Proactive Mental Fatigue Detection of Traffic Control Operators using Bagged Trees and Gaze-bin analysis

Fan Li [1, 2], Chun-Hsien Chen [1], Gangyan Xu [3, 2*], Li Pheng Khoo [1], Yisi Liu[4]

[1]*School of Mechanical and Aerospace Engineering, Nanyang Technological University, Singapore*

[2]*Maritime Institute, Nanyang Technological University, Singapore*

[3]*School of Architecture, Harbin Institute of Technology, Shenzhen, China*

[4]*FraunhoferIDM@NTU Center, Singapore*

Abstract: Most of existing eye movement-based fatigue detectors utilize statistical analysis of fixations, saccades and blinks as inputs. These parameters require long recording time and depend heavily on eye trackers. As a result, they cannot timely and effectively discriminate fatigue. In an effort to facilitate proactive detection of mental fatigue, we introduced a novel fatigue indicator, named gaze-bin analysis. Instead of identifying events from eye tracking data, the novel fatigue indicator simply presents the eye tracking data with histograms. We developed an innovative method which engaged gaze-bin analysis as inputs of Semi-supervised Bagged trees. The approach could alleviate the burden of manual label, eliminate the problem of overlapped data, as well as improve the performance of fatigue detection model. It was demonstrated by a case study in a vessel traffic service center. The results showed that the approach could achieve an excellent accuracy of 89% which outperformed other methods. In general, this work paved an alternative way to detect mental fatigue as well as enabled the application of a low sampling rate eye tracker in traffic control center.

Keywords: Human fatigue; bagged tree; eye movement; bin analysis; time window

1. Introduction

Traffic Control (TC) refers to surveil, control and manage traffic via monitoring real-time traffic data and providing instructions or advice to traffic operators [1-3]. It aims at improving traffic flow as well as promoting fluent and safe traffic. Currently, it has been implemented in almost all transport modes such as air traffic management, freeway traffic management, vessel traffic service and railway traffic control. Its main operations include 24-hour passive monitoring [4-6]. The passive monitroing could easily induce high possibility of mental fatigue [7, 8], leading to disastrous consequences in public safety [9, 10]. Recently, mental fatigue is regarded among the top 10 safety issue, causing around 20% of accidents in all modes of transport [7].

Potentially, the risk of mental fatigue can be reduced by proactively monitoring fatigued Traffic Control Operators (TCOs) using eye movements. Eye movements have long been found to be applicable indicators of mental fatigue. Moreover, with technology improvement, eye movements currently could be remotely tracked by contactless eye trackers [11, 12]. Hence, eye movements can be measured continually during extended cognitive tasks. Changes in eye movements can be promptly captured to understand the interactions of human interface. Hence, eye movement-based fatigue detectors have lately received great attention [13-16]. Researchers have tried to use eye-tracking data to detect real-time human fatigue in watching video, driving, surgical operating and airline operating [16-19].

Existing methods in eye movement-based mental fatigue detection focused on using long-time recorded eye movements to generate descriptive fatigue indicators, such as mean, standard deviation and median of fixations and saccades [18]. These indicators impeded warning fatigued subjects timely. Moreover, these indicators depend heavily on the selection of parsing methods and the quality of eye-tracking data [21]. As a result, there are many contradictions results in studies of eye movement-based mental fatigue detection. For example, though many studies indicated that saccadic parameters vary with fatigue level [27, 33, 34], Saito [35] did not find any significant quantitative changes in saccadic eye movement in five hours of eye-tracking tasks.

Hence, instead of focusing on fixations and saccades, we proposed to capture dynamic information from short-period eye-tracking data using bin analysis. Bin analysis refers to counting how many values fall into a specified interval. It could accurately present the distribution of numerical data and capture dynamic features from short-period data [25]. By extending bin analysis to eye-tracking data, we introduced a novel problem for fatigue study namely using gaze-bin analysis of short-period data to develop fatigue detection model.

There are several challenges in addressing this problem. First, by splitting long recorded eye-tracking data into short-period data, we would introduce overlapped data. The overlapped data could induce fake performance of fatigue detection model. Second, excessive efforts required for labeling short-period eye-tracking data. Emerging studies in eye movement-based fatigue detection adopts manually labeling eye tracking data with subjective fatigue scales. However, in this work, the numerous short-period data require excessive efforts in manually labeling them with fatigue levels. Third, the performance of eye movement-based human fatigue detection impaired by the existence of classification noise, which is caused by the fairly large variance in eye movements [16]. The commonly used method is support vector machine. Nevertheless, its performance in eye movement-based human fatigue detection is still far from satisfactory. As a result, the existing eye-tracking fatigue detectors could produce a great amount of false alarms [23], inducing "wolf" effects. Forth, how to determine the period of time-window? In existing studies, the time window used for eye-movement-based human fatigue detection ranges from 8 seconds to 30 minutes. It is believed that the period of time window would have great effects on the performance of human fatigue models, such as detection accuracy and delayed time [22]. However, the appropriate time window received limited investigation.

In order to deal with these challenges and achieve the aim of proactively and non-invasively monitoring mental fatigue of TCOs, we proposed a method by extending Bagged trees with semi-supervised training to gaze-bin analysis. Bagged trees adopts the concept of assembling multiple decision trees. It could perform well in analyzing data with substantial classification noise [24]. Semi-supervised training enable us to use both labeled and unlabeled training data. We conducted a case study in a vessel traffic service center to illustrate the proposed method. The results showed that our method dominated other methods, and achieved an accuracy of 89% using 10-second eye movement data.

The paper is organized as follows. In section 2, as the limited studies in TC, the existing eye movement-based detectors of human fatigue in other fields were reviewed to figure out the methods and parameters that are commonly used. In addition, the development of Bagged trees and bin analysis are described. In Section 3, the novel problem of using gaze-bin analysis to discriminate human fatigue is stated. Section 4 describes the innovate approach for detecting human fatigue using gaze-bin analysis. The case study in vessel traffic service and the performance of the proposed approach are presented in Section 5. In addition, Section 5 presents the effects of model characterizes as well as the comparison results of the proposed method with other classic methods. Section 6 summarizes the contributions and limitations of this work.

2. Related Works

This section provides a review on eye movement-based human fatigue detection from two aspects: eye movement-based indicators and detection algorithms. In addition, the development and applications of Bagged Trees and bin analysis are examined to provide theoretic background for following sections.

2.1 Eye Movement-based Human Fatigue Detection

As early as 1996 [26], researchers had found that eye movement parameters could be potential indicators of human fatigue. In general, eye movement parameters including blinks, Percent Eye Closure (PERCLOS), saccades and fixations. Blinks related parameters are the most commonly utilized indicators [19, 20, 27]. In fatigued state, operators usually shows high rate of blink. Du et al. [28] indicated the possibility of using blink rate variance to discriminate human fatigue. Jin et al. [29], Azim et al. [30], and Ahlstrom et al. [20] investigated into detecting fatigue using blink frequency. Besides blinks, Percent Eye Closure (PERCLOS) has been widely used to detect human fatigue, too [29-31]. PERCLOS commonly increases with the increase of human fatigue. Hartley et al. [23] indicated that PERCLOS is among the best ocular measures for assessing fatigue. However, both blink related parameters and PERCLOS are of limited value [32]. The high rate of blink and large PERCLOS leads to visual information loss and the high possibility of human errors. Hence, using them to detect human fatigue would be too late to reduce the risk of human fatigue. Other commonly used parameters are saccadic parameters [27, 33, 34]. Finke et al. [33] indicated that saccadic parameters such as saccade amplitude, saccade rate and saccade duration, especially saccade peak velocity were reliable indicators of human fatigue. Di Stasi et al. found that the saccade peak velocity seems to greatly decrease with the increase of human fatigue [32]. All the parameters mentioned above depend heavily on the paring methods and the eye tracker [21]. For example, only the eye tracker with high sampling frequency could record saccade peak velocity. These drawbacks made it challenging to realize eye movement-based human fatigue detection using fixations and saccades parameters.

Currently, emerging studies found that it is possible to connect eye movement data to human fatigue by using machine learning techniques [19, 29, 30, 36, 37]. Yamada and Kobayashi [19] adopted support vector machine to detect mental fatigue. They recorded eye movements of participants who were watching video. Then they engaged gaze allocations as inputs of support vector machine to measure fatigue. Zhu et al. [36] applied convolutional neural networks in fatigue detection using blinks, slow eye movement and rapid eye movement. In general, the most commonly used technique is support vector machine, which was first proposed in 1995 [38]. It has shown good performance in detecting and recognition of text, speech and even credits [39, 40] and has been found to be a suitable method for measuring the cognitive states of humans. Nevertheless, the performance of support vector machine in eye movement-based human fatigue detection is still far from satisfactory (around 70%) [19]. We believed that the result may be caused by overlook of eye movement dynamic information and large classification noise. Hence, we proposed to address these problems by using the following theories.

2.2 Bin analysis

Data binning refers to segmenting data into several intervals and replacing values within one interval with a representative value [41]. The interval is called as "bin". This is a data pre-processing technique used to reduce the number of values to analyze [42]. It has been widely used in many fields such as visual diagnosis, genome architecture and metabolomics data analysis [41-43]. In general, the representative values of a give interval is the central value. In this study, we proposed to use the probability of the values fallen with a given interval as the presentative value. In this way, a histogram can be established. It could accurately represent the distribution of continuous data [25] with P=($p_1, p_2, \dots, p_N$). n refers to the number of the given interval. p_n refers to the probability of the values fallen within the *n*th interval. Entropy which indicates the state of the system can be calculated following the definition of Shannon et al. [44].

2.3 The bagged trees

Bagged predictor was first proposed in 1996 [24]. It is a method for manipulating training data to generate multiple predictors and aggregating results of multiple predictors. It has been successfully applied to construction material classification and diesel-electric locomotive train [45, 46]. Bagged trees approach attempts to assemble multiple trees so as to improve the model performance [24]. Assuming that the training data has *N* instances and each of these instances is labeled with one of *K* classes. A learning system, such as the decision-tree algorithm, C4.5, can be used to construct a predictor by taking a bootstrap sample from the training data. The size of the bootstrap sample is same with the size of the training data (*N*), but some instances are duplicated and some are removed. Multiple predictors can be constructed by generating repeatedly and training the bootstrap sample from the original training data. In general, the number of repetitions (*T*) can be fixed or be determined by cross-validation [47]. To classify an instance, *x*, all predictors constructed in *T* trails would indicate the class of the instance *x*. The instance *x* will be classified as Class *k* which has most votes from these predictors [48].

This concept of assembling multiple trees is not new and both the boosted trees and random forest make predictions based on multiple trees [49]. Comparing with the boosted trees and random forest, the bagged trees approach can perform much better in analyzing data with substantial classification noise [49]. Breiman [24] indicated that '*If perturbing the learning set can cause significant changes in the predictor constructed, then bagging can improve accuracy*'. This advantage makes it very suitable for eye movement-based human fatigue detection [19].

3. Gaze-bin Analysis-based Indicators of Human Fatigue

In this section, we first introduce the definition of gaze-bin analysis. Based on this definition, we give a description of the research problem waiting to be addressed in this work.

3.1 Gaze-bin Analysis

Given a series of eye tracking data, extracting parameters aims to identifying indicator of human fatigue. As a fundamental problem of eye movement-based human fatigue detection, eye parameters have been investigated in many studies [15, 16, 18]. Figure 1(a) shows the basic idea of these existing parameters. Generally, previous studies focused on identify fixations and saccades first and then generate statistical parameters. Basically, they transform the eye tracking data $\Re = \{G, T\}, G = \{X, Y\}$ to $\Re = \{F, S\}$, where G refers to the x and y coordinates set of gaze points, T is the set of time stamp, F is the set of fixations, and S is the set of saccades.

Different from existing studies, we investigate a novel research problem: gaze-bin analysis. We aim to capture the dynamic velocities of short-period eye-tracking data. The gaze velocity of each gaze point could be calculated by determining the difference between the central of two points and multiply by the sampling frequency (f) of the eye tracker .

$$v_t = f \times \sqrt[2]{(x_t - x_{t-1})^2 + (y_t - y_{t-1})^2} \tag{1}$$

$$t \in T, x \in X, y \in Y$$

Hence, we presented the eye tracking data as $\Re = \{V, T\}$, V is the set of gaze velocity.

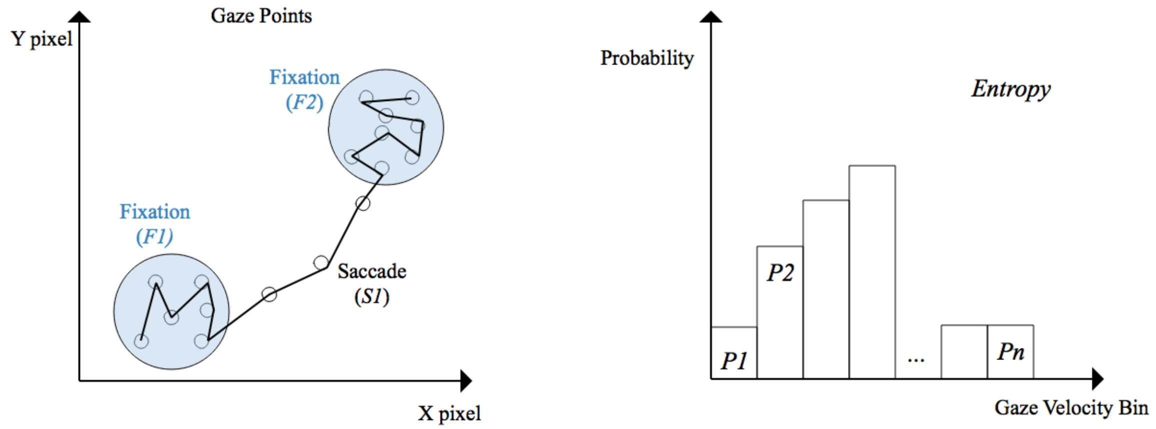
We split the eye-tracking data into short-period data packets first and then conduct bin analysis. The time series data of eye movement were split into N packets based on time window and shift window. Time window (T_w) denotes the period of each packet. Shift window (S_w) refers to the length that time window moves forward. T refers to the length of the eye tracking data.

$$N = \left\lfloor \frac{T}{S_w} \right\rfloor - \left\lfloor \frac{T_w}{S_w} \right\rfloor + 1 \tag{2}$$

$$M = T_w \times f \tag{3}$$

In this way, the eye tracking data $\Re = \{V, T\}$ are presented as $\Re = \{D_1, D_2, \dots D_N\}$. M refers to the number of gaze points belonging to a data packets. f is the sampling rate frequency of the eye tracker.

The basic idea of gaze-bin analysis is shown in Figure 1 (b). We present the eye-tracking data $\Re = \{D_1, D_2, \dots D_N\}$ as $\Re = \{P\} = \{p_1, p_2, \dots p_N\}$, where P is the set of gaze velocity probability vector.



(a) Convential parameters (b) Gaze-bin analysis

Figure 1: Representation of eye tracking data

We extended the bin analysis to gaze velocity analysis by describing each data packet as a histogram of all velocities the eye travel during the time window (*Tw*). For the gaze velocity data belonging to the data packet, it is transformed into a discrete probability mass function and entropy, $p=(p(v_1), p(v_2),\dots, p(v_B), e)$, where $p \in P, p(vb) >= 0$ for all b, $\sum_b p(v_b) = 1$. The probability and entropy can be calculated by Eq. (3) and (4), respectively.

$$\begin{cases} p(v_b) = \dfrac{n_f}{Tw \times f} \\ p(v_b) = \dfrac{n_s}{Tw \times f}, \quad for\ b = B \end{cases} , for\ b = 1,2,\dots,B-1 \tag{4}$$

$$\mathrm{e} = \sum_{b=1}^{B} p(v_b) \times log(p(v_b)) \tag{5}$$

where

n_f refers to the number of gazes whose velocity meet the requirement : $B \times \frac{V}{B} \geq v > (b-1) \times \frac{V}{B}$.

n_s refers to the number of gazes whose velocity meet the requirement : $\mathrm{v} > \mathrm{V}$.

V is the maximum velocity of the fixation;

B is the number of bins; f refers to the sampling rate of the eye tracker;

e is the entropy of the gaze velocities belonging to the time window.

Since velocities of saccade are much faster than velocities of fixations, and the amount of time spent on saccades is substantially shorter than on fixation, just one bin (bin *B*) was set for saccades.

3.2 Problem Statement

Unlike the existing studies, we proposed a novel analysis of short-period eye-tracking data instead of statistical analysis of fixations and saccades, which can only be obtained from long-recorded eye-tracking data. In this way, we introduced a new problem of using gaze-bin analysis to discriminate human fatigue in TCCs. The problem is stated as follows: given a set of eye-tracking data $\mathfrak{R} = \{p_1, \dots, p_m\}$, a set of fatigue levels $FL = \{fl_1, \dots, fl_j\}$, we aim to: (1) train a fatigue model $\{\mathfrak{R}, FL\} \to FM$; as well as (2) predict human fatigue level using short-period eye tracking data $fl \overset{FM}{\longleftarrow} \{p\}$. This problem is based on the following assumptions:

Assumption 1: The dimension of FL is much smaller than the amount of data packets belonging to $\mathfrak{R}$. In other words: $j \ll m$.

Assumption 2: The eye-tracking data collected from the same participant are subject to the same distribution.

Assumption 3: There are some overlaps among data packets, $p_d \cap p_{d+1} \neq \emptyset, where\ \{p_d, p_{d+1}\} \in P$.

Compare with existing human fatigue detection works that focused on fixations and saccades from long-recorded eye-tracking data, our solution of using gaze-bin analysis is expected to capture dynamic features from short-period eye-tracking data. Moreover, the gaze-bin analysis is independent on event detection methods and eye trackers. Hence, it is expected to reduce the contradictory results caused by eye trackers and event detection methods.

4. Semi-supervised Bagged Trees for Human Fatigue Detection

We proceed to present an innovative method to address the challenges mentioned in Section 1 for gaze-bin analysis-based human fatigue detection. In general, training data and test data were randomly selected. Nevertheless, simple selection is not applicable in this work due to the overlapped data packets. Hence, we introduced our proactively selection of training and testing data first. Then, we developed a semi-supervised bagged trees method to train fatigue model using labeled and unlabeled data. Last, we state the procedures of human fatigue detection.

4.1 Proactively Selection of Training and Testing Data

In total, *N* data packets are utilized for training and testing the bagged trees model. First, we randomly selected 80% of data packets for training and remains for testing. To avoid overfitting caused by the overlaps of data packets, the training data and testing data were split before training. The overlapped data packets were deleted, as shown in Figure 2. A random number *s* was generated by Matlab R2018a. To avoid s is out of index, s was set to be larger than N_P and smaller than 0.8*N*-N_P-1. The data packets numbered *s* to *s*+0.2*N* were utilized as testing data. The data packets whose number range from s-0.5N_P to *s*+0.2*N* +0.5N_P were deleted. After deleting the overlapped data, the remaining data were used to train the model.

$$N_P = 2 * (\left\lfloor \frac{T_w}{S_w} \right\rfloor - 1) \quad (6)$$

where, N_P is the number of overlapped data packets.

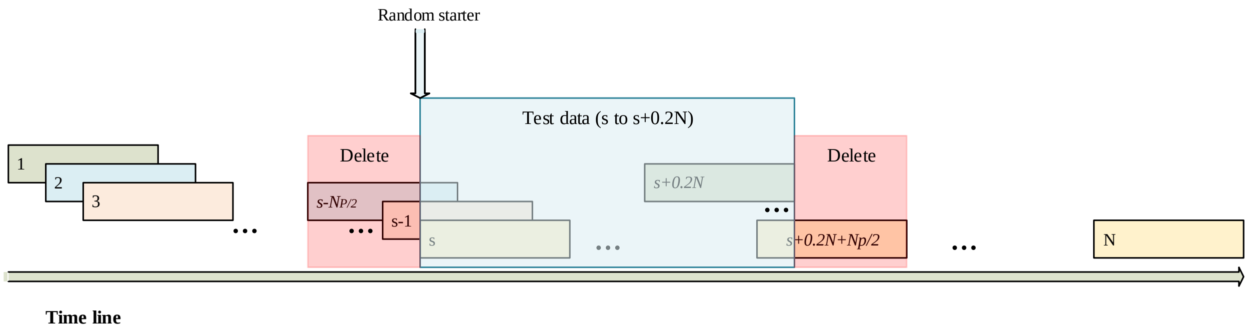


Figure 2: Generating data for training and testing

In this way, we obtained training data set $Tr = \left\{D_{1,} \ldots, D_{s-N_p-1}, D_{s+0.2N+N_p+1,} \ldots, D_N\right\}$, and testing data set $Ts = \left\{D_{s,} \ldots, D_{s+0.2N}\right\}$.

4.2 Semi-supervised Training with Decision Tree

Given a set of eye-tracking data $Tr = \left\{D_{1,} \ldots, D_{s-N_p-1}, D_{s+0.2N+N_p+1,} \ldots, D_N\right\}$, a set of fatigue levels FL, the amount of data packets in Tr is much smaller than the dimension of FL. We present $\{Tr, FL\}$ as $\{D^l, D^u\}$, where D^l refers to the set of labeled eye-tracking data and D^u refers to the set of unlabeled eye-tracking data. Basically, $\{D^l, D^u\} = \{d_1^l, d_2^l, \ldots d_j^l, d_{j+1}^u, \ldots, d_m^u\}$, $d^l = \{p, fl\}$ and $d^u = \{p\}$. Hence, we search for a labelling vector $FL^* = (fl_{j+1}^*, \ldots, fl_m^*)^T$ for the unlabelled dataset.

We first train a decision tree using the labeled data set D^l, then use the decision tree to generate a labelling vector $FL^* = (fl_{j+1}^*, \ldots, fl_m^*)^T$ for the unlabelled dataset. A subset fl_{sub}^* with high classification confidence of the FL^* would be selected to label the D_{sub}^u. Then we retrain decision tree using the data set $\{D^l, D_{sub}^u\}$ and reselect a subset data from the unlabelled dataset. The procedure is repeated until it reaches a stopping condition. We used the C4.5 algorithm to build decision trees. The C4.5 algorithm is the most well-known algorithm for building decision trees and has been successfully used in many fields [50]. This has motivated us to develop trees using the C4.5 algorithm. Algorithm 1 presents the main structure of semi-supervised training algorithm. The output training dataset can be represented as $Tr = \{d_1^l, \ldots, d_m^l\}$

Algorithm 1: Semi-supervised training algorithm

```
T: number of iterations.   C: prediction confidence threshold
t=1
while t<T
```

Dt ← Decision tree(D^l)

FL^*= Bt (D^u)

select D^u_{sub} whose p(FL^*_{sub})>C

FL ← {FL+FL^*_{sub}}

D^l ← {D^l+D^u_{sub}}

D^u ← {D^u-D^u_{sub}}

t=t+1

end

Output: generate final training dataset D^l and a labelling vector FL^*

4.3 Bagged trees-enabled Human Fatigue Detection

In order to build A trees for bagging, we generated A copies of data set $BT = \{BT_1, BT_2, \dots BT_A\}$ from the training data set Tr using Sampling with replacement. Based on each of the data set BT_a, a decision tree F^{*a} could be built using Classification And Regression Tree (CART) algorithm.

Given the training data set $BT_a = \{(p^a_1, fl^a_1), \dots, (p^a_m, fl^a_m)\}$ where $p^a \in P, fl^a \in FL$, the decision tree F^{*a} divide P into several feature subspace $\{\mathcal{R}_1, \dots, \mathcal{R}_K\}$. On each subspace $\mathcal{R}_k$, the same prediction is made for all $p_w \in \mathcal{R}_k$. The estimated probability of each class F on subspace $\mathcal{R}_k$ is:

$$\widehat{Pro_{kF}} = \sum\nolimits_i \mathrm{I}(fl^a_i = F) \cdot \mathrm{I}(p^a_i \in \mathcal{R}_k) \Big/ \sum\nolimits_i \mathrm{I}(p^a_i \in \mathcal{R}_k) \quad (7)$$

We aims at finding F^{*a} which could achieve minimized Gini index:

$$Gini = \sum_{F=0}^{1} \widehat{Pro_{kF}}(1 - \widehat{Pro_{kF}}) \quad (8)$$

The final prediction for a give observation p is calculated using eq.(6) .

$$\hat{F}_{bag}(p) = \left\lfloor \frac{1}{2A} \sum_{a=1}^{A} \hat{F}^{*a}(p) \right\rfloor \quad (9)$$

In general, out-of-bag errors are usually used to evaluate the bagged model. Although the out-of-bag error estimation is particularly convenient, it is not suit for this study due to the overlaps of data packets. As illustrated in Figure 2, each data packet overlaps with several other data packets. In this way, it is too complex to identify out-of-bag observation. Hence, we extended k-fold cross-validation method in this study. We used the aforementioned testing data set $Ts = \{D_{s,} \dots, D_{s+0.2N}\}$, where $D_s = (P_s, FL_s)$ to test the bagged model. For each fold l, we could obtained the $Accuracy_l$ based on Eq. (10).

$$Accuracy_l = 1 - \frac{1}{0.2N} \sum_{r=s}^{s+0.2N} \left| (FL_r - \hat{F}_{bag}(P_r)) \right| \quad (10)$$

Beside accuracy, sensitivity and specificity were tested. Basically, accuracy measures the percentage of correctly classified observations. As for sensitivity, which is also known as the true positive rate, it determines the proportion of actual positives that are correctly identified as positives. On the other hand, specificity measures the proportion of actual negatives that are correctly identified as negative. In this study, 'alert' is defined as positive and 'fatigue' is defined as negative.

5. Case Study

In the following, the data collection and the performance of the proposed method are described. The effects of the model characteristics including time window, bin number, and input features were investigated and presented. In addition, we compared our method with other commonly used methods such as decision tree, linear regression and support vector machine.

5.1 Data collection and model construction of VTS operators

Eight Vessel Traffic Service Operators (VTSOs) comprising seven males and one female with normal vision were recruited for this study. Their age ranges from 30-years old to 50-years old. All of them were working in the morning shift, which starts from 7:30 am to 15:30 pm. None of them suffered from sleep disorders.

The data collection phase comprised four sessions (Figure 3). The participants were asked to conduct their daily work and to do 5-min Mackworth Clock Test after every two hours of work. The participants were instructed to complete the Samn-Perelli Fatigue Scale before and after the Mackworth Clock Test. Their eye movements were recorded using Tobii X3-120.

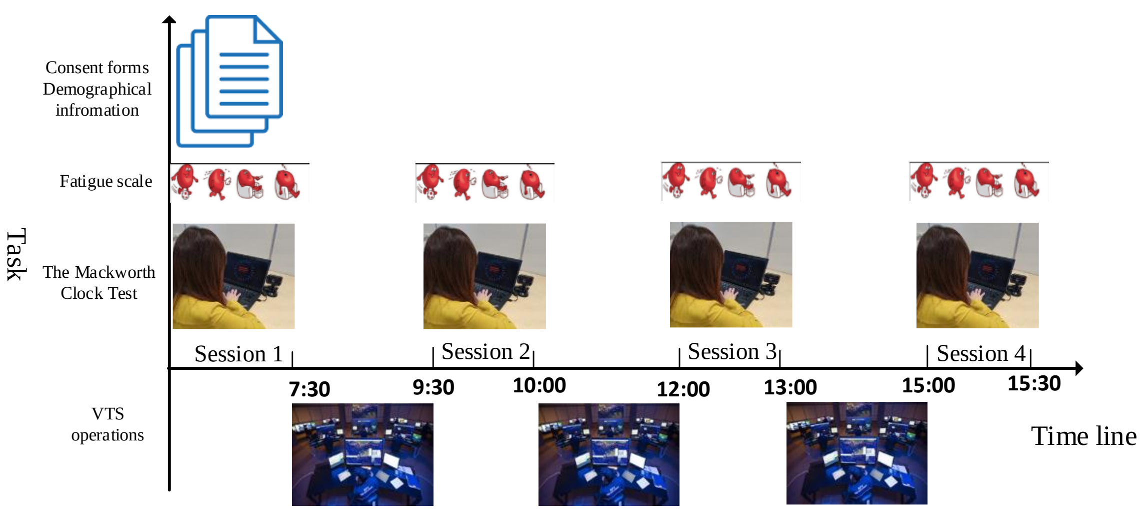


Figure 3: The procedures of data collection

In total, we collected eye movements data of 160 minutes (8 participants * 5 minutes * 4 sessions) from eight VTSOs. The human fatigue data were summarized from the results of the Samn-Perelli Fatigue Scale and the Mackworth Clock test.

For generating data packets, we set the time window ranges from 4 seconds to 10 seconds. The shift window was set as a quarter of the time window. According to the Eq (6), the number of the overlapped data packets is 6. The amount of data packets *N* varies with the length of the time window, as shown in Table 1. Besides the time window, the value of bin number (b) requires more investigation, as it can affect the performance of human fatigue detection model. Specifically, if b is too large, it may induce data redundancy and low efficiency of computing. If b is too small, some dynamic information of the gaze data could not be captured by the bin analysis. In this study, we set b as 6, 8, 10, 12, 14 and 16.

Table 1: The amount of data packets with different time window (for each participant)

Time window (seconds)	**N**	**Np**	**N test**
4	1196	6	1199
6	796	6	159
8	596	6	119
10	476	6	95

5.2 Performance of the human fatigue detection models

We established 192, i.e. 6 bin numbers x 4 time windows x 8 participants, bagged tree models and tested their performance from three aspects, namely accuracy, sensitivity and specificity. For each model, 10 observations were obtained from 10-fold cross-validation. Multiple ANOVAs were conducted to test the differences in observation of all the models.

Figure 4 shows the performance of human fatigue detection models across different time windows. We conducted multiple ANOVAs to test the differences in performance of the models with different time window. The results showed that the time window had a significant effect on the accuracy ($F_{3, 168}$=16.062, p<0.001) and specificity ($F_{3, 168}$=33.891, p<0.001) of the human fatigue detection model, but not on sensitivity ($F_{3, 168}$=1.319, p=0.270). It can be deduced that the time window has no significant effects on detecting ‘alert’ state, while has significant effects on detecting ‘fatigue’ state. In other words, using longer periods to summarize the data made the fatigue state easier to be detected. With the increase of time window, the human fatigue detection model can detect ‘fatigue’ state with higher accuracy and more sensitivity. While the main effects of the bin number on the model performance is not significant (accuracy: $F_{5, 168}$=0.013, p=1.000; sensitivity: $F_{5, 168}$=0.55, p=0.998; specificity: $F_{5, 168}$=0.113, p=0.989). Hence, the performance of human fatigue detection model is not sensitive to the bin number. Figure 5 shows the line graph of the model performance across different bin number. The best performance (accuracy=0.90; sensitivity=0.81; specificity=0.90) was achieved when the time window was set to 10 seconds and the bin number was set as 8.

Figure 6 shows the cloud graph of human fatigue detection accuracy with different bin number and time window. As shown in Figure 6, longer time window could achieve better performance. Nevertheless, larger in number showed no significant improvement. The interaction effect of bin number and time window appeared to be insignificant (accuracy: $F_{15,168}$=0.03, p=1.000; sensitivity: $F_{15,168}$=0.30, p=1.00;specificity: $F_{15,168}$=0.091, p=1.000). In the followings sections, the bin number is fixed at 8.

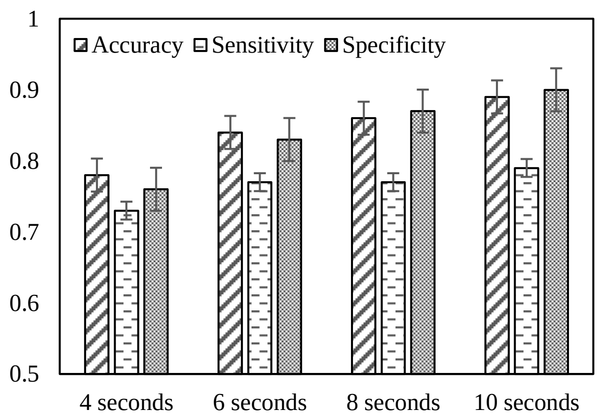


Figure 4: Model performance vs time window

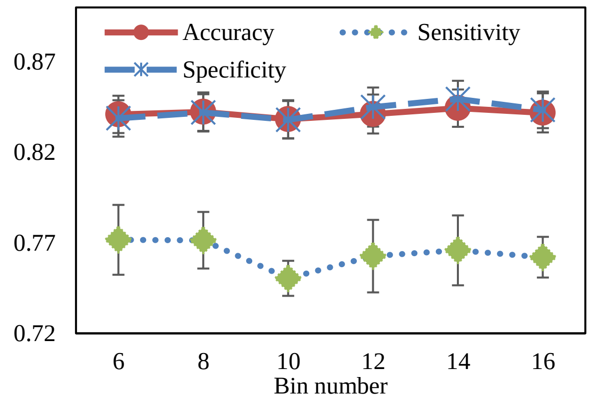


Figure 5: Model performance vs bin number

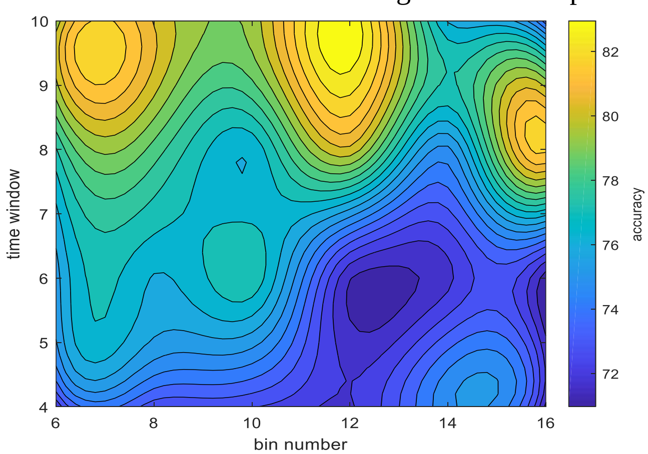


Figure 6: Effects of bin number and time window on detecting accuracy

5.3 Comparison with other common fatigue indicators

In this section, we compared the performance of four kinds of feature combinations, namely gaze bin analysis, fixations, saccades and combination data. Table 2 shows their parameters. Specifically, gaze bin analysis includes two kinds of parameters, namely velocity probability of each bin and entropy. For fixations and saccades, many statistical parameters could be generated from the raw eye movement data.

Hence, we selected several frequently used fixation parameters and saccade parameters based on literature review.

Table 2: Feature combinations used as model input

	Feature combination			
Eye movement parameters	Gaze bin analysis (G)	Fixation data (F)	Saccade data (S)	combination data (C)
Mean fixation duration		✓		✓
Mean fixation velocity		✓		✓
Mean fixation stability		✓		✓
Fixation count		✓		✓
Saccade count			✓	✓
Mean saccade peak velocity			✓	✓
Mean saccade velocity			✓	✓
Mean saccade amplitude			✓	✓
Mean saccade duration			✓	✓
Gaze velocity probability	✓			
Entropy	✓			

To investigate the contribution of feature combinations on human fatigue detection, 128, i.e. 4 feature combinations x 4 time windows x 8 participants bagged tree models were built.

For accuracy：

Mauchly's t

Table 3 shows the performance of fatigue models with four kinds of feature combination. The results showed that it had a significant effect on the performance of human fatigue detection (accuracy: $F_{3, 112}=2.283$, $p=0.083<0.05$; sensitivity: $F_{3, 196}=1.489$, $p<0.05$; specificity: $F_{3, 196}=7.772$, $p<0.001$). The multiple comparisons showed that the gaze bin analysis could achieve a higher accuracy in human fatigue detection than other inputs. There is no significant difference in model performance among fixation data, saccade data and combination data (p=0.065). The results indicated that fixation data could contribute to human fatigue detection. What's more, the results showed that the combination of fixation data with saccade data cannot greatly improve the accuracy of human fatigue detection. It could be explained in two aspects: first, the combination of fixation and saccade introduces too many input features and impairs the performance of the bagged tree. Second, there should be some relations between fixation and saccades.

There was no interaction effect of time window and feature combinations (accuracy: $F_{18, 196}=0.092$, $p=1.000$; sensitivity: $F_{18, 196}=0.124$, $p=1.000$; specificity: accuracy: $F_{18, 196}=0.247$, $p=0.999$). For all kinds of inputs, the model performance increased with time window.

Table 3: Model performance (time window=4, 6, 8, 10 seconds, bin number=8)

Performance	Gaze bin analysis	Fixation data	Saccade data	Combination data

Accuracy	0.84	0.77	0.79	0.81
Sensitivity	0.77	0.71	0.67	0.75
Specificity	0.85	0.75	0.78	0.79

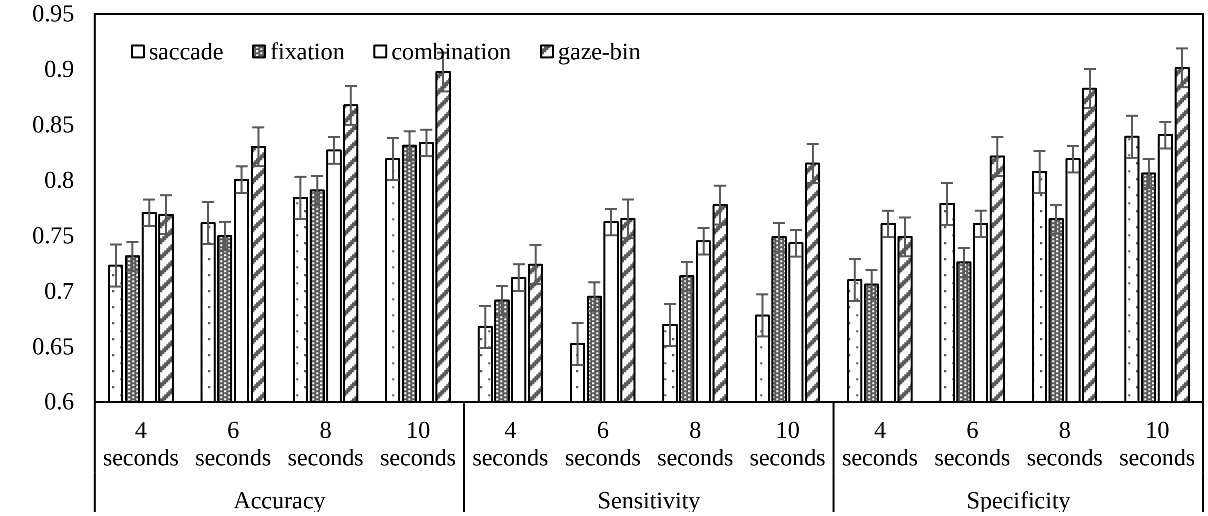


Figure 7: Effects of feature combination and time window on the performance of human fatigue detection

5.4 Comparison with other methods

In this section, the proposed approach was compared with other classic methods. The results gathered from 128, i.e. 4 time windows x 4 methods x 8 participants models were tested in the comparative study.

Table 4 shows the accuracy of different methods using gaze bin analysis as inputs. The methods used to detect human fatigue had significant effects on accuracy and sensitivity (accuracy: $F_{3, 112}=7.771, p<0.001$; sensitivity: $F_{3, 112}=3.52, p<0.05$). Multiple comparisons show that Bagged trees outperformed other three methods in terms of accuracy and sensitivity. Nevertheless, there is no significant effect on the accuracy and sensitivity of the other three methods. No significant effects had been found on specificity ($F_{3, 112}=0.729, p=0.537$). Therefore, Bagged trees cannot significantly improve the efficiency in detecting 'fatigue' state. Time window affected specificity ($F_{3, 112}=11.311, p<0.001$), but not accuracy ($F_{3, 112}=0.735, p=0.533$) and sensitivity ($F_{3, 112}=0.541, p=0.655$). The model's specificity increased with time window size, suggesting that using longer periods of eye movement data could improve the model's ability in detecting fatigued state. The interaction effects of time window and methods are not significant, too (accuracy: $F_{9, 112}=0.125, p=0.999$; sensitivity: $F_{9, 112}=0.26, p=0.982$; specificity: $F_{9, 112}=0.311, p=0.970$). Compared with the results of Section 4.2, we could conclude that the time effects on Bagged trees are more significant than on the other three methods. That's why no main effect of time window was found in this Section.

Table 4: Model performance (bin number=12, time window=4, 6, 8,10seconds)

Performance	Decision Tree	Linear regression	Support vector machine	The Bagged trees

Accuracy	0.76	0.71	0.70	0.84
Sensitivity	0.65	0.63	0.68	0.75
Specificity	0.82	0.81	0.81	0.85

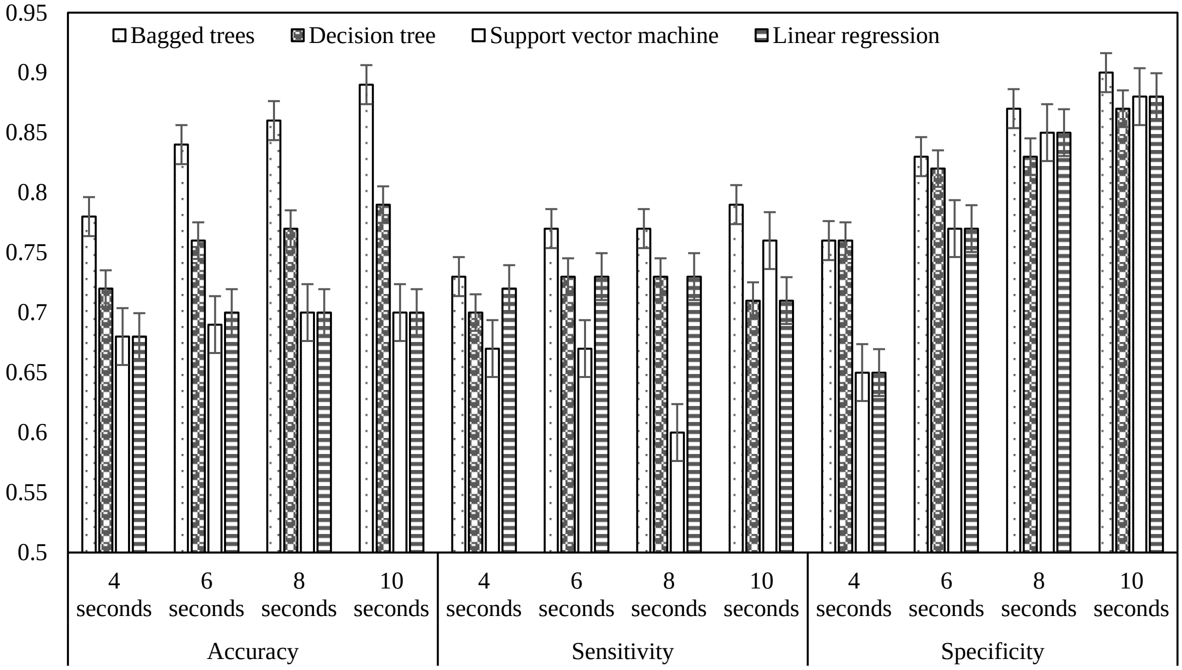


Figure 8: Effects of methods and time window on the performance of human fatigue detection

6. Conclusion

The performance of TCOs is usually impaired by mental fatigue, resulting in great threats to public safety. To minimize its risks, we proposed an innovative approach to non-invasively detect human fatigue of TCOs using gaze-bin analysis. A case study was conducted to test the effectiveness and performance of the proposed approach.

The contributions of this paper can be concluded as follows. First, we paved an innovative way to detect mental fatigue of TCOs. Instead of using brain dynamics, this work explored the possibility of using short-period eye movement in this field. This work provides an alternative way to reduce the risk of mental fatigue in TCCs. Second, we proposed an innovative approach to timely detecting mental fatigue using short-period eye movement data. The utilization of short-period data unlocks the critical bottleneck of time delay in detecting human fatigue. Third, we did a pioneering work of using gaze-bin analysis as human fatigue indicators. The utilization of gaze-bin analysis enables the application of a low sampling rate eye tracker. Hence, it can significantly reduce the cost and improve the usability of eye movement-based human fatigue detector. Fourth, we extended the application of Bagged trees on human factors study. It outperformed the commonly used machine learning methods, such as decision tree and support vector machine. Last, we took the lead of testing the effects of time window and bin number on model performance. The results provided a reference and guidance for future studies.

One limitation of this study is that the proposed method cannot clearly detect medium fatigue from alert. The proposed method can only warn fatigued users. It cannot be used in the fields that required operators to be quite alert. What’s more, to guarantee the performance of human fatigue model, training data of each user is required. For future studies, a general model which can detect more levels of human fatigue would be developed.

Acknowledgement

This work was supported by Singapore Maritime Institute Research Project (SMI-2014-MA-06). The authors would like to thank all VTS operators and students participated in this study.

Reference

1. Praetorius, G., *Safety within the Vessel Traffic Service (VTS) Domain. Understanding the role of the VTS for safety within maritime traffic management.* 2012.
2. Härmä, M., et al., *The effect of an irregular shift system on sleepiness at work in train drivers and railway traffic controllers.* Journal of sleep research, 2002. **11**(2): p. 141-151.
3. Hou, X., et al., *EEG-based human factors evaluation of conflict resolution aid and tactile user interface in future Air Traffic Control systems*, in *Advances in Human Aspects of Transportation*. 2017, Springer. p. 885-897.
4. Metzger, U. and R. Parasuraman, *The role of the air traffic controller in future air traffic management: An empirical study of active control versus passive monitoring.* Human Factors, 2001. **43**(4): p. 519-528.
5. XU, G., et al., *Toward Resilient Vessel Traffic Service: A Sociotechnical Perspective.* Transdisciplinary Engineering: A Paradigm Shift, 2015: p. 829.
6. Praetorius, G., E. Hollnagel, and J. Dahlman, *Modelling Vessel Traffic Service to understand resilience in everyday operations.* Reliability engineering & system safety, 2015. **141**: p. 10-21.
7. Roets, B. and J. Christiaens, *Shift work, fatigue, and human error: An empirical analysis of railway traffic control.* Journal of Transportation Safety & Security, 2017: p. 1-18.
8. Hsu, W.-K.K., *Ports’ service attributes for ship navigation safety.* Safety science, 2012. **50**(2): p. 244-252.
9. Lerman, S.E., et al., *Fatigue risk management in the workplace.* Journal of Occupational and Environmental Medicine, 2012. **54**(2): p. 231-258.
10. Ray, S.J. and J. Teizer, *Real-time construction worker posture analysis for ergonomics training.* Advanced Engineering Informatics, 2012. **26**(2): p. 439-455.
11. Cyganek, B. and S. Gruszczyński, *Hybrid computer vision system for drivers' eye recognition and fatigue monitoring.* Neurocomputing, 2014. **126**: p. 78-94.
12. Olsen, A., *The Tobii I-VT fixation filter.* Tobii Technology, 2012.
13. Bodala, I.P., et al. *Measuring vigilance decrement using computer vision assisted eye tracking in dynamic naturalistic environments*. in *Engineering in Medicine and Biology Society (EMBC), 2017 39th Annual International Conference of the IEEE*. 2017. IEEE.

14. Bodala, I.P., et al., *EEG and eye tracking demonstrate vigilance enhancement with challenge integration.* Frontiers in human neuroscience, 2016. **10**: p. 273.
15. Cazzoli, D., et al., *Eye movements discriminate fatigue due to chronotypical factors and time spent on task–a double dissociation.* PloS one, 2014. **9**(1): p. e87146.
16. Di Stasi, L.L., et al., *Saccadic eye movement metrics reflect surgical residents' fatigue.* Annals of surgery, 2014. **259**(4): p. 824-829.
17. Ji, Q. and X. Yang, *Real-time eye, gaze, and face pose tracking for monitoring driver vigilance.* Real-time imaging, 2002. **8**(5): p. 357-377.
18. McKinley, R.A., et al., *Evaluation of eye metrics as a detector of fatigue.* Human Factors: The Journal of the Human Factors and Ergonomics Society, 2011. **53**(4): p. 403-414.
19. 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.
20. Ahlstrom, U. and F.J. Friedman-Berg, *Using eye movement activity as a correlate of cognitive workload.* International Journal of Industrial Ergonomics, 2006. **36**(7): p. 623-636.
21. Nyström, M. and K. Holmqvist, *An adaptive algorithm for fixation, saccade, and glissade detection in eyetracking data.* Behavior research methods, 2010. **42**(1): p. 188-204.
22. Liang, Y., M.L. Reyes, and J.D. Lee, *Real-time detection of driver cognitive distraction using support vector machines.* IEEE transactions on intelligent transportation systems, 2007. **8**(2): p. 340-350.
23. Hartley, L., et al., *Review of fatigue detection and prediction technologies.* National Road Transport Commission, 2000.
24. Breiman, L., *Bagging predictors.* Machine learning, 1996. **24**(2): p. 123-140.
25. Pearson, K., *Contributions to the mathematical theory of evolution. II. Skew variation in homogeneous material.* Philosophical Transactions of the Royal Society of London, 1895. **186**(Part I): p. 343-424.
26. Morris, T. and J.C. Miller, *Electrooculographic and performance indices of fatigue during simulated flight.* Biological psychology, 1996. **42**(3): p. 343-360.
27. Schleicher, R., et al., *Blinks and saccades as indicators of fatigue in sleepiness warnings: looking tired?* Ergonomics, 2008. **51**(7): p. 982-1010.
28. Du, L.-H., et al. *Detecting driving fatigue with multimodal deep learning*. in *Neural Engineering (NER), 2017 8th International IEEE/EMBS Conference on*. 2017. IEEE.
29. Jin, L., et al., *Driver sleepiness detection system based on eye movements variables.* Advances in Mechanical Engineering, 2013. **5**: p. 648431.
30. Azim, T., M.A. Jaffar, and A.M. Mirza, *Fully automated real time fatigue detection of drivers through fuzzy expert systems.* Applied Soft Computing, 2014. **18**: p. 25-38.
31. Han, W., et al. *Driver drowsiness detection based on novel eye openness recognition method and unsupervised feature learning*. in *Systems, Man, and Cybernetics (SMC), 2015 IEEE International Conference on*. 2015. IEEE.
32. Di Stasi, L.L., et al., *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.
33. Finke, C., et al., *Dynamics of saccade parameters in multiple sclerosis patients with fatigue.* Journal of neurology, 2012. **259**(12): p. 2656-2663.
34. Hirvonen, K., et al., *Improving the saccade peak velocity measurement for detecting fatigue.* Journal of neuroscience methods, 2010. **187**(2): p. 199-206.
35. Saito, S., *Does fatigue exist in a quantitative measurement of eye movements?* Ergonomics, 1992. **35**(5-6): p. 607-615.
36. Zhu, X., et al. *EOG-based drowsiness detection using convolutional neural networks*. in *IJCNN*. 2014.

37. Moacdieh, N.M. and N. Sarter, *Using Eye Tracking to Detect the Effects of Clutter on Visual Search in Real Time.* IEEE Transactions on Human-Machine Systems, 2017. **47**(6): p. 896-902.
38. Sain, S.R., *The nature of statistical learning theory*. 1996, Taylor & Francis Group.
39. Jun, M. and J.C. Cheng, *Selection of target LEED credits based on project information and climatic factors using data mining techniques.* Advanced Engineering Informatics, 2017. **32**: p. 224-236.
40. Golparvar-Fard, M., A. Heydarian, and J.C. Niebles, *Vision-based action recognition of earthmoving equipment using spatio-temporal features and support vector machine classifiers.* Advanced Engineering Informatics, 2013. **27**(4): p. 652-663.
41. Saunders, D.G., et al., *Two-dimensional data binning for the analysis of genome architecture in filamentous plant pathogens and other eukaryotes*, in *Plant-Pathogen Interactions*. 2014, Springer. p. 29-51.
42. Davis, R.A., et al., *Adaptive binning: An improved binning method for metabolomics data using the undecimated wavelet transform.* Chemometrics and intelligent laboratory systems, 2007. **85**(1): p. 144-154.
43. Lavielle, M. and K. Bleakley, *Automatic data binning for improved visual diagnosis of pharmacometric models.* Journal of pharmacokinetics and pharmacodynamics, 2011. **38**(6): p. 861-871.
44. Shannon, C.E., W. Weaver, and A.W. Burks, *The mathematical theory of communication.* 1951.
45. Zhang, C.-Y., et al., *Data-driven train operation models based on data mining and driving experience for the diesel-electric locomotive.* Advanced Engineering Informatics, 2016. **30**(3): p. 553-563.
46. Son, H., et al., *Classification of major construction materials in construction environments using ensemble classifiers.* Advanced Engineering Informatics, 2014. **28**(1): p. 1-10.
47. Quinlan, J.R. *Bagging, boosting, and C4. 5*. in *AAAI/IAAI, Vol. 1*. 1996.
48. De'Ath, G., *Boosted trees for ecological modeling and prediction.* Ecology, 2007. **88**(1): p. 243-251.
49. Dietterich, T.G., *An experimental comparison of three methods for constructing ensembles of decision trees: Bagging, boosting, and randomization.* Machine learning, 2000. **40**(2): p. 139-157.
50. Quinlan, J.R., *C4. 5: programs for machine learning*. 2014: Elsevier.