https://dx.doi.org/10.1016/j.knosys.2021.108047

Artificial Intelligence-enabled Non-intrusive Vigilance Assessment Approach to Reducing Traffic Controller's Human Errors

Fan LI [1], Chun-Hsien CHEN [2], Ching-Hung LEE[3*], Shanshan FENG[4]

[1]*Department of Aeronautical and Aviation Engineering, The Hong Kong Polytechnic University, Hong Kong, China,* fan-5.li@polyu.edu.hk

[2]*School of Mechanical and Aerospace Engineering, Nanyang Technological University, Singapore*， MCHchen@ntu.edu.sg

[3]*School of Public Policy and Administration, Xi'an Jiaotong University, China*

leechinghung@xjtu.edu.cn

[4]*School of Computer Science and Technology, Harbin Institute of Technology (Shenzhen), China*

sanlang333@gmail.com

* Corresponding author

E-mail: leechinghung@xjtu.edu.cn

Abstract:

To be vigilant is highly required for traffic controllers in transportation fields, such as air traffic management, vessel traffic service, and railway management, as they need to monitor traffic conditions and notice any potential hazards. Hence, emerging studies have been conducted to develop an objective and non-intrusive approach to assessing vigilance levels and generate warnings if needed. This study aims to investigate the effects of impaired vigilance on human performance via non-intrusive data analysis, namely spatial and temporal gaze pattern analytics, and develop an objective model for vigilance assessment accordingly. A novel four-phase framework, including vigilance test design, non-intrusive data collection, spatial and temporal gaze pattern analytics, and a shallow neural network-based model was proposed to achieve this aim. Meanwhile, an illustrative experiment in the maritime industry was conducted to verify the proposed method. The spatial and temporal gaze patterns analytics revealed that low vigilance levels impacted comprehension time but not perception time, with longer fixations duration but stable time-to-the-nearest-fixation under a low vigilance level. It is found that even a person with impaired vigilance can quickly notice abnormal events. The effectiveness and empirical implications of this model can help traffic controllers avoid fatigue-induced vigilance reduction. In addition, it provides evidence, references, and solutions for designing human-computer interfaces to reduce human errors caused by low vigilance.

Keywords: Eye-tracking; gaze pattern; fatigue; maritime; human performance, shallow neural network.

1. Introduction:

The term vigilance normally is used to represent the sustained attention over a lengthy period of cognitive work [1]. With the increased adoption of automation technology in the workplaces, the role of an operator tends to change from a controller to a supervisor [2-4], whose sustained attention are critical in detecting potential threats and coping with the possible failures of the automated systems [5, 6]. Accordingly, the operations tend to change from physical tasks to prolonged cognitive tasks, such as decision-making and monitoring [7, 8]. These prolonged cognitive tasks would cause impaired vigilance, which induces the increased response time and a high probability of human errors [9-11]. An operator with impaired vigilance may fail to notice the warnings and miss a potentially unsafe situation developing. Statistics of Europe showed that more than 10% of traffic accidents were caused by vigilance decrement [12]. In 1993, an American cargo plane crashed and burned due to the pilots' impaired vigilant situational awareness of the airplane. The accident caused huge economic loss, it was luck that all crew members survived.

Owing to the huge consequence of impaired vigilance, extensive studies have been conducted to assess vigilance levels, improve vigilance, and design interventions to recovery from low vigilance states [13]. Among them, assessing the vigilance levels is the first step and one of the most important steps. Vigilance assessments serve as reactive strategies that alert operators once vigilance decrement occurs. Classically, vigilance can be objectively measured using physiological responses [14], such as brain dynamics, blood flow velocity, and eye movements, which can provide relatively reliable assessments for detecting the subtle changes in vigilance [13-16]. For example, brain dynamics data from 2-hour virtual reality-based simulated driving tasks were used to build a vigilance model [17]. The eye-tracking data from 40-minutes air traffic control tasks were utilized to assess the vigilance levels [18].

This study proposes to use the eye-tracking technique to analyze the vigilance stages [19, 20], as it can collect non-intrusive data. This idea is supported by decades of scientific evidence indicated that there could be a link between mental activities and gaze movements [21-23]. With advanced infrared technology, gaze movements can be remotely captured in a non-intrusive manner [24]. In addition, as an important information channel, vision presents most of the human-computer interactions in conducting cognitive tasks, especially in supervisory operations [25]. Previous studies found that the fixation duration is correlated with comprehension efficiency and the capacity of cognitive sources [26]. Saccades are required during the process of target search. In the authors' previous study, we have confirmed that the relations between saccade peak velocity and the reaction time are affected by cognitive states [27]. Besides, it has long been found that there are close relations between vigilance stages and gaze patterns. For example, Zheng and Lu [17] proposed an eye-tracking-based indicator of vigilance, called as PERCLOS, by calculating the duration of blinks, fixations, and saccades. Hence, it is expected that gaze movements analytics can be used to assess vigilance stages.

Nevertheless, to achieve the gaze patterns-based assessment of vigilance, several challenges have to be addressed. First, gaze patterns would be significantly affected by task types [28]. The gaze patterns of monitoring would be different from gaze patterns of control operations. Nevertheless, there are limited studies that analyzed gaze movement in vigilance tests. Hence, it is necessary to design a general and basic task for fully understanding the general gaze patterns across different vigilance stages. Second, the complex context, such as time-on-task and fatigue would affect gaze movements [13, 29], and thusly should be considered during data collection. A systematic way to collect data should be proposed. Third, it is hard to figure out the suitable indicators of vigilance from gaze patterns, as they have close relations with the decision-making process. In other words, it is difficult to determine whether the specific gaze patterns are caused by impaired vigilance or a special decision-making process. Finally, the relations between vigilance stages and gaze movements are dynamic and complex [30], the traditional linear regression is not suitable for dealing with gaze movements. Hence, the recently widely applied data mining and machine learning methods may be applied [31, 32].

To address these challenges, a four-phase framework is proposed, including vigilance test design, non-intrusive data collection, spatial and temporal gaze movement analytics, and artificial intelligence (AI)-based model for vigilance assessment. Vigilance stages normally have great effects on hazard identification [13]. To simulate the hard identification operations, we designed two tasks, namely simple response tasks and multichoice tasks. Both of them require subjects to notice a sudden change displayed in the interface and make a corresponding response. The response time, which is one of the most widely used vigilance labels [33], is collected as the performance of vigilance. In the second phase, a general structure for collecting behavior data, context data, gaze movement data is proposed. To well understanding how vigilance stages impact the cognitive process of human performance, a spatial and temporal gaze pattern analytic method is proposed in the third phase. Finally, a shallow neural network is adopted to establish a vigilance model.

An experiment in vessel traffic service operations was conducted to demonstrate and verify the proposed framework and test the following hypotheses: (1) H1: Vigilance stages has significant effects on gaze patterns; (2) H2: The task complexity and fatigue level have interaction effects on gaze patterns; (3) H3: the gaze patterns can be used to assess vigilance. The following sections are structured as follows. The state-of-the-art literature is reviewed in Section 2. Section 3 describes the proposed framework. Section 4 presents the case study and results and is followed by discussions in Section 5. The contributions, limitations, and future works are concluded in Section 6.

2. Literature Review

2.1 Vigilance Assessment

Vigilance has a similar meaning to watchfulness, it is a kind of state or quality of being vigilant. Keeping vigilant is greatly important in safety-security industries, such as air traffic management, vessel traffic service, and nuclear power plant [13]. Hence, several studies have been conducted to assess vigilance and tried to alert operators with low vigilance. Historically, vigilance assessment was conducted via measuring task performance, such as lane departure and local error rates [17, 34]. In lab-based experiments, subjects are normally asked to the vigilance tests, such as the psychomotor vigilance task and Mackworth Clock test, which can induce impaired vigilance and the collected response is the measurement of vigilance states [35].

Recently, as the wearable devices for collecting psychological data matured, much of the related research was geared toward bio signal-based vigilance assessment [13]. Classically, the vigilance stages are described in terms of variations in our psychophysiological data, such as brain dynamics, eye movements, and heart rate [36-38]. Among them, brain dynamics have received the most research attention. A study predicted vigilance stages by using the 256-channel electroencephalogram (EEG) functional connectivity metrics [36], and the different features extracted from EEG were tested. Besides multi-channel-based assessment, recent studies tried to assess vigilance using EEG data collected from two channels or a single channel [39]. A recent study proposed two EEG-channel configurations, which can classify vigilance changes of ATCOs' vigilance [35]. Although the EEG-based approach seems can accurately estimate the vigilance level, it is not currently suitable for practical applications owing to the uncomfortableness of wearing an EEG device.

Though some studies have tried to use gaze movements to assess vigilance, none of them has deeply investigated the gaze patterns along with the decision-making under the states of high or low vigilance.

2.2 Gaze patterns and cognitive states

The eye-tracking technique normally captures two events, including fixations and saccades [40]. Fixations mean maintaining gazes on a location with limited ranges and low gaze velocity. A saccade is a quick movement from one fixation to another with a high gaze velocity [41, 42]. Movements of the eyeballs are the primary way to direct attention [40, 43]. Reading and comprehension usually occur during fixations [44]. It has long been found that gaze movement data can provide objective and reliable indicators for cognitive states [45, 46]. Several parameters would be generated from the two basic events, such as the saccadic speed [46], the number of fixations [12], the saccade rate, the peak saccade velocity, the amplitude and duration of saccades [47], and the latency [48]. Besides, by considering the dynamic or static aggregations of fixations and saccades, some parameters are generated, such as fixation sequences [49] and heat maps [34].

A number of studies have been conducted to investigate the correlations between gaze movements parameters and cognitive states, such as fatigue, workload, vigilance, and situation awareness [49-52]. For example, accumulated fatigue will decrease the saccadic speed, the peak saccade velocity, and the amplitude of saccades [48, 53] but increase the saccade duration [47]. In addition, gaze spatial distribution can be used to monitor or assess the inattention level of users. Normally the optimal variations in eye fixation movement depend on the research problem [54].

3. AI-enabled non-intrusive vigilance assessment

According to the literature review presented in Section 2, gaze patterns analytics may reveal the cognitive states of human beings and is expected to be used for vigilance assessment. Hence, a general framework with four phases is proposed to assess vigilance based on gaze patterns. The overall architecture of the proposed framework is presented in Figure 1. Four phases are involved in it: the vigilance test, data collection, data analysis, and AI-based model. The user interfaces for phase 1 and 2 includes a Tobii X3-120 with a sampling frequency of 120 Hz and a computer for displaying the vigilance test. A data management system was developed to collect and keep eye-tracking data, human behaviors, and other related information. The four modules are discussed in detail as follows.

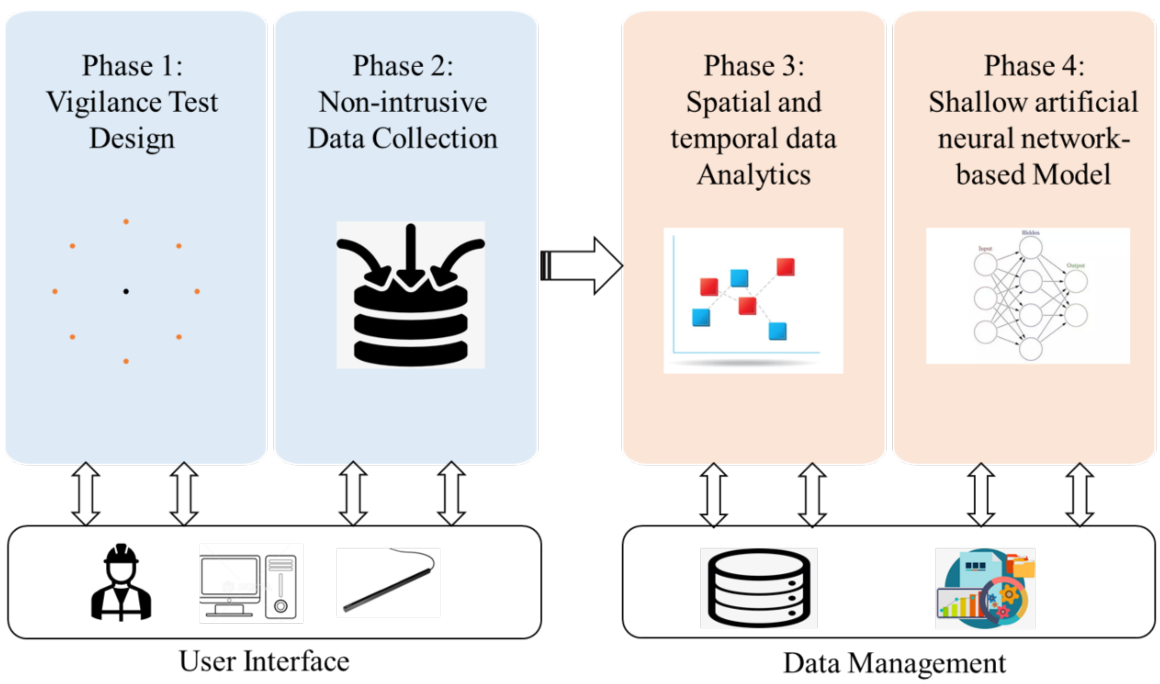

Figure 1: The vigilance assessment system framework

3.1 Vigilance test design

Though the gaze patterns reveal a part of the mental process, the complex and dynamic interactions among the mental process are still hard to be analyzed. Since the mental process may be fleeting, the perception of the stimulus and comprehension may occur simultaneously. Even the advanced eye-tracking technique cannot clearly distinguish them. To address this problem, we designed two tasks with different task complexities, namely a simple reaction time test and a multi-choice reaction time test. For the simple reaction time test, participants are asked to press the J key whenever an orange circle appeared. The participant is only required to acknowledge the presence of the visual cue. The location of the orange circle is randomized, but it is always either at one of the pre-defined eight locations, as shown in Figure 2. The locations are chosen to be as far away as possible and have the same distance to the center dot. For the multi-choice reaction time test, participants should press the J key when vowel letters (A, I, U, E, O) appear and press the F key when consonant letters (K, L, T, V, X) appear. Similar to the first test, the locations of the vowel letters and consonant letters are randomized. In the easy tasks, only the perception of stimulus and actions are required. In complex tasks, the whole mental process, including the perception of stimulus, comprehension, and actions is required.

The two reaction time tests were designed with OpenSesame, a widely utilized program to create experiments for psychology, neuroscience, and experimental economics. This task was similar to most of the traditional and classic reaction time tests [27]. The differences lie in the characteristics and location of the target cue. Roman alphabets and multiple choices were used in this experiment, as they have higher complexity and can force subjects to fixate on the visual cues. To induce saccades, the locations of the visual cues vary in each trail. Human performance was assessed in terms of the percentage of correct signals detected and reaction time. These variables were averaged every 120 trials.

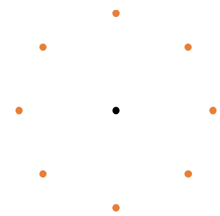

Figure 2: The eight predefined locations and the center dot

3.2 Non-intrusive data collection

Data are collected from several sources, namely eye tracker, questionnaire, OpenSesame, and experiment notes, as shown in Figure 3. The raw data of the eye tracker include timestamps (t), gaze locations (x, y), and gaze velocities. In this study, the gaze locations (x, y) generated by Tobii X3-120 are measured in pixels and transferred into millimeters. The velocity threshold identification algorithm (I-VT) [55] was used to extract fixations and saccades from the raw data. The fixation is a point where one focuses on and precedes its detailed information. The fixation is aggregated by many consecutive gaze points located in a specified area [56].

The questionnaire collects demographic data, such as name, age, gender, experience, and fatigue level. The Samn-Perelli Fatigue Scale [57, 58] was used to measure fatigue level. It uses the 7-point Likert scale, where "1" represents "fully alert, wide awake," and "7" denotes "completely exhausted, unable to function effectively."

The vigilance performance, namely response time and hit rate is collected by OpenSesame. The reaction time (RT) measures the interval time between the appearance of a stimulus and the response of participants on a button [59]. It is a widely used index of psychomotor vigilance performance [60, 61]. Besides the reaction time, the hit rate is commonly used to assess vigilance, too [18, 38]. The hit rate refers to the percentage of correct signals detected. Other context information, such as time-on-task is recorded in experiment notes by experts.

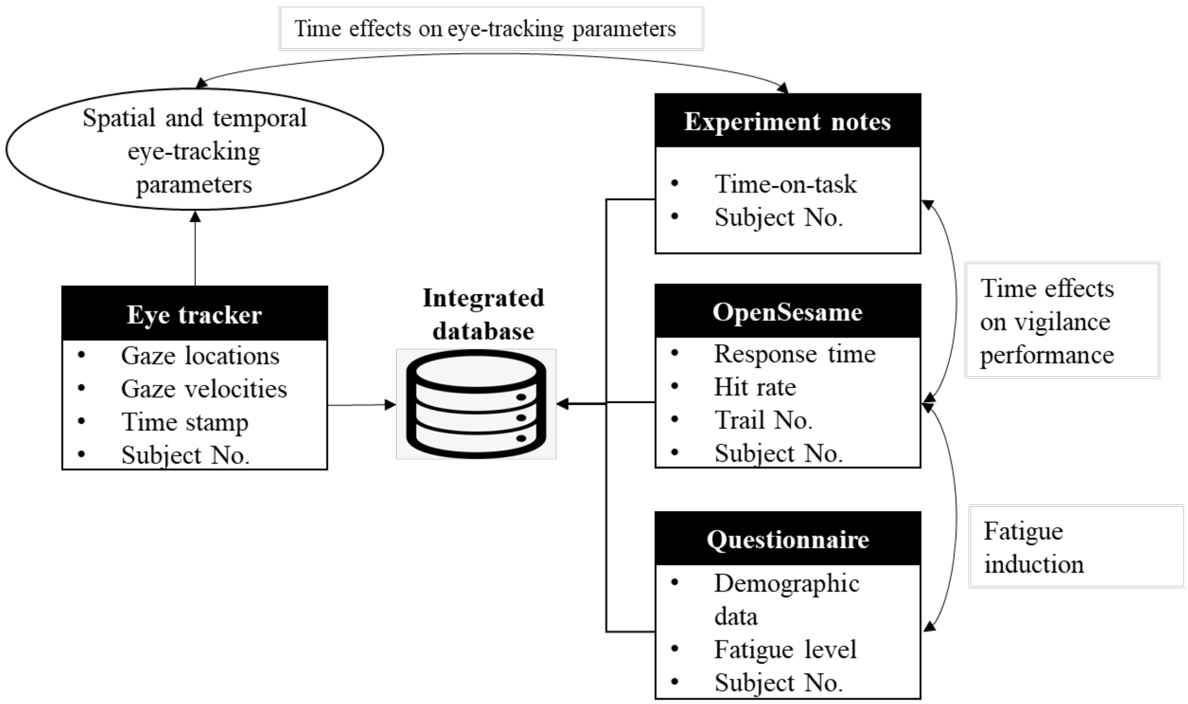

Figure 3: The diagram of data collection and analysis

3.2 Spatial and temporal gaze pattern analytics

The gaze movements during the task were presented in Figure 4. The fixation that occurs at the appearance of the target is named F0, and the fixation that locates nearest the target is named F1. Several spatial and temporal parameters are generated based on the two fixations. For the temporal parameters, the time to F1(T_{F1}) and F1 duration (D_{F1}) are extracted. As shown in Figure 4, T_{F1} means the time elapsed from the appearance of the target to the fixation that locates nearest the target (F1). As shown in Figure 5, for the spatial parameters, the distance between F0 and target (DT_{F0}) and distance between F1 and target (DT_{F1}) are calculated based on the following equations:

$$D = \sqrt[2]{(\hat{x} - x_t)^2 + (\hat{y} - y_t)^2} \tag{1}$$

(x_t, y_t) refer to the position of the target which is measured in millimeters.

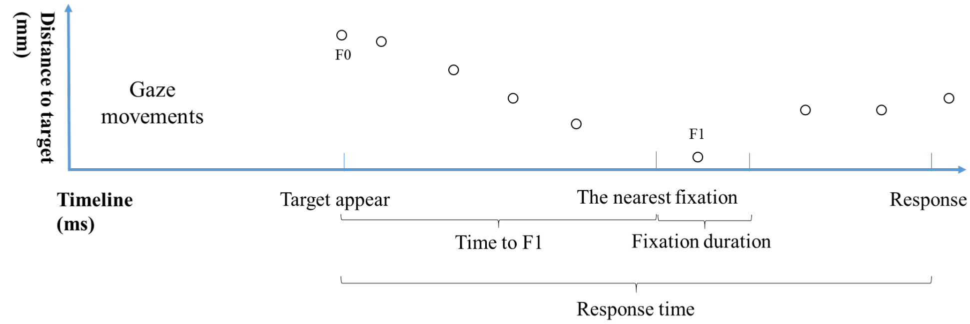

Figure 4: Temporal analysis of gaze movements in reaction time test

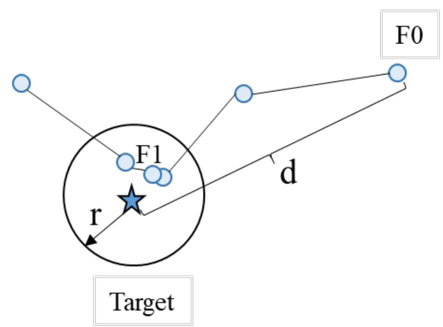

Figure 5: Spatial analysis of gaze movements in reaction time test

Also, we proposed to consider the number of valid fixations. The valid fixations were defined as the ones that were located near targets. Depending on the nature of the task, it is normally recommended that the dispersion of fixations should be a radius of 0.25° to 0.5° [62, 63]. For stimuli that contain mostly pictures, Tobii Technology [55] recommends a fixation radius of 50 pixels. Considering the recommendations and the experiment setup mentioned in Section 2.3, the radius of fixations is defined as 40 pixels. Hence, fixations that are located in 40 pixels of targets are defined as valid fixations. The ratio of valid fixations (RVF) is the number of valid fixations divided by the number of all fixations that occur between F0 and F1.

3.3 Shallow artificial neural network-based vigilance model

In this section, we develop a deep learning method to assess human performance based on gaze patterns. The input and output variables are summarized in Figure 10. The input variables include indicators that are extracted from the spatial and temporal gaze pattern analytics, namely the time to F1(TF1), F1 duration (DF1), the distance between F0 and target (DT_{F0}), the distance between F1 and target (DT_{F1}), and the ratio of valid fixations (RVF). The output variable is the response time, as shown in Figure 10.

Considering the great individual difference, a subject-dependent model was developed with a shallow artificial neural network (SANN). It has a layered structure and a set of interconnected nodes, as presented in Figure 11. The SANN has been widely used in recognition, classification, and prediction. Comparing with the deep artificial neural network, it can save computational time and improve efficiency. In this study, the quantitative model is developed with MATLAB and Deep Learning Toolbox in MATLAB 2019b. Several steps are conducted to process the data:

1. Normalization: Scale inputs and targets into the range [0,1]

2. Data division: Divide data into three parts, training (70%), validation (15%), and testing (15%).

3. Optimization: Update the weights and bias according to the training loss.

4. Validation: Stop training when the validation loss has been relatively low.

5. Test: Test the trained model with a testing dataset.

To guarantee the generalization of our model, two different data sets, namely the training and validation set are utilized to determine the stop point and avoid overfitting.

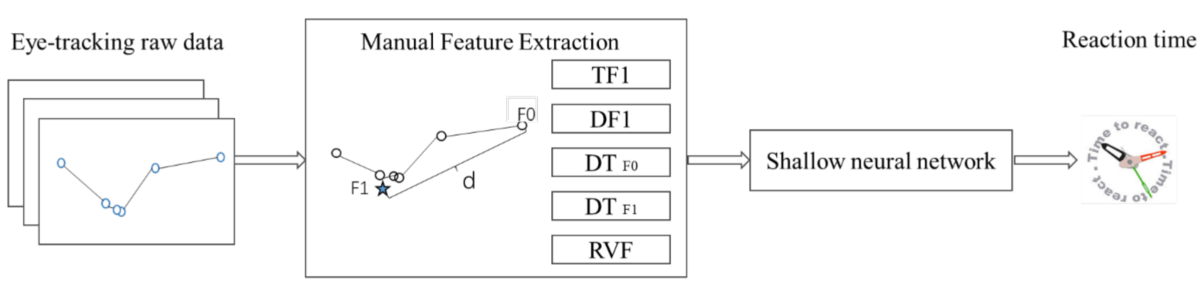

Figure 10: The process of quantitative model development

The 5-fold cross-validation was selected to calculate the performance of the predictions. Cross-validation splits the training sets into several sets. One set of the data is utilized to test the performance of the model trained by other remain data datasets. The 5-fold cross-validation can fully use the data and address the problem of insufficient samples of training data [64].

All parameters are set as default. For example, the maximum number of epochs to train is 1,000. The minimum performance gradient is 1e-6.

$$Z_1 = f(W_1 X + B_1) \tag{2}$$

The linear activation function is used in the output layer. W_1 is the weight matrix of the hidden nodes. B_1 is the bias vector of the hidden nodes. The weights and biases of the SANN are initialized using the normal distribution. The activation function used in the hidden layers is hyperbolic tangent:

$$\emptyset(z) = \frac{e^z - e^{-z}}{e^z + e^{-z}} \tag{3}$$

The dimension of the hidden layers is tested and determined to minimize the model loss, which is mean square error,

$$Loss = \frac{1}{n}\sum_{i=1}^{n}(R_{SANN} - R_{Exp})^2 \tag{4}$$

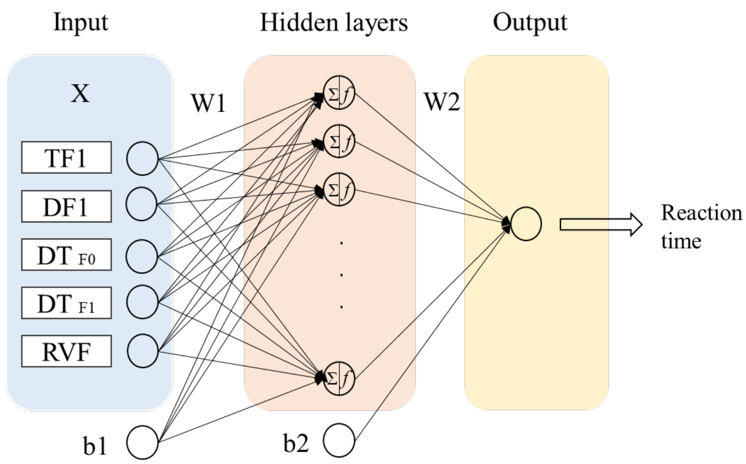

Figure 11: The structure of the shallow neural network

4. Demonstrating experiment

An experiment was conducted at Nanyang Technological University to collect data and demonstrate the proposed framework. This experiment complied with the American Psychological Association Code of Ethics and was approved by the Institutional Review Board of Nanyang Technological University (NTU-IRB), Singapore. The reference number is IRB-2018-04-007. All methods were carried out following the guidelines and regulations of NTU-IRB. 20 subjects were recruited. They were above 21 years old and signed the informed consent before participating in the experiment.

4.1 Experimental tasks

The two designed tasks were adopted to test vigilance in this case study. Both two tests have two sessions: the practice session and the test session. The practice session, which includes 10 trials, was given to reduce learning effects and to let the participant familiarizes themselves with the task. The test session includes 4 blocks of trials, with each block consisting of 27 trials. In each trial, the participants were required to fixate their eyes on the center of the display (a black dot) when there is no visual cue. One of the letters was presented randomly on the display. The visual cue is set to appear randomly between 800 – 1500 milliseconds to guarantee the occurrence of fixations and saccades.

To guarantee the changes in vigilance performance, a fatigue-inducing task is added, which is a simulation task of the vessel traffic service operations, as shown in Figure 12. Participants were required to monitor the videos of vessel traffic and record the name of vessels whose speed is larger than a pre-defined threshold. The vessel traffic videos are real traffic recordings of Gibraltar on 1 May 2016, Port of Rotterdam on 7 May 2016, and OSCL Jupiter agro on 14 Aug 2017. The vessel traffic monitoring task is monotonous and tedious, requiring attention, engagement, and information processing. The time given for each task is around one hour. According to our previous studies, at least 40-minutes monotonous task is required to induce human fatigue, which can significantly affect vigilance performance.

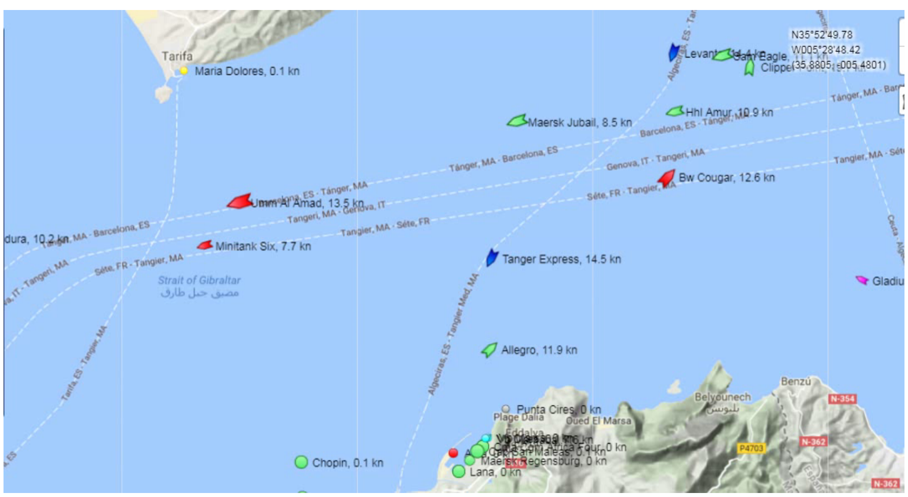

Figure 12: Snap of the vessel traffic video

The flow of the experiment started with the subject received a verbal briefing and signed the consent form in the brief session. The verbal brief describes the objectives, procedures, apparatus, and tasks. Followed by the verbal brief, the participants were instructed to calibrate the eye tracker. For each participant, the calibration has to be done at least once. On calibration, a red moving dot moves to 5 different positions with a “z” shape. The participants should track and fixate on the moving dot. After calibration, the participant was instructed to rate their fatigue score with the Samn-Perelli Mental Fatigue Scale and do the reaction time test. Upon the completion of the reaction time test, participants rate their fatigue level again. Then, the fatigue induce-task was presented. The mental fatigue was assessed before and after the participant completed the vigilance task. Next, the reaction time test and fatigue induce task would be done again before the last mental fatigue assessment and the last reaction time test. In total, each participant rated the fatigue scale six times, did six reaction time tests, and two fatigue-induction tasks. In general, this experiment lasts around 120 minutes.

In this study, we focused on human fatigue-induced vigilance performance impairment. Hence, the reaction time test only lasted six minutes to reduce the possibility of mind wandering. We recorded the fatigue level before and after the reaction time test to confirm that the reaction time test would not induce or recovery mental fatigue. Figure 13 shows the experimental procedures.

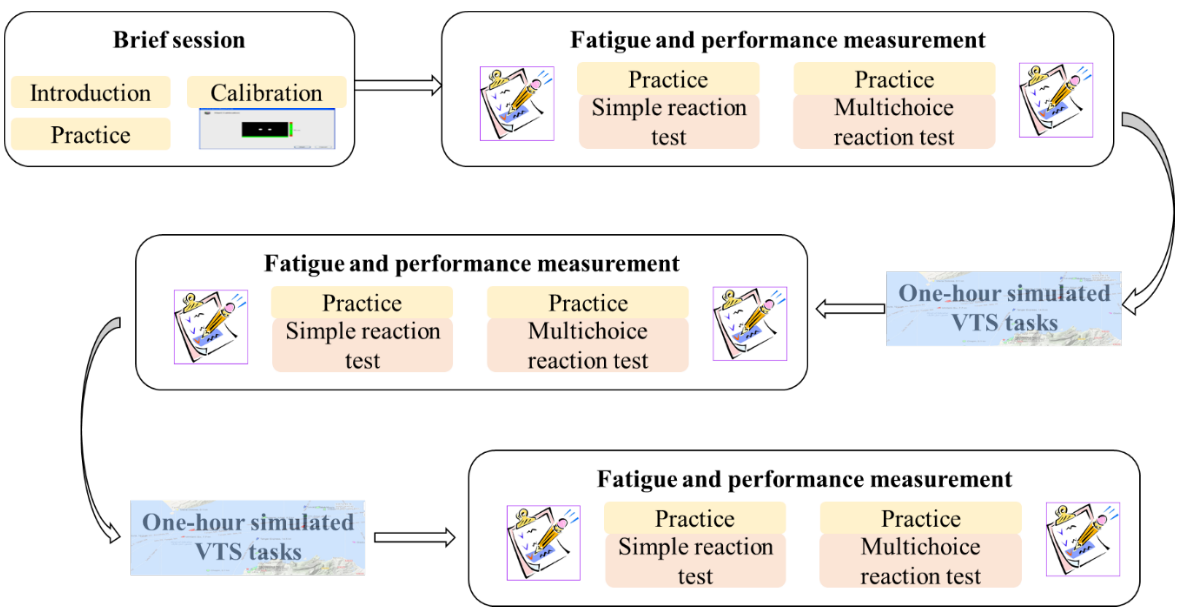

Figure 13: Experimental procedures

4.2 Data Collection

Due to the trace problem of the eye-tracker, the eye-tracking data normally suffer from quality problems. Hence, some collected eye-tracking data were deleted according to the procedures mentioned in our previous study [41]. In this study, only eye-tracking data collected from 14 subjects were utilized for statistical analysis. As mentioned in Section 4.1, subjects conducted 4*28=108 trials for both simple reaction time test and multiple-choice reaction time test. The eye-tracking data and vigilance performance data were collected during the test. The gaze pattern parameters and performance parameters were averaged across each trail. Table list the structured data.

The collected data were structured for spatial and temporal data analysis and AI-model development, as shown in Table 1. For spatial and temporal data analysis, all gaze pattern parameters and performance parameters were averaged across time-on-task to figure out how they change with time-on-task. For AI-model development, this study yields 108*3=324 data points for each subject on both simple reaction time test and multiple-choice reaction time test, as subjects conducted the tests three times.

Table 1: Sample of the structured data

Subject No	Trail No	Time on task	TF1	DF1	DT_{F0}	DT_{F1}	RVF	Response time	Hit rate
1	1	0	312.2222	298.1333	134.8333	40.42164	0.63	387.5111	1
1	2	0	405.8235	278.4706	134.1849	89.06549	0.74	532.9118	1
1	3	0	289.9783	214.913	143.5596	57.62113	0.62	385.6087	1
1	4	0	295.2083	197.5625	113.6245	32.22488	0.6	372.2708	1
1	5	0	221.7333	142.0333	119.0238	69.04201	0.78	313.9667	1

4.3 Spatial and temporal data analysis

4.3.1 Vigilance performance

All subjects conducted the reaction time test three times. Hence, a repeated measure Analysis of Variance (ANOVA) was performed to test the changes in response time and hit rate. Since the hit rate of the simple reaction time test is always 1, we did not conduct a statistical analysis on it. Both time-on-task and task complexity showed significant effects on response time (time-on-task: $F_{(2, 38)} = 8.27$, $p = 0.01$; task complexity: $F_{(1, 19)} = 9.73$, $p = 0.003$), where F value is the results of ANOVA test, it is a value on the F distribution, and p value is the probability of the test results would occur by random chance. According to Figure 14 (a), the response time of both two tests decreased with time. Similarly, time-on-task had a significant effect on hit rate ($F_{(2, 38)} = 4.596$, $p = 0.014$). The hit rate was decreased with time going on.

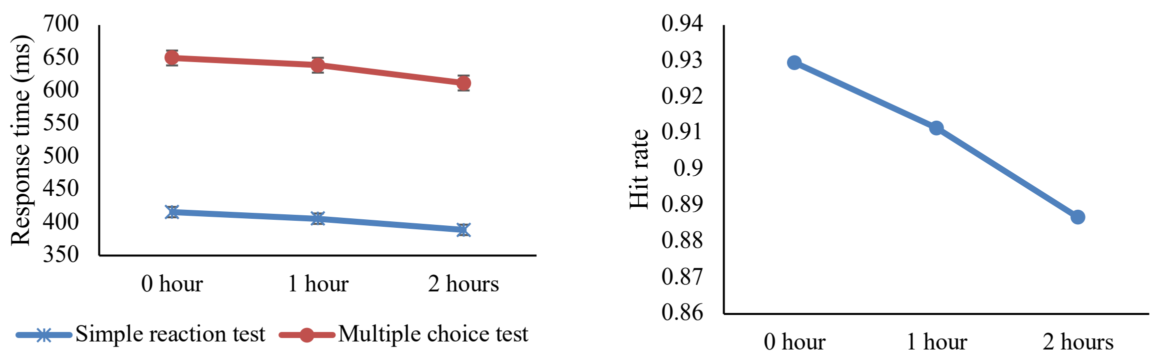

Figure 14: Response time across two reaction time tests and time (a); Hit rate of the multiple-choice test across time (b)

4.3.2 The temporal analysis

The task complexity had significant main effects on T_{F1} ($F_{(1, 13)} = 20.9$, $p = 0.0005$), while the effects of time on T_{F1} were not significant ($F_{(2, 26)} = 0.97$, $p = 0.39$) and the interaction effects on T_{F1} were not significant ($F_{(2, 26)} = 1.39$, $p = 0.27$). As shown in Figure 15(a), subjects spent much more time to arrive F1 in the multi-choice reaction time test (mean = 302.7ms) than in the simple reaction time test (mean = 371.6ms). For the duration of F1 (D_{F1}), the task complexity had significant effects on it ($F_{(1, 13)}$=20.6, p=0.0006), while the effects of time were not significant ($F_{(2, 26)}$=1.94, p=0.16). Comparing with the simple reaction time test (mean = 296ms), subjects needed much more time to understand the visual cues in the multi-choice reaction time test (mean D_{F1} = 363ms). It is worth noting that the interaction effects of time and task complexity on D_{F1} were significant ($F_{(2, 26)} = 5.42$, $p = 0.01$). With time going on, subjects presented much longer F1 in the complex multiple-choice test, while for the simple reaction time test, the D_{F1} was stable during the whole experiment.

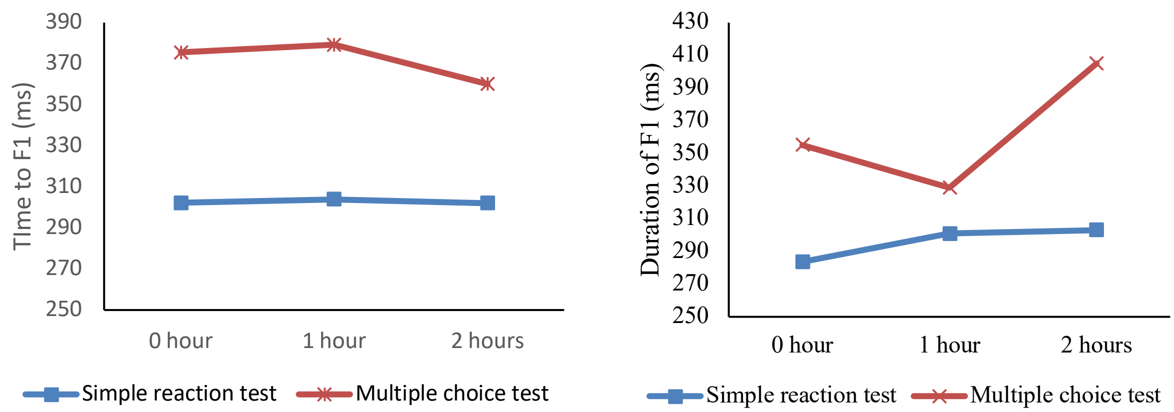

Figure 15: The time to F1 (TF1) across times and two tests (a); Duration of F1 (DF1) across times and two tests (b)

4.3.3 The spatial analysis

As shown in Figure 16 (a), the distance between F0 and target (DT_{F0}) stay unchanged across times and two tests. No significant result was observed in the main effects of time ($F_{(2, 26)}$=0.34, p=0.71) and task complexity $F_{(1, 13)}$=0.33, p=0.57, nor in their interaction effects ($F_{(2, 26)}$=1.39, p=0.27). The task complexity has significant effects on (DT_{F1}) ($F_{(1, 13)}$=5.58, p=0.034), while the effects of time ($F_{(2, 26)}$=1.68, p=0.2) and their interaction effects were not significant ($F_{(2, 26)}$=2.36, p=0.11), as shown in Figure 8 (b).

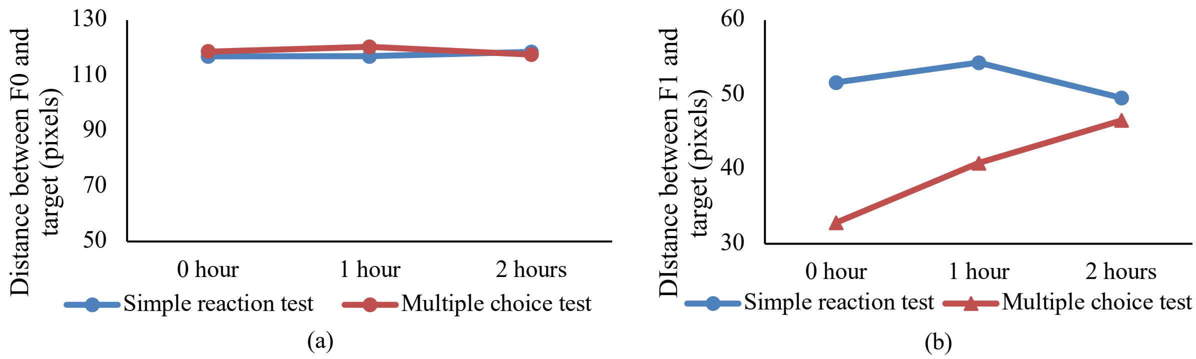

Figure 16: DT_{F0}: the distance between F0 and target (a); DT_{F1}: the distance between F1 and target (b)

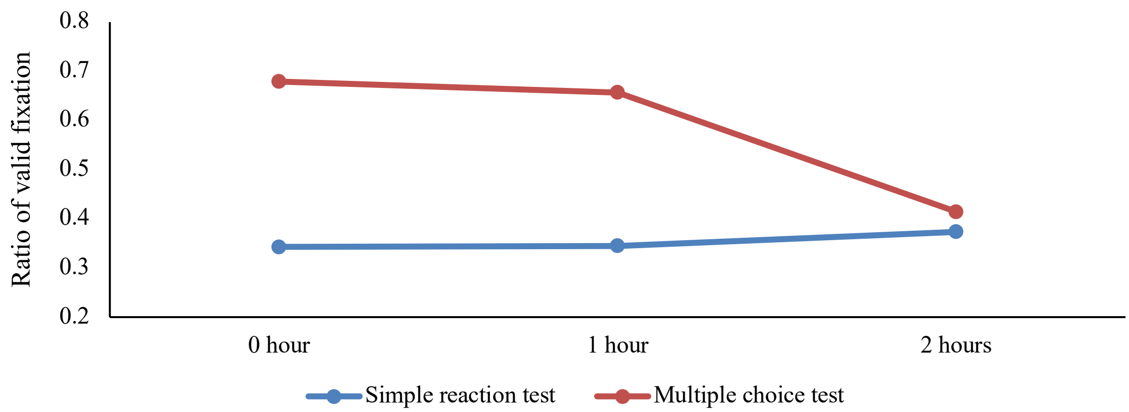

Figure 17: The ratio of valid fixation (RVF)

The task complexity had significant effects on RVF ($F_{(1, 13)} = 9.54$, $p = 0.0086$), while the effects of time on RVF were not significant ($F_{(2, 26)} = 1.32$, $p = 0.28$), as shown in Figure 17. Their interaction effects on RVF were significant ($F_{(2, 26)} = 3.52$, $p = 0.044$).

4.4 Shallow neural network-based model development

As discussed in Section 4.3, there are great associations between gaze patterns, time-on-task, and vigilance performance. Hence, it is expected that spatial and temporal gaze pattern analytics can be adopted for assessing vigilance. In this study, a regression neural network is developed to fit eye-tracking parameters with vigilance performance, namely reaction time. To measure the performance of regression methods, the mean squared error (MSE) and a goodness-of-fit measure between outputs and targets (R^2) are utilized [13]. Typically, we should minimize MSE and maximize R^2 for achieving good performance. In this study, MSE and R^2 are utilized as performance indicators of the developed SANN and other benchmarks.

A shallow artificial neural network (SANN) was developed via MATLAB deep learning toolbox. As mentioned in Mathworks, to develop the SANN, we run through the following steps, (1) create a network, (2) configure the network, (3) initialize weights and biases, (3) training network, and (4) verify the network.

In the first step, a shallow neural network was created with Matlab syntax "network." Then, in network configuration, the number of hidden layers was set to "1," considering the dimension of the input data was just five, as shown in Figure 11. A pilot experiment was conducted to investigate how the performance of SANN varied with the number of hidden nodes and training algorithms, as shown in Figure 18. To minimize MSE and maximize R, the Bayesian regularization was selected as the training algorithm, as it can resulting in relatively good generalization for small and noisy datasets. Though the scaled conjugate gradient and Levenberg-Marquardt algorithms require less memory and can save computational time, their performance is much worse than the Bayesian regularization, as shown in Table 2. It can be found that with hidden nodes of "10" and "15," a smaller MSE can be obtained, but the R^2 value was greatly reduced. Considering the computational efficiency, we determined to set the hidden nodes as "5."

In the step of initializing weights and biases, biases were initialized with 0 and weights were initialized with random numbers. Weights were not initialized with zero to avoid fitting the eye-tracking data with the linear regression. For each subject, 648 data points, including eye-tracking parameters and reaction time, were obtained. In this study, to establish a subject-independent method, data collected from all subjects were mixed and randomly classified into three parts, training (70%), validation (15%), and testing (15%). As mentioned in Section 3.3, all the manually extracted spatial and temporal eye-tracking parameters were utilized as inputs of the SANN, and the output was the reaction time. The SANN was trained with 75% of the collected data points. Its hyperparameters were tuned during training and verification.

Table 3 shows the testing results of applying the trained SANN on the randomly selected 15% test data. It was compared with linear regression to verify the proposed method, which is the most widely used quantitative model in traditional human factors studies. Moreover, the proposed method was compared with other machine learning methods, such as “decision tree”, “support vector machine”, and “bagged trees.” These benchmarks were initially developed with default hyperparameters, some of which were tuned later to achieve better performance. Specifically, the sigmoid kernel function was adopted in developing the support vector machine. These methods utilized the same input features, labels, and same training/testing data with the SANN. Table 3 presented the performance of all these methods. It can be found that the SANN performance is much better than the traditional machine learning methods in terms of MSE and R. Specifically, the SANN achieved the smallest MSE and largest R^2.

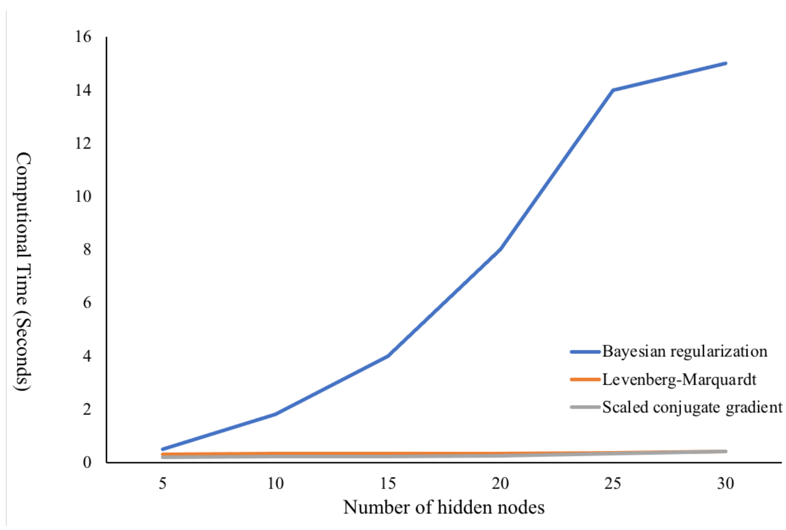

Figure 18: The training time across the number of hidden nodes

Table 2: The model performance across the number of hidden nodes and training algorithms

	MSE			R^2		
Number of hidden nodes	Levenberg-Marquardt	Bayesian regularizatio n	Scaled conjugate gradient	Levenberg-Marquardt	Bayesian regularizatio n	Scaled conjugate gradient
5	0.004	0.003	0.004	0.162	0.792	0.013
10	0.002	0.002	0.008	0.084	0.494	0.130
15	0.006	0.002	0.002	0.102	0.476	0.005
20	0.003	0.011	0.016	0.003	0.672	0.096
25	0.012	0.004	0.004	0.303	0.078	0.102
30	0.004	0.006	0.005	0.058	0.024	0.001

Table 3: The performance of traditional machine learning methods

	MSE	R^2
SANN	0.003	0.792
Linear Regression	0.0042	0.2
Decision Tree	0.0064	0.053
Support Vector Machine	0.0048	0.36
Bagged Trees	0.0052	0.1

5. Discussion and verification

5.1 Effectiveness of the spatial and temporal analytics

According to the spatial analysis of gaze movements, gazes were always fixated far away from the target in the simple reaction time test although the visual cue appeared. Specifically, gazes are located between the center dot and the target cue most of the time. The average DT_{F1} is 52 pixels, which is much larger than the radius of a fixation defined in this study. The average RVF is 0.35. Since no significant eye movements were observed, it can be concluded that only covert attention was paid in the simple reaction time test, and users can notice the stimulus, even the cues lying in the subjects' peripheral view. In other words, to acknowledge the presence of the visual cue, no significant gaze movements and valid fixations are needed. Compared with the simple reaction time test, the frequency of the subjects' gazes fixating on the targets increased in the multiple-choice reaction time test. Specifically, the average DT_{F1} is 40 and the average RVF is 0.58. The DT_{F0} is stable across time and task complexities. Hence, we can conclude that subjects followed our instruction to fixate in the center when there is no visual cue, and our analysis of all parameters is practical. In addition, much more overt attention was required in the multiple-choice reaction time test.

According to the mental process theory, subjects need to perceive, understand, and then react to the target. In this study, the simple reaction time test mainly requires perception and reaction, as the task is too simple to need information processing. We found that the time effects on T_{F1} and response time of the simple reaction time test were not significant. It can be concluded that the perception ability of subjects was not impaired by the impaired vigilance. Fixation duration is highly associated with information processing. We can find that D_{F1} of the simple reaction time test was not affected by vigilance levels, while D_{F1} of the multiple-choice test significantly increased under the fatigue condition. As mentioned in the spatial analysis of gaze movement, the simple reaction time test seems only requires perception and needs limited time on information processing. Hence, the stable D_{F1} of the simple reaction time test is reasonable. The increased D_{F1} of the multiple-choice test indicated that the time required for processing the cues' information was significantly increased with impaired vigilance levels. Considering that the response time decreased with time, the perception time kept still, and the information processing time increased with time, we can conclude that the decreases were caused by the decreased action time.

Decreased action time may result from learning effects and arousal effects. In our study, we have arranged a series of practices to make subjects to be familiar with the tests. Moreover, a post-analysis was conducted to test the human performance differences within each session. We divided each session into four parts and compared the hit rates across the four parts. No significant results were observed in this analysis. Hence, we concluded that the decreased action time was not caused by learning effects. For the arousal effects, as mentioned in [13], a lengthy period of work has multi aspects, such as impaired vigilance and emotional fatigue. These aspects are highly correlated and interact with each other. The emotional fatigue induced in this study may make subjects to be impatient and impetuous. Subjects were eager to finish the tasks as soon as possible, resulting in a shorter response time. Though participants achieved high performance in the response time, their hit rate decreased. The result of the decreased hit rate is in line with many studies [18, 38]. According to the temporal analysis of gaze movement, the psychomotor performance was not affected by impaired vigilance in this work. At the same time, the cognitive capacity was impaired. As a result, they achieved the shortest reaction time with the lowest hit rate after two hours of vessel traffic service operations.

5.2 Effectiveness of the proposed AI-based method

In this study, the SANN was developed to fit the spatial and temporal eye-tracking data with vigilance performance, namely reaction time. The proposed method was compared with other AI methods. As shown in Table 3, it is unexpected to find out that decision trees and bagged trees generated the poorest predictions. Since, the bagged tree algorithm ensembles a number of decision trees to generate the final result, it is reasonable that the bagged tree algorithm has a similar but better performance than the decision tree algorithm [44]. The problem lies in the weak correlation between predicted results and the observed data. The poor results may be caused by the limited dimension of the input data and a large number of training data. The decision tree is not suitable for continuous data prediction and may be overfitted. The performance of the linear regression is not so good, too. The results may be caused by the complex and dynamic relations between the eye-tracking parameters and vigilance performance [27]. In addition, there are generally great individual differences in eye-tracking parameters [41]. By extracting parameter from the spatial and temporal gaze movement analytics, the method is expected to be generalized and can be applied to other traffic controller's working environments with similar enterprise context and risks. It is hoped that this work provides practical guidances for achieving a more advanced AI enabled human-computer interfaces and a consolidated assessment approach.

The proposed method achieved relatively high performance with an MSE of 0.003 and R^2 of 0.792. The performance is much better than most of the existing eye-tracking-based vigilance models. For example, a study published in 2019 [13] combined the context data and traditional eye-tracking parameters, such as fixation duration, fixation number, saccade amplitude to predict the reaction time. It achieved an RMSE of 0.119, namely an MSE of 0.014. A recent study published in 2021 [65] utilized the linear regression model and the saccade number to predict the driver reaction time to handover request, and it reported R^2 of 0.399, which is much poorer than the performance of the proposed method. The proposed method can achieve better performance due to several reasons. On the one hand, most of the existing studies utilize parameters that are averaged from a long period of recording. These parameters highly depend on the context [66]. As a result, models developed based on these eye-tracking parameters have limited robustness. On the other hand, this study proposes to analyze the eye-tracking data with SANN, which is dynamic to significant variance and has good fault tolerance [67].

6. Conclusion

Nowadays, high levels of vigilance are required in many fields, such as vessel traffic service, high-speed train, and aviation. Hence, an objective way to non-intrusively assess vigilance is urgently needed. This study proposes a four-phase framework to achieve a non-intrusive assessment of vigilance based on gaze movement pattern analytics. The gaze patterns were analyzed temporally and spatially to figure out how vigilance stages affect human performance in terms of perception, comprehension, and action. We found that the psychomotor performance was not affected under impaired vigilance, while the cognitive capacity was impaired under a moderate-high level of impaired vigilance. The temporal analysis of our gaze movements explained the impaired vigilance-motor performance paradox. A shallow neural network model has been developed to predict the reaction time based on the proposed spatial and temporal gaze movement parameters. With the same input features, the proposed model achieved relatively better performance than other machine learning methods. In addition, the proposed method outperformed several recently published eye-tracking-based vigilance models in terms of the mean squared error and a goodness-of-fit measure between outputs and targets.

The limitations to our study stemmed from the experiment design and laboratory tests. Instead of real-word tasks, two general and basic tests were designed to measure vigilance. Even the tests simulated the common hazard identification tasks, the gaze patterns of the designed tasks should be different from the real-world tasks. In future works, some practical scenarios should be applied for vigilance assessment. Furthermore, manually extracted eye-tracking parameters were utilized as the input features of a shallow neural network. To reduce the dependence on manual work, the deep neural network can be applied to deal with the raw eye-tracking data in future works.

Acknowledgments:

This research was partially supported by the Xi'an Jiaotong University [grant number: 7121192301], National Natural Science Foundation of China [grant number 72174168], and The Hong Kong Polytechnic University [grant number A0038827]. We appreciate the support from the Human factor and design lab of MAE at Nanyang Technological University.

Reference

1. Parasuraman, R., J.S. Warm, and W.N. Dember, *Vigilance: Taxonomy and utility*, in *Ergonomics and human factors*. 1987, Springer. p. 11-32.
2. Lee, C.-H., C.-H. Chen, C. Lin, F. Li, and X. Zhao, *Developing a Quick Response Product Configuration System under Industry 4.0 Based on Customer Requirement Modelling and Optimization Method.* Applied Sciences, 2019. **9**(23): p. 5004.
3. Golec, A. and E. Kahya, *A fuzzy model for competency-based employee evaluation and selection.* Computers & Industrial Engineering, 2007. **52**(1): p. 143-161.
4. Pacaux-Lemoine, M.-P., D. Trentesaux, G.Z. Rey, and P. Millot, *Designing intelligent manufacturing systems through Human-Machine Cooperation principles: A human-centered approach.* Computers & Industrial Engineering, 2017. **111**: p. 581-595.
5. Helton, W.S., G. Matthews, and J.S. Warm, *Stress state mediation between environmental variables and performance: The case of noise and vigilance.* Acta psychologica, 2009. **130**(3): p. 204-213.
6. Jarosch, O., H. Bellem, and K. Bengler, *Effects of Task-Induced Fatigue in Prolonged Conditional Automated Driving.* Human factors, 2019: p. 0018720818816226.
7. Trapsilawati, F., X. Qu, C.D. Wickens, and C.-H. Chen, *Human factors assessment of conflict resolution aid reliability and time pressure in future air traffic control.* Ergonomics, 2015. **58**(6): p. 897-908.
8. Trapsilawati, F., C.D. Wickens, X. Qu, and C.-H. Chen, *Benefits of imperfect conflict resolution advisory aids for future air traffic control.* Human factors, 2016. **58**(7): p. 1007-1019.
9. Rasmussen, M. and K. Laumann, *The evaluation of fatigue as a performance shaping factor in the Petro-HRA method.* Reliability Engineering & System Safety, 2020. **194**: p. 106187.
10. Casner, S.M. and J.W. Schooler, *Vigilance impossible: Diligence, distraction, and daydreaming all lead to failures in a practical monitoring task.* Consciousness and cognition, 2015. **35**: p. 33-41.
11. Chuang, C.-H., C.-S. Huang, L.-W. Ko, and C.-T. Lin, *An EEG-based perceptual function integration network for application to drowsy driving.* Knowledge-Based Systems, 2015. **80**: p. 143-152.
12. Bergasa, L.M., J. Nuevo, M.A. Sotelo, R. Barea, and M.E. Lopez, *Real-time system for monitoring driver vigilance.* IEEE Transactions on Intelligent Transportation Systems, 2006. **7**(1): p. 63-77.
13. Li, F., C.-H. Lee, C.-H. Chen, and L.P. Khoo, *Hybrid data-driven vigilance model in traffic control center using eye-tracking data and context data.* Advanced Engineering Informatics, 2019. **42**: p. 100940.
14. Zandi, A.S., A. Quddus, L. Prest, and F.J. Comeau, *Non-Intrusive Detection of Drowsy Driving Based on Eye Tracking Data.* Transportation Research Record, 2019: p. 0361198119847985.
15. An, P., Z. Yuan, J. Zhao, X. Jiang, and B. Du, *An effective multi-model fusion method for EEG-based sleep stage classification.* Knowledge-Based Systems, 2021. **219**: p. 106890.
16. Yadav, S.K., K. Tiwari, H.M. Pandey, and S.A. Akbar, *A review of multimodal human activity recognition with special emphasis on classification, applications, challenges and future directions.* Knowledge-Based Systems, 2021: p. 106970.

17. Zheng, W.-L. and B.-L. Lu, *A multimodal approach to estimating vigilance using EEG and forehead EOG.* Journal of Neural Engineering, 2017. **14**(2): p. 026017.
18. McIntire, L.K., R.A. McKinley, C. Goodyear, and J.P. McIntire, *Detection of vigilance performance using eye blinks.* Applied ergonomics, 2014. **45**(2): p. 354-362.
19. Behrend, J. and F. Dehais, *How role assignment impacts decision-making in high-risk environments: Evidence from eye-tracking in aviation.* Safety science, 2020. **127**: p. 104738.
20. Syrjämäki, A.H., P. Isokoski, V. Surakka, T.P. Pasanen, and J.K. Hietanen, *Eye contact in virtual reality–A psychophysiological study.* Computers in Human Behavior, 2020. **112**: p. 106454.
21. Wang, X., L. Lin, M. Han, and J.M. Spector, *Impacts of cues on learning: Using eye-tracking technologies to examine the functions and designs of added cues in short instructional videos.* Computers in Human Behavior, 2020. **107**: p. 106279.
22. Ku, H.-Y., C.A. Harter, P.-L. Liu, L. Thompson, and Y.-C. Cheng, *The effects of individually personalized computer-based instructional program on solving mathematics problems.* Computers in human behavior, 2007. **23**(3): p. 1195-1210.
23. Li, F., C.-H. Lee, S. Feng, A. Trappey, and F. Gilani. *Prospective on Eye-Tracking-based Studies in Immersive Virtual Reality*. in *2021 IEEE 24th International Conference on Computer Supported Cooperative Work in Design (CSCWD)*. 2021. IEEE.
24. Robus, C.M., C.J. Hand, R. Filik, and M. Pitchford, *Investigating effects of emoji on neutral narrative text: Evidence from eye movements and perceived emotional valence.* Computers in Human Behavior, 2020. **109**: p. 106361.
25. Zhang, X., S. Mahadevan, N. Lau, and M.B. Weinger, *Multi-source information fusion to assess control room operator performance.* Reliability Engineering & System Safety, 2020. **194**: p. 106287.
26. Tan, S.Y., C.H. Chen, S.W. Lye, and F. Li. *Task Measures for Air Traffic Display Operations*. in *International Conference on Intelligent Human Systems Integration*. 2020. Springer.
27. Renata, V., F. Li, C.-H. Lee, and C.-H. Chen. *Investigation on the Correlation between Eye Movement and Reaction Time under Mental Fatigue Influence*. in *In Proc. of the 17th Cyberworlds International Conference (CW 2018)*. 2018. Singapore: IEEE.
28. Iqbal, S.T. and B.P. Bailey. *Using eye gaze patterns to identify user tasks*. in *The Grace Hopper Celebration of Women in Computing*. 2004.
29. Conati, C. and C. Merten, *Eye-tracking for user modeling in exploratory learning environments: An empirical evaluation.* Knowledge-Based Systems, 2007. **20**(6): p. 557-574.
30. Chen, N.T. and P.J. Clarke, *Gaze-based assessments of vigilance and avoidance in social anxiety: a review.* Current psychiatry reports, 2017. **19**(9): p. 1-9.
31. dos Santos, B.S., M.T.A. Steiner, A.T. Fenerich, and R.H.P. Lima, *Data mining and machine learning techniques applied to public health problems: A bibliometric analysis from 2009 to 2018.* Computers & Industrial Engineering, 2019. **138**: p. 106120.
32. MassirisFernández, M., J.Á. Fernández, J.M. Bajo, and C.A. Delrieux, *Ergonomic risk assessment based on computer vision and machine learning.* Computers & Industrial Engineering, 2020. **149**: p. 106816.

33. Lin, C.-T., K.-C. Huang, C.-F. Chao, J.-A. Chen, T.-W. Chiu, L.-W. Ko, and T.-P. Jung, *Tonic and phasic EEG and behavioral changes induced by arousing feedback.* NeuroImage, 2010. **52**(2): p. 633-642.
34. Zheng, W.-L., K. Gao, G. Li, W. Liu, C. Liu, J.-Q. Liu, G. Wang, and B.-L. Lu, *Vigilance estimation using a wearable EOG device in real driving environment.* IEEE transactions on intelligent transportation systems, 2019. **21**(1): p. 170-184.
35. Sebastiani, M., G. Di Flumeri, P. Aricò, N. Sciaraffa, F. Babiloni, and G. Borghini, *Neurophysiological vigilance characterisation and assessment: laboratory and realistic validations involving professional air traffic controllers.* Brain sciences, 2020. **10**(1): p. 48.
36. Imperatori, L.S., J. Cataldi, M. Betta, E. Ricciardi, R.A. Ince, F. Siclari, and G. Bernardi, *Cross-participant prediction of vigilance stages through the combined use of wPLI and wSMI EEG functional connectivity metrics.* Sleep, 2021. **44**(5): p. zsaa247.
37. Guo, Z., Y. Pan, G. Zhao, S. Cao, and J. Zhang, *Detection of driver vigilance level using EEG signals and driving contexts.* IEEE Transactions on Reliability, 2017. **67**(1): p. 370-380.
38. Abe, T., T. Nonomura, Y. Komada, S. Asaoka, T. Sasai, A. Ueno, and Y. Inoue, *Detecting deteriorated vigilance using percentage of eyelid closure time during behavioral maintenance of wakefulness tests.* International Journal of Psychophysiology, 2011. **82**(3): p. 269-274.
39. Cui, J., Z. Lan, Y. Liu, R. Li, F. Li, O. Sourina, and W. Müller-Wittig, *A compact and interpretable convolutional neural network for cross-subject driver drowsiness detection from single-channel EEG.* Methods, 2021.
40. Wang, C.-Y., M.-J. Tsai, and C.-C. Tsai, *Predicting cognitive structures and information processing modes by eye-tracking when reading controversial reports about socio-scientific issues.* Computers in Human Behavior, 2020. **112**: p. 106471.
41. Li, F., C.-H. Chen, G. Xu, and L.-P. Khoo, *Hierarchical Eye-Tracking Data Analytics for Human Fatigue Detection at a Traffic Control Center.* IEEE Transactions on Human-Machine Systems, 2020. **50**(5): p. 465-474.
42. Han, L., R. Sun, Y. Sun, F. Gao, D. Xie, and M. Jou, *Reactions and gender differences to online pictures of covered sexual organs among heterosexual young adults—Studies based on behavior, eye movement and ERP.* Computers in Human Behavior, 2020. **111**: p. 106425.
43. Findlay, J.M. and I. Gilchrist, *Visual selection, covert attention and eye movements.* Active vision: The psychology of looking & seeing, 2003: p. 35-54.
44. Li, F., C.-H. Chen, G. Xu, L.P. Khoo, and Y. Liu, *Proactive mental fatigue detection of traffic control operators using bagged trees and gaze-bin analysis.* Advanced Engineering Informatics, 2019. **42**: p. 100987.
45. Waite, S., S. Kolla, J. Jeudy, A. Legasto, S.L. Macknik, S. Martinez-Conde, E.A. Krupinski, and D.L. Reede, *Tired in the Reading Room: The Influence of Fatigue in Radiology.* Journal of the American College of Radiology, 2016.
46. Schleicher, R., N. Galley, S. Briest, and L. Galley, *Blinks and saccades as indicators of fatigue in sleepiness warnings: looking tired?* Ergonomics, 2008. **51**(7): p. 982-1010.
47. Di Stasi, L.L., R. Renner, A. Catena, J.J. Cañas, B.M. Velichkovsky, and S. Pannasch, *Towards a driver fatigue test based on the saccadic main sequence: A partial validation by subjective report data.* Transportation research part C: emerging technologies, 2012. **21**(1): p. 122-133.

48. Finke, C., L.M. Pech, C. Sömmer, J. Schlichting, S. Stricker, M. Endres, F. Ostendorf, C.J. Ploner, A.U. Brandt, and F. Paul, *Dynamics of saccade parameters in multiple sclerosis patients with fatigue.* Journal of neurology, 2012. **259**(12): p. 2656-2663.
49. Yang, Y., M. McDonald, and P. Zheng, *Can drivers' eye movements be used to monitor their performance? A case study.* IET Intelligent Transport Systems, 2012. **6**(4): p. 444-452.
50. Zhang, J.-Y., S.-L. Liu, Q.-M. Feng, J.-Q. Gao, and Q. Zhang, *Correlative Evaluation of Mental and Physical Workload of Laparoscopic Surgeons Based on Surface Electromyography and Eye-tracking Signals.* Scientific reports, 2017. **7**(1): p. 11095.
51. Yamada, Y. and M. Kobayashi. *Detecting mental fatigue from eye-tracking data gathered while watching video*. in *Conference on Artificial Intelligence in Medicine in Europe*. 2017. Springer.
52. Appel, T., C. Scharinger, P. Gerjets, and E. Kasneci. *Cross-subject workload classification using pupil-related measures*. in *Proceedings of the 2018 ACM Symposium on Eye Tracking Research & Applications*. 2018. ACM.
53. Kim, D., S. Choi, J. Choi, H. Shin, and K. Sohn. *Visual fatigue monitoring system based on eye-movement and eye-blink detection*. in *IS&T/SPIE Electronic Imaging*. 2011. International Society for Optics and Photonics.
54. Hirvonen, K., S. Puttonen, K. Gould, J. Korpela, V.F. Koefoed, and K. Müller, *Improving the saccade peak velocity measurement for detecting fatigue.* Journal of neuroscience methods, 2010. **187**(2): p. 199-206.
55. Olsen, A., *The Tobii I-VT fixation filter.* Tobii Technology, 2012.
56. Kim, M., B.H. Kim, and S. Jo, *Quantitative evaluation of a low-cost noninvasive hybrid interface based on EEG and eye movement.* IEEE transactions on neural systems and rehabilitation engineering, 2015. **23**(2): p. 159-168.
57. Samn, S. and L. Perelli, *Estimating Aircrew Fatigue: A Technique with Implications to Airlift Operations. USAF School of Aerospace Medicine, Brooks AFB*. 1982, TX. Technical Report No. SAM-TR-82-21.
58. Ferguson, S.A., G.M. Paech, C. Sargent, D. Darwent, D.J. Kennaway, and G.D. Roach, *The influence of circadian time and sleep dose on subjective fatigue ratings.* Accident Analysis & Prevention, 2012. **45**: p. 50-54.
59. Danna-Dos-Santos, A., S. Mohapatra, M. Santos, and A.M. Degani, *Long-term effects of mild traumatic brain injuries to oculomotor tracking performances and reaction times to simple environmental stimuli.* Scientific reports, 2018. **8**(1): p. 4583.
60. Shattuck, N.L. and P. Matsangas, *Psychomotor vigilance performance predicted by E pworth S leepiness S cale scores in an operational setting with the United States N avy.* Journal of sleep research, 2015. **24**(2): p. 174-180.
61. Phipps-Nelson, J., J.R. Redman, D.-J. Dijk, and S.M. Rajaratnam, *Daytime exposure to bright light, as compared to dim light, decreases sleepiness and improves psychomotor vigilance performance.* Sleep, 2003. **26**(6): p. 695-700.
62. Manor, B.R. and E. Gordon, *Defining the temporal threshold for ocular fixation in free-viewing visuocognitive tasks.* Journal of neuroscience methods, 2003. **128**(1-2): p. 85-93.
63. Jacob, R.J. and K.S. Karn, *Eye tracking in human-computer interaction and usability research: Ready to deliver the promises*, in *The mind's eye*. 2003, Elsevier. p. 573-605.
64. developers, s.-l. *Cross-validation: evaluating estimator performance*. 2019 [cited 2020 14/3/2020]; Available from: https://scikit-learn.org/stable/modules/cross_validation.html.

65. Wu, Y., K. Kihara, Y. Takeda, T. Sato, M. Akamatsu, S. Kitazaki, K. Nakagawa, K. Yamada, H. Oka, and S. Kameyama, *Eye movements predict driver reaction time to takeover request in automated driving: A real-vehicle study.* Transportation research part F: traffic psychology and behaviour, 2021. **81**: p. 355-363.
66. Pokojski, J. *Evaluation of Humanoid Robot Design Based on Global Eye-Tracking Metrics*. in *Transdisciplinary Engineering for Complex Socio-technical Systems–Real-life Applications: Proceedings of the 27th ISTE International Conference on Transdisciplinary Engineering, July 1–July 10, 2020*. 2020. IOS Press.
67. Duan, J. and F. Li, *Transient heat transfer analysis of phase change material melting in metal foam by experimental study and artificial neural network.* Journal of Energy Storage, 2021. **33**: p. 102160.