physiological measures such as heart rate, pupil dilation, and driving performance data would be considered and used as a reference measure in future work. This will help in reducing the complexity of measuring these physiological measures in real driving situations.
c. In combining the sub-decisions for the final decision in the data fusion method, other methods, such as majority voting and classification algorithms, would be considered instead of the weighted average method. These and other different approaches could be explored in future work.
d. There is also a need to consider the effects of neuro-cognitive load on gender-based learning using a VR driving system.

Author Contributions: Conceptualization and methodology: U.A.A. and S.-C.Y.; software, validation, formal analysis, investigation, and writing: U.A.A.; review, editing, and supervision: S.-C.Y.; resources and data curation: U.A.A., Y.W., L.W. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the China Scholarship Council (CSC) (2018GXZ021733).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Britannica and Inc Encyclopaedia. *Encyclopaedia Britannica*; Britannica and Inc Encyclopaedia: Chicago, IL, USA, 1957.
2. Si, J.; Dongsik, K. How do instructional sequencing methods affect cognitive load, learning transfer, and learning time. *Educ. Res.* **2011**, *2*, 1362–1372.
3. De Jong, T. Cognitive load theory, educational research, and instructional design: Some food for thought. *Instr. Sci.* **2010**, *38*, 105–134. [CrossRef]
4. Shibli, D.; Rachel, W. Cognitive Load Theory and Its Application in the Classroom. Available online: https://impact.chartered.college/article/shibli-cognitive-load-theory-classroom/ (accessed on 23 November 2020).
5. Choi, H.; Van Merriënboer, J.J.; Fred, P. Effects of the physical environment on cognitive load and learning: Towards a new model of cognitive load. *Educ. Psychol. Rev.* **2014**, *26*, 225–244. [CrossRef]
6. Sweller, J. Cognitive load during problem-solving: Effects on learning. *Cogn. Sci.* **1988**, *12*, 257–285. [CrossRef]
7. Miller, G.A. The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychol. Rev.* **1956**, *2*, 81. [CrossRef]
8. Bulling, A.; Thorsten, O.Z. Cognition-aware computing. *IEEE Pervasive Comput.* **2014**, *13*, 80–83. [CrossRef]
9. Paas, F.G.; Van Merriënboer, J.J. Instructional control of cognitive load in the training of complex cognitive tasks. *Educ. Psychol. Rev.* **1994**, *6*, 351–371. [CrossRef]
10. Paas, F.; Tuovinen, J.E.; Tabbers, H.; Van Gerven, P.W. Cognitive load measurement as a means to advance cognitive load theory. *Educ. Psychol.* **2003**, *38*, 63–71. [CrossRef]
11. Sweller, J. Cognitive load theory, learning difficulty, and instructional design. *Learn. Instr.* **1994**, *4*, 295–312. [CrossRef]
12. Schnotz, W.; Christian, K. A reconsideration of cognitive load theory. *Educ. Psychol. Rev.* **2007**, *19*, 469–508. [CrossRef]

13. Meshkati, N.; Hancock, P.A.; Rahimi, M.; Dawes, S.M. *Techniques in Mental Workload Assessment*; Taylor & Francis: Abingdon-on-Thames, UK, 1995.
14. Naismith, L.M.; Cavalcanti, R.B. Validity of cognitive load measures in simulation-based training: A systematic review. *Acad. Med.* **2015**, *90*, S24–S35. [CrossRef]
15. Naismith, L.M.; Cheung, J.J.; Ringsted, C.; Cavalcanti, R.B. Limitations of subjective cognitive load measures in simulation-based procedural training. *Med. Educ.* **2015**, *49*, 805–814. [CrossRef]
16. Imhoff, S.; Lavallière, M.; Teasdale, N.; Fait, P. Driving assessment and rehabilitation using a driving simulator in individuals with traumatic brain injury: A scoping review. *NeuroRehabilitation* **2016**, *39*, 239–251. [CrossRef]
17. Ettenhofer, M.L.; Guise, B.; Brandler, B.; Bittner, K.; Gimbel, S.I.; Cordero, E.; Nelson Schmitt, S.; Williams, K.; Cox, D.; Roy, M.J.; et al. Neurocognitive driving rehabilitation in virtual environments (NeuroDRIVE): A pilot clinical trial for chronic traumatic brain injury. *NeuroRehabilitation* **2019**, *44*, 531–544. [CrossRef]
18. Lundqvist, A.; Rönnberg, J. Driving problems and adaptive driving behavior after brain injury: A qualitative assessment. *Neuropsychol. Rehabil.* **2001**, *11*, 171–185. [CrossRef]
19. Kosch, T.; Hassib, M.; Buschek, D.; Schmidt, A. Look into My Eyes: Using Pupil Dilation to Estimate Mental Workload for Task Complexity Adaptation. In Proceedings of the Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems, Montreal, QC, Canada, 21–26 April 2018.
20. Pomplun, M.; Sunkara, S. Pupil dilation as an indicator of cognitive workload in human-computer interaction. In Proceedings of the International Conference on HCI, Crete, Greece, 22–27 June 2003.
21. Mehler, B.; Reimer, B.; Coughlin, J.F.; Jeffery, A. Dusek. Impact of incremental increases in cognitive workload on physiological arousal and performance in young adult drivers. *Transp. Res. Rec.* **2009**, *2138*, 6–12. [CrossRef]
22. Kennedy, D.O.; Andrew, B.S. Glucose administration, heart rate, and cognitive performance: Effects of increasing mental effort. *Psychopharmacology* **2000**, *149*, 63–71. [CrossRef]
23. Carroll, D.; Rick Turner, J.; Hellawell, J.C. Heart rate and oxygen consumption during the g active psychological challenge: The effects of level of difficulty. *Psychophysiology* **1986**, *23*, 174–181. [CrossRef]
24. Jorna, P.G. Spectral analysis of heart rate and psychological state: A review of its validity as a workload index. *Biol. Psychol.* **1992**, *34*, 237–257. [CrossRef]
25. Charlton, S.G.; O'Brien, T.G. (Eds.) *Handbook of Human Factors Testing and Evaluation*; CRC Press: Boca Raton, FL, USA, 2019.
26. Lenneman, J.K.; Shelly, A.R.; Backs, R.W. Deciphering psychological-physiological mappings while driving and performing a secondary memory task. In Proceedings of the 3rd International Driving Symposium on Human Factors in Driver Assessment, Training, and Vehicle Design, Rockport Maine, ME, USA, 27–30 June 2005.
27. Backs, R.W.; Seljos, K.A. Metabolic and cardiorespiratory measures of mental effort: The effects of level of difficulty in a working memory task. *Int. J. Psychophysiol.* **1994**, *16*, 57–68. [CrossRef]
28. Brookhuis, K.A.; De Waard, D. *Assessment of Drivers' Workload: Performance and Subjective and Physiological Indexes*; Stress, Workload and Fatigue; Laurence Erlbaum Associates Inc.: Mahwah, NJ, USA, 2001.
29. Veltman, J.A.; Gaillard, A.W.K. Physiological workload reactions to increasing levels of task difficulty. *Ergonomics* **1998**, *41*, 656–669. [CrossRef] [PubMed]
30. Zhang, L.; Wade, J.; Bian, D.; Fan, J.; Swanson, A.; Weitlauf, A.; Warren, Z.; Sarkar, N. Cognitive load measurement in a virtual reality-based driving system for autism intervention. *IEEE Trans. Affect. Comput.* **2017**, *8*, 176–189. [CrossRef] [PubMed]
31. Barua, S. Multivariate Data Analytics to Identify Driver's Sleepiness, Cognitive Load, and Stress. Ph.D. Dissertation, Mälardalen University, Västerås, Sweden, 2019.
32. Hussain, S.; Chen, S.; Calvo, R.A.; Chen, F. Classification of cognitive load from task performance & multichannel physiology during affective changes. In Proceedings of the Conference on Multimodal Interaction, Alicante, Spain, 14–18 November 2011.
33. Putze, F.; Jarvis, J.P.; Schultz, T. Multimodal recognition of cognitive workload for multitasking in the car. In Proceedings of the 20th International Conference on Pattern Recognition, Istanbul, Turkey, 23–26 August 2010; pp. 3748–3751.
34. Zhang, X.; Lyu, Y.; Qu, T.; Qiu, P.; Luo, X.; Zhang, J.; Fan, S.; Shi, Y. Photoplethysmogram-based Cognitive Load Assessment Using Multi-Feature Fusion Model. *ACM Trans. Appl. Percept.* **2019**, *16*, 1–7. [CrossRef]
35. Autodesk. Available online: https://www.autodesk.com/ (accessed on 21 December 2020).
36. ArcGIS CityEngine. Available online: https://www.esri.com/en-us/arcgis/products/arcgis-cityengine/ (accessed on 21 December 2020).
37. Unity3D. Available online: www.unity3d.com (accessed on 20 December 2020).
38. VIVE. Available online: https://www.vive.com/eu/product/vive-pro-eye/ (accessed on 23 February 2020).
39. Komogortsev, O.V.; Gobert, D.V.; Jayarathna, S.; Gowda, S.M. Standardization of automated analyses of oculomotor fixation and saccadic behaviors. *IEEE Trans. Biomed. Eng.* **2010**, *57*, 2635–2645. [CrossRef]
40. Benedetto, S.; Pedrotti, M.; Minin, L.; Baccino, T.; Baccino, T.; Re, A.; Montanari, R. Driver workload and eye blink duration. *Transp. Res. Part F Traffic Psychol. Behav.* **2011**, *14*, 199–208. [CrossRef]
41. Lahiri, U.; Warren, Z.; Sarkar, N. Design of a gaze-sensitive virtual social interactive system for children with autism. *IEEE Trans. Neural Syst. Rehabil. Eng.* **2011**, *19*, 443–452. [CrossRef]
42. Atrey, P.K.; Hossain, M.A.; El Saddik, A.; Kankanhalli, M.S. Multimodal fusion for multimedia analysis: A survey. *Multimed. Syst.* **2010**, *16*, 345–379. [CrossRef]

43. Causse, M.; Baracat, B.; Pastor, J.; Dehais, F. Reward and uncertainty favor risky decision-making in pilots: Evidence from cardiovascular and oculometric measurements. *Appl. Psychophysiol. Biofeedback* **2011**, *36*, 231–242. [CrossRef] [PubMed]
44. Boutcher, Y.N.; Stephen, H.B. Cardiovascular response to Stroop: Effect of verbal response and task difficulty. *Biol. Psychol.* **2006**, *73*, 235–241. [CrossRef]
45. Gabaude, C.; Baracat, B.; Jallais, C.; Bonniaud, M.; Fort, A. Cognitive load measurement while driving. In *Human Factors: A View from an Integrative Perspective*; Human Factors and Ergonomics Society: Washington, DC, USA, 2012; p. 67.
46. Verwey, W.B.; Veltman, H.A. Detecting short periods of elevated workload: A comparison of nine workload assessment techniques. *J. Exp. Psychol. Appl.* **1996**, *2*, 270. [CrossRef]
47. Mohanavelu, K.; Poonguzhali, S.; Ravi, D.; Singh, P.K.; Mahajabin, M.; Ramachandran, K.; Singh, U.K.; Jayaraman, S. Cognitive Workload Analysis of Fighter Aircraft Pilots in Flight Simulator Environment. *Def. Sci. J.* **2020**, *70*, 131–139. [CrossRef]
48. Palinko, O.; Kun, A.L.; Shyrokov, A.; Heeman, P. Estimating cognitive load using remote eye tracking in a driving simulator. In Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications, New York, NY, USA, 22–24 March 2010; pp. 141–144.
49. Miller, S. *Workload Measures*; National Advanced Driving Simulator: Iowa City, IA, USA, 2001.
50. Engström, J.; Markkula, G.; Victor, T.; Merat, N. Effects of cognitive load on driving performance: The cognitive control hypothesis. *Hum. Factors* **2017**, *59*, 734–764. [CrossRef] [PubMed]
51. Casali, J.G.; Walter, W.W. A comparison of rating scale, secondary-task, physiological, and primary-task workload estimation techniques in a simulated flight task emphasizing communications load. *Hum. Factors* **1983**, *25*, 623–641. [CrossRef] [PubMed]