Assessment of Pilots’ Cognitive Competency Using Situation Awareness Recognition Model Based on Visual Characteristics

Shaoqi Jiang,1,2 Ruifang Su,3 Zhenzhen Ren,4 Weijiong Chen,2,4 and Yutao Kang2

1College of Information Engineering, Jinhua Polytechnic, Jinhua, Zhejiang, China
2College of Environment and Engineering, Shanghai Maritime University, Shanghai, China
3Department Information Center, Jinhua Maternal and Child Health Care Hospital, Jinhua, Zhejiang, China
4College of Merchant Marine, Shanghai Maritime University, Shanghai, China

Correspondence should be addressed to Ruifang Su; yzzhang997@163.com

Received 28 July 2023; Revised 8 January 2024; Accepted 22 January 2024; Published 12 February 2024

Visual characteristics have the potential to assess the navigational proficiency of ship pilots. A precise assessment of ship piloting competence is imperative to mitigate human errors in piloting. An exhaustive examination of cognitive capabilities plays a pivotal role in developing an enhanced and refined system for classifying, selecting, and training ship piloting proficiency. Insufficiency in situation awareness (SA), denoting the cognitive underpinning of hazardous behaviors among pilots, may lead to subpar performance in ship pilotage when faced with adverse conditions. To address this issue, we propose an SA recognition model based on the random forest-support vector machine (RF-SVM) algorithm, which utilizes wearable eye-tracking technology to detect pilots’ at-risk cognitive state, specifically low-SA levels. We rectify the relative error (RE) and root mean square error (RMSE) and employ principal component analysis (PCA) to enhance the RF algorithm, optimizing the combination of salient features in greater depth. Through the utilization of these feature combinations, we construct a SVM algorithm using the most suitable variables for SA recognition. Our proposed RF-SVM algorithm is compared to RF or SVM alone, and it achieves the highest accuracy in recognizing at-risk cognitive states under poor visibility conditions (an improvement of 86.79% to 93.43% in accuracy). Taken collectively, the present findings offer vital technical support for developing a technique-based intelligent system for adaptively evaluating the cognitive accomplishment of pilots. Furthermore, they establish the groundwork and framework for the surveillance of cognitive processes and capabilities in marine pilotage operations within China.

1. Introduction

Situation awareness (SA) enhancement among ship pilots is critical to lowering anthropogenic errors, which have occupied 75%–96% of maritime accidents in recent years [1, 2]. With the increasing size, speed, and traffic density of ships, enhancing operational safety for pilots has become a pressing issue [3]. However, due to the complexity of marine systems and the growing integration of intelligence and automation, pilots face an increasingly daunting task in comprehending the current situation and predicting future changes [4, 5]. To effectively prevent unsafe behaviors among pilots, it is necessary to recognize their SA levels from a cognitive perspective, particularly in emergency situations during ship pilotage [6]. However, the requirement for pilots to maintain high-SA levels, along with individual differences and empirical aspects of pilotage, can create measurement gaps [7]. Therefore, recognizing SA levels as a means of avoiding human errors becomes a complex process that requires further investigation, especially in ship pilotage emergencies.

Current research primarily focuses on statistical methods to determine the correlation between poor visibility and real-world marine navigation accidents [3]. Since poor visibility deteriorates pilots’ perception of nearby contextual changes, it contributes prominently to accidents by leading to heightened risk of accidents [8]. Chauvin et al. [9] identified visibility as a significant contributing factor to
collision accidents, accounting for 56.51% of contextual factors according to the Human Factors Analysis and Classification System (HFACS). Supporting this conclusion, Bye and Aalberg [10] conducted a study that assessed the impact of different visibility conditions on accident causation. Their findings indicated that the most pronounced effect occurred when visibility dropped below 0.25 nautical miles. Although the current research has successfully examined whether poor visibility is more likely to trigger accidents compared to other environmental variables (such as flow rate and water depth), it has neglected to evaluate the cognitive state changes in pilots when anticipating potential hazards under poor visibility conditions [11]. Currently, a widely recognized cognitive framework for SA is the three-tiered architecture, which encompasses three stages: perception, comprehension, and projection of the present environment into the future [12]. Hence, in situations characterized by reduced visibility, the application of the three-tier cognitive framework to precisely evaluate the situation awareness (SA) levels of pilots becomes a critical challenge that requires attention, as pointed out by [13].

Eye-tracking techniques have been used in literature studies to examine the relationship between cognitive states and visual behaviors, including pupil diameter, saccade frequency, and fixation time [14, 15]. Louw and Merat [16] demonstrated evident scattering of a driver’s visual attention during automated driving by simulation experiments. Nonetheless, the internal processes connecting the level of automation to attention were not taken into account. Given this observation, the connection between eye movement measurements and situation awareness (SA) level, as well as the fundamental influencing factors, has emerged as a central research area in the transportation domain [17]. In many works, fixation metrics are linked to SA [17–19]. Moderate associations between saccade metrics and SA have also been verified in various studies [20, 21]. However, no work has identified a connection between pupil dilation or blink rate and SA [22]. It is worth noting that although some studies have utilized multiple eye-tracking metrics [23], not all of them were associated with SA. In general, it is acknowledged that longer time spent by subjects in a specific area of interest (AOI) indicates higher SA [24]. However, a consensus remains elusive concerning the correlation between eye-tracking metrics and situation awareness (SA) due to discrepancies in application objectives and task conditions.

Despite the extensive research on the correlation between eye-tracking metrics and SA in various application domains, the problem of SA identification based on eye-tracking technology remains unresolved. From the aforementioned literature investigation, it is evident that some studies did not consider SA measurement as their primary research goal [25, 26] or only included it as one of their research goals [15, 17]. Only a limited number of studies explicitly mentioned that their primary aim was to establish the link between eye-tracking metrics and situation awareness (SA). These studies include those conducted by [23, 27]. Furthermore, to the best of our knowledge, the direct investigation of eye tracking-based SA recognition approaches is scarce.

Therefore, this study has two main objectives. The primary aim was to evaluate the correlations between eye movement features and SA. Through correlation analyses, we confirm significant associations between saccade and fixation metrics and SA levels, consistent with previous findings [28]. The ultimate goal of the present study is to explore an SA recognition approach based on relevant eye-tracking metrics, with the hope of reducing pilotage risks by enhancing the selection and training of pilots. With this objective in mind, this paper introduces a novel approach for assessing pilots’ SA using eye-tracking metrics. It is anticipated that this method could contribute to the development of an SA assessment technique based on physiological measurements and offer insights into how to ‘train and enhance pilots’ SA.

2. Experimental Methods

Following the eye-tracking experiment conducted on a bridge simulator, a research framework was developed to detect the correlation between eye movement metrics and SA levels (Figure 1). Initially, it was hypothesized that visual attention is significantly associated with SA, particularly in specific scenarios such as poor visibility. This hypothesis has been confirmed in a previously accepted paper, which also specifies the participants, situations, areas of interest (AOIs), and experimental procedures [28]. Expanding on the utilization of the eye-tracking technique and the SART questionnaire, the SA level groups were established as independent variables, with eye movement metrics designated as dependent variables. To guarantee the questionnaire’s professionalism and measurement accuracy, safety engineering and management procedures were employed for SART questionnaire validation. This process included regulatory oversight from maritime authorities and input from seasoned pilots. Subsequently, heatmaps and scan paths were used to illustrate the visual distribution of pilots across various areas of interest (AOIs) as part of an initial cognitive analysis. Permutation simulation was then utilized to verify eye movement metrics that exhibited a significant correlation with SA during ship pilotage. The eye movement metrics showing significant differences were divided into testing and training sets, serving as input for the random forest-support vector machine (RF-SVM) algorithm. This innovative approach allows for the categorization of SA levels to screen pilots and has been preliminarily validated in a previously accepted paper [29].

In the aforementioned simulation experiments, the collection of synchronous real-time eye-tracking data was the initial step. However, due to the susceptibility of the eye tracker to illumination and pilots’ head movements, the data often contain significant outliers, leading to identification errors [30, 31]. To effectively reduce the influence of noise on performance when employing wearable eye-tracking technology, a SA recognition model was created using the random forest-support vector machine (RF-SVM) algorithm. This model consists of modules for data input, SVM, modified RF, and verification [32]. RF utilizes a voting integration approach based on decision tree classifier
predictions. As an ensemble learning technique, RF is known for its accuracy and robustness in recognizing noise and outliers [33]. On the other hand, SVM is a machine learning approach based on statistical theory that offers distinct advantages in addressing small-sample recognition and high-dimensional nonlinear pattern recognition challenges [34]. Utilizing the RF-SVM approach we introduced, pertinent eye-tracking feature sets can be derived by taking into account the feature importance sequence in RFs as input for SVMs. The validation module assesses the accuracy in identifying the SA level of pilots, as described by [35].

2.1. Data Analysis. During the data collection procedures, the Tobii Glasses 2, a wireless wearable eye-tracking device, was employed to gather eye movement data from 25 ship pilots. The pilots engaged in ship piloting tasks in a bridge simulator for a minimum of 40 min, including over 25 min of poor visibility conditions, while being monitored by the eye-tracking device. This process unfolded in three stages: first, the calibration phase of the device was conducted before the experiment; subsequently, the testing phase was carried out throughout the entire piloting task; and finally, posttest interviews were conducted to validate the SA measurement results.

To analyze eye movement features at different levels of SA, preliminary extraction and analysis of eye-tracking features were conducted. The eye-tracking device was initially used to gather data, including pupil diameter and fixation count and duration, as well as saccade count and duration. Figure 2 displays the eye-tracking data spanning a 50 s duration, while in practice, the device records data at a sampling rate of 50 Hz, resulting in the collection of approximately 2,496,000 samples. However, the limited spatiotemporal sampling capabilities of eye-tracking devices pose restrictions on acquiring visual information from the peripheral environment. To address this, interpolation was employed to fill in missing data points, and noise reduction was achieved through filtering. After filtration, outliers were effectively removed from the gaze data fragment, which can be found in Figure 2.

Then, there was a fast drop in fixation accuracy upon deviation of the sight line from the central vision field. Accordingly, the gaze frequency-based recognition of eye-tracking types was accomplished by the velocity-threshold identification (I-VT) approach [36]. The gaze data underwent classification using the I-VT algorithm, which segregated the data into fixation samples (below a certain threshold) and saccade samples (exceeding the threshold). The coordinates of the fixation samples were determined in relation to the visual perspective of the subjects. Gaze data with fixation times ranging from 50 to 600 ms and a frequency range greater than 3 Hz were selected. The threshold for setting the velocity of ocular movements was established at 30°/s. In this way, visual behaviors could be identified with
the utilization of eye-tracking data based on registered coordinates, which, as displayed in Figure 2, stood for smooth fixation and fluctuant saccade coordinates. In addition, since the gaze data location was evident, the quantity and duration computation of fixations and saccades was possible.

With respect to pupil processing, the lowest acquisition parameter of the eye-tracking device was assigned as a default 2 mm value. In addition, linear interpolation was applied to address the presence of a few extreme or missing values in the raw data [37]. Figure 3 presents the processed data. On the whole, through integration of the eye-tracking devices’ output types and the identification and calculation techniques, the classification of signal features into fixation, saccade, and noise types was possible. Noise constituted 16.8% (equivalent to 87.6 minutes) of the total experimental duration, falling within the expected range.

2.2. Modified RF Module. RFs, a machine learning approach, are efficient tools for classification and assessment tasks [38]. In this investigation, training subsets were created using the bootstrap sampling method from preprocessed samples that had undergone interpolation and filtration [29]. For each subset \( (S_1, S_2, \ldots, S_k) \), a decision tree model was established, and the classification outcomes were assessed using the majority voting principle. The RF algorithm aimed to create a decision tree that depended on a stochastic variable \( \theta \), using the data sample \( X \) and the recognition variable \( Y \) as inputs. Suppose that the recognition result of a single decision tree classifier \( h(x, \theta_k) \) is \( h_i(X) \). The formula for the model’s final recognition result is as follows:

\[
H(X) = \frac{1}{k} \sum_{i=1}^{k} h_i(X).
\] (1)

To evaluate the importance of features in RF, the inclusion of noise in a specific feature was considered, and a significant decrease in recognition accuracy was taken into account [39]. The importance of eigen-parameters was assessed using the residual mean square for the out-of-bag (OOB) score during the computational process. The formula for calculating the OOB score is as follows:

\[
\text{importance} = \frac{\sum (\text{errOOB}2 - \text{errOOB}1)}{N_{\text{tree}}}. \tag{2}
\]

The formulas for \( \text{errOOB}1 \) and \( \text{errOOB}2 \) represent the OOB recognition errors before and after introducing noise interference into all sample features. Model stability was guaranteed through a grid search methodology aimed at determining the most suitable variables. To avoid overfitting of the feature data, we employed the root mean square error (RMSE). The relevant computational formula is provided as follows:

\[
\text{MSE} = \frac{\sum_{i=1}^{n}(y_{\text{obs}}^i - y_{\text{pred}}^i)^2}{\sum_{i=1}^{n}(y_{\text{obs}}^i - \bar{y}_{\text{pred}}^i)^2}, \tag{3}
\]

where \( y_{\text{obs}} \) and \( y_{\text{pred}} \) separately denote the observed and forecasted values of the corresponding samples, respectively, and \( n \) stands for the sample quantity. In addition, to address the potential issue of partial information overlaps that may arise from initial feature correlations, the optimal combination of features was obtained using principal component analysis (PCA) as follows:

\[
\begin{align*}
F_1 &= e_{11}X_1 + e_{12}X_2 + \cdots + e_{1z}X_z \\
F_2 &= e_{21}X_1 + e_{22}X_2 + \cdots + e_{2z}X_z \\
&\vdots \\
F_z &= e_{z1}X_1 + e_{z2}X_2 + \cdots + e_{zz}X_z
\end{align*}
\] (4)
where \( e_1^2 + e_2^2 + \cdots + e_z^2 = 1 \), \( i = 1, 2, \ldots, z \). No correlation is noted between the 2 principal components, that is, \( F_i \neq F_j \) \((i \neq j; i, j = 1, 2, \ldots, z)\). The initial principal component, denoted as \( F_1 \), signifies the linear combination that demonstrates the most significant variation among the entire sequence of initial importance values \((X_1, X_2, \ldots, X_z)\). On the other hand, the second principal component, \( F_2 \), is a linear combination of \( X_1 - X_z \), independent of \( F_1 \), and demonstrates the next-largest difference from the initial sequence. Likewise, the remaining principal components were identified as new input sets for the SVM.

\[ f(x) = w^T \varphi(x) + b, \]

where \( w^T \) is the weight, \( \varphi(x) \) is the kernel function, and \( b \) is an offset term. Evaluating the precision of the classification procedure is of utmost importance. To minimize the error, an insensitive loss function with two relaxation factors (\( \xi \) and \( \xi^* \)) was introduced. Assuming that all training samples are completely separated with an accuracy of \( \varepsilon \), the classification condition is as follows:

\[ y_i - w^T x_i - b \leq \varepsilon + \xi_i, \]
\[ w^T x_i + b - y_i \leq \varepsilon + \xi^*_i. \]

To optimize the SVM classification function, equation (6) was reformulated as a minimization problem. Subsequently, the Lagrange equation method, along with the dual principle, was utilized to transform the minimization issue into a dual optimization problem. The expressions for the minimization and dual optimization problems are as follows:

\[ \min \frac{1}{2} ||w||^2 + C \sum_{i=1}^N (\xi_i + \xi^*_i), \]
\[ \min \frac{1}{2} \sum_{i,j=1}^N (\alpha_i - \alpha^*_i)(\alpha_j - \alpha^*_j)x_i x_j + \sum_{i=1}^N (\alpha_i - \alpha^*_i)y_i - \sum_{i=1}^N (\alpha_i - \alpha^*_i)\varepsilon, \]
where C represents the penalty coefficient and $\alpha_i$, $\alpha_i^*$, $\alpha_j$, and $\alpha_j^*$ are Lagrangian multipliers.

**Step 2.** Validation of the kernel function: In practice, the training samples often do not meet the requirement for linear separability. To address the classification of nonlinear features, a kernel function was introduced to map the training samples from the original space to a high-dimensional Hilbert feature space. This process enabled the determination of a discriminative hyperplane with the greatest margin between categories. By reducing the feature dimensionality, a nonlinear classification boundary could be established. In this investigation, the selected kernel function was the Gaussian radial basis function (RBF). This offers good generalization ability and has the kernel bandwidth ($\sigma^2$) as the main parameter. The RBF can be expressed as follows:

$$K(x_i, x_j) = \exp\left(-\frac{\|x - x_i\|^2}{2\sigma^2}\right). \quad (8)$$

**Step 3.** Parameter optimization of SVM and voting integration: The two important parameters of the Gaussian RBF kernel are the penalty coefficient C and the kernel bandwidth $\sigma^2$. To enhance the learning and generalization capabilities of the SVM method, these parameters were optimized using a grid search approach. Consequently, the SVM model obtained, based on the optimal parameters, yielded the classification equation shown in equation (9). The final identification result was determined through the voting integration of the base classifiers.

$$f(x) = \sum_{i=1}^{N}(\alpha - \alpha^*)K(x_i, x) + b. \quad (9)$$

### 2.4. Verification Module.

The test dataset, containing samples with indeterminate categories, produced a confusion matrix that included true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN) as the model’s output. To assess the recognition performance of the R-SVM approach, a comparison was made with RF and SVM without optimized feature combinations. The assessment criteria for each classifier included the true positive rate (TPR), true negative rate (TNR), and general accuracy (ACC). The computational formulas for these metrics were as follows:

$$\text{ACC} = \frac{TP + TN}{TP + TN + FP + FN} \times 100\%,$$

$$\text{TPR} = \frac{TP}{TP + FN} \times 100\%,$$

$$\text{TNR} = \frac{TN}{TN + FP} \times 100\%. \quad (10)$$

In these formulas, TP represents samples where both the observed and forecasted values are 1. FP and TN refer separately to the samples with observed values of 0 and 1 and predicted values of 1 and 0. FN represents samples where both the observed and forecasted values are 0.

### 3. Results

To construct the SA recognition model using eye-tracking metrics, groups with different levels of SA were initially created. For the purpose of analysis, pilots’ SA was hypothetically divided into two levels on the basis of their SART score: high (>mean SART score) and low (<mean SART score). We divided the pilots into either the high-SA group ($n = 13$; mean = 24.5; and standard deviation = 5.13) or the low-SA group ($n = 12$; mean = 15.2; and standard deviation = 4.37) depending on the SART scores (mean = 20.13 and standard deviation = 5.83).

#### 3.1. AOI Analysis.

To compare the allocation of attention and scanning techniques among subjects in different situation awareness (SA) groups, we conducted an initial assessment of pilot visual behavior under conditions of poor visibility. This assessment was based on thermograms and scan paths of eye-tracking metrics. The thermogram evaluation revealed two key findings: during marine pilotage, participants primarily focused on the electrical chart (AOI-1) and outside the window (AOI-2). Notably, significant differences were observed in the time spent fixating on the same AOIs across different SA groups, as illustrated in Figures 4(a) and 4(b). Specifically, participants in the low-SA group allocated more attention to AOI-1, while those in the high-SA group directed their focus more toward AOI-2. These observations indicate a potential link between the selective distribution of attention between different AOIs and the SA level of pilots.

Furthermore, the findings of the scan path analysis showed that AOI-2 was the main area where pilots scanned back and forth. Interestingly, the high-SA group exhibited a significantly higher scanning frequency than the low-SA group (Figures 4(c) and 4(d)). This can be attributed to the fact that in real-life situations, the high-SA population requires frequent confirmation of perceptual elements and timely updates of their mental model to accurately forecast and anticipate behavior. Taken together, these findings provide evidence for a probable relationship between scanning strategies and the level of SA in poor visibility situations.

#### 3.2. Correlation Evaluation.

To avoid confusion about the influence of varying AOIs on the significant disparities in the eye-tracking metrics and SA levels of pilots, a statistical evaluation was undertaken by emphasizing AOI-1 and AOI-2. We selected these two AOIs because they represented the main points of fixation and saccadic movement during the poor visibility scenario, as indicated by the eye-tracking
The association between eye-tracking metrics and SA levels in AOI-1 and AOI-2 was analyzed individually using permutation simulations, as outlined in Table 1. The statistical analysis results for AOI-1 demonstrated an obvious correlation between SA level and fixation count ($p = 0.024 < 0.05$), fixation duration ($p = 0.036 < 0.05$), saccade count ($p = 0.075 < 0.1$), and saccade duration ($p = 0.05 \leq 0.05$). The obtained findings suggest that high-SA pilots exhibited more visual engagement. Likewise, the AOI-2 analysis revealed that the SA level was linked prominently to 3 eye-tracking metrics: fixation quantity ($p = 0.034 < 0.05$), saccade time ($p = 0.075 < 0.1$), and fixation time ($p = 0.047 < 0.05$). Nevertheless, pupil diameter was not significantly linked to SA, probably because of individual disparities.

Considering the adverse influence of reduced visibility on pilots’ ability to gather real-time environmental information through scene scanning, it is worth mentioning that the group with low SA exhibited the longest average fixation duration in AOI-1. This suggests that the low-SA group might have concentrated their attention on acquiring primary perceptual information in AOI-1. In contrast, the average fixation time of the high-SA group was longer in AOI-2 than in AOI-1, indicating that the priority of this group was still the necessary feedforward information processing in AOI-2. Considering the correlation evaluation between AOI-1 and AOI-2 under the poor visibility scenario, the foregoing findings provide potential for future exploration into the SA level detection of pilots with the utilization of relevant fixation and saccade metrics in such a setting. In addition, data segmentation was performed by utilizing a sliding time window, set at 5 s (i.e., epoch length) to minimize data noise and volume. The computational results obtained from this segmentation were regarded as the features for selection in the subsequent recognition model, which can be found in Table 2.

3.3. SA Identification. In the present research, the state of cognition was categorized into high- and low-SA levels by introducing a nonlinear RF-SVM algorithm, which exploited the eye-tracking data derived from a marine piloting experiment. Following the data preprocessing via filtration and interpolation, the classification of features proceeded into 9 groups, as detailed in Table 2. Next, the characteristics were assigned to the testing and training sets, with the letter accounting for 75% of the whole samples. The RF-SVM approach was formulated in accordance with this training set. The best parameters for the RF model were identified through the grid search method, as depicted in Figure 5. A maximum search efficiency of 0.9252 was achieved by configuring the model with 151 estimators and a maximum depth of 20. Subsequently, verification of optimal RF variables was carried out; in addition, importance ranking of initial features was accomplished according to the average score (Table 3).

The feature with the minimum importance score in Table 3 was discarded so that the recognition model overfitting could be avoided. Subsequently, we fed back the remaining features into the RF model. The repetition of such a feature screening process was accomplished for 8 iterations, and the RMSE and relative error (RE) of 1–8 features were estimated, as detailed in Table 4. According to the obtained findings, the RMSE and RE were minimized when using six features. Consequently, the features ND, SC, and MN were excluded, leaving six valid features. Furthermore, to address potential information overlap in the visual...
features from the same participants, dimension reduction was performed using the PCA algorithm. Table 5 lists the eigenvalues for the correlation coefficient matrix, as well as the contributions of various principal components. The initial four principal components were chosen as the extraction criterion, as they collectively accounted for over 90% of the variance rate. These four principal components are as follows:

Table 1: Correlation results in poor visibility.

<table>
<thead>
<tr>
<th>Eye-tracking metrics</th>
<th>High SA</th>
<th>Low SA</th>
<th>Permutation results</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>Std</td>
<td>Mean</td>
</tr>
<tr>
<td>AOI-1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fixation count</td>
<td>9.096</td>
<td>1.063</td>
<td>15.042</td>
</tr>
<tr>
<td>Fixation duration</td>
<td>6.28</td>
<td>2.915</td>
<td>19.595</td>
</tr>
<tr>
<td>Saccade count</td>
<td>0.445</td>
<td>0.323</td>
<td>0.674</td>
</tr>
<tr>
<td>Saccade duration</td>
<td>0.449</td>
<td>0.322</td>
<td>0.674</td>
</tr>
<tr>
<td>Pupil diameter</td>
<td>4.214</td>
<td>0.775</td>
<td>4.356</td>
</tr>
<tr>
<td>AOI-2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fixation count</td>
<td>19.911</td>
<td>7.567</td>
<td>5.282</td>
</tr>
<tr>
<td>Fixation duration</td>
<td>17.563</td>
<td>10.156</td>
<td>3.476</td>
</tr>
<tr>
<td>Saccade count</td>
<td>13.404</td>
<td>8.256</td>
<td>4.216</td>
</tr>
<tr>
<td>Saccade duration</td>
<td>0.724</td>
<td>0.339</td>
<td>0.723</td>
</tr>
<tr>
<td>Pupil diameter</td>
<td>4.214</td>
<td>0.826</td>
<td>4.348</td>
</tr>
</tbody>
</table>

*p < 0.1 and *p < 0.05. The bold values given in Table 1 indicate p values corresponding to eye-tracking metrics associated with SA levels.

Table 2: List of calculated features.

<table>
<thead>
<tr>
<th>Signal type</th>
<th>Features</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fixation</td>
<td>Median fixation duration in 5 s (MF)</td>
</tr>
<tr>
<td></td>
<td>Fixation duration in 5 s (FD)</td>
</tr>
<tr>
<td></td>
<td>Fixation count in 5 s (FC)</td>
</tr>
<tr>
<td>Saccade</td>
<td>Median saccade duration in 5 s (MS)</td>
</tr>
<tr>
<td></td>
<td>Saccade duration in 5 s (SD)</td>
</tr>
<tr>
<td></td>
<td>Saccade count in 5 s (SC)</td>
</tr>
<tr>
<td>Noise</td>
<td>Median noise duration in 5 s (MN) and noise duration in 5 s (ND)</td>
</tr>
<tr>
<td></td>
<td>Noise frequency in 5 s (NF)</td>
</tr>
</tbody>
</table>

Figure 5: Parameter optimization of the RF method.
Extraction of principal components \((F_i)\) was accomplished through feature combinations. New input sets were generated using these principal components for the SVM method. The optimization of SVM parameters was performed using the grid search technique, and the results are illustrated in Figure 6. A maximum efficiency score of 0.9328 was attained with \(C = 6\) when the bandwidth \(g\) of the kernel function was 0.005. After calculating the base classifiers for the SVM, the optimal identification result was obtained through voting integration.

By comparatively analyzing RF-SVM, RF, and SVM, the optimized feature combinations were assessed for validity with 3 performance evaluation metrics. Figure 7 illustrates the distributions of classification precision measures for the 3 classifiers that vary in performance metrics. The central mark in each box indicates the median, with the boxes representing the 25th and 75th percentiles. When utilizing the best eye-tracking feature combination in the RF-SVM algorithm, the average accuracy (ACC) across 100 iterations reached 0.934, with a TNR of 0.875 and a TPR of 0.940. Overall, the RF-SVM outperformed the SVM and RF models without feature optimization.

Since receiver operating characteristic (ROC) curves can be plotted to visualize the accuracy of the classification method, the ROC combined with the area under the curve (AUC) is usually used to solve the evaluation problem of binary classification [41]. The RF-SVM algorithm was retrained and retested using the optimal parameters, and the ROC graphs with the AUC were analyzed for comparative purposes. Figure 8 illustrates the results, with an AUC score of 0.894 for RF, 0.907 for SVM, and 0.942 for RF-SVM, indicating the higher stability of RF-SVM. TPR was used to confirm the superior sensitivity of RF-SVM, while its specificity was supported by the TNR. Performance

### Table 3: Initial feature importance score and ranking.

<table>
<thead>
<tr>
<th>Number</th>
<th>Feature</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>FD</td>
<td>0.248</td>
</tr>
<tr>
<td>2</td>
<td>SD</td>
<td>0.199</td>
</tr>
<tr>
<td>3</td>
<td>FC</td>
<td>0.135</td>
</tr>
<tr>
<td>4</td>
<td>NF</td>
<td>0.133</td>
</tr>
<tr>
<td>5</td>
<td>MF</td>
<td>0.093</td>
</tr>
<tr>
<td>6</td>
<td>MS</td>
<td>0.090</td>
</tr>
<tr>
<td>7</td>
<td>MN</td>
<td>0.040</td>
</tr>
<tr>
<td>8</td>
<td>SC</td>
<td>0.034</td>
</tr>
<tr>
<td>9</td>
<td>ND</td>
<td>0.027</td>
</tr>
</tbody>
</table>

### Table 4: Error values with varying feature quantities.

<table>
<thead>
<tr>
<th>Number of features</th>
<th>RMSE</th>
<th>RE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.912</td>
<td>0.476</td>
</tr>
<tr>
<td>2</td>
<td>0.843</td>
<td>0.427</td>
</tr>
<tr>
<td>3</td>
<td>0.882</td>
<td>0.391</td>
</tr>
<tr>
<td>4</td>
<td>0.787</td>
<td>0.340</td>
</tr>
<tr>
<td>5</td>
<td>0.704</td>
<td>0.283</td>
</tr>
<tr>
<td>6</td>
<td>0.674</td>
<td>0.247</td>
</tr>
<tr>
<td>7</td>
<td>0.756</td>
<td>0.263</td>
</tr>
<tr>
<td>8</td>
<td>0.722</td>
<td>0.255</td>
</tr>
</tbody>
</table>

### Table 5: Principal components of features.

<table>
<thead>
<tr>
<th>Component</th>
<th>Eigenvalues</th>
<th>Contribution (%)</th>
<th>Cumulative contribution (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(F_1)</td>
<td>6.773</td>
<td>53.213</td>
<td>53.213</td>
</tr>
<tr>
<td>(F_2)</td>
<td>3.349</td>
<td>20.297</td>
<td>73.510</td>
</tr>
<tr>
<td>(F_3)</td>
<td>1.568</td>
<td>12.447</td>
<td>85.957</td>
</tr>
<tr>
<td>(F_4)</td>
<td>1.012</td>
<td>4.643</td>
<td>90.597</td>
</tr>
</tbody>
</table>

\[
\begin{align*}
F_1 &= 0.0431X_1 + 0.1408X_2 + 0.3083X_3 + 0.3241X_4 + 0.0787X_5 + 0.2672X_6, \\
F_2 &= 0.2293X_1 + 0.3341X_2 + 0.2692X_3 + 0.1042X_4 + 0.1782X_5 + 0.0763X_6, \\
F_3 &= 0.0744X_1 + 0.1988X_2 + 0.0835X_3 + 0.2765X_4 + 0.3038X_5 + 0.0243X_6, \\
F_4 &= 0.3112X_1 + 0.3937X_2 + 0.1108X_3 + 0.3681X_4 + 0.1186X_5 + 0.3412X_6, \\
F_i &= 0.5874F_1 + 0.2240F_2 + 0.1374F_3 + 0.0512F_4.
\end{align*}
\]
information for the three classification algorithms based on the assessment methodology is presented in Table 6. The RF-SVM algorithm, using optimized features as input data, achieved an average accuracy of 0.934, an average sensitivity of 0.940, an average specificity of 0.876, and an AUC score of 0.942. The obtained findings suggest that RF-SVM, with optimized parameters, outperforms classical models in recognizing eye movement features associated with different levels of SA. This provides valuable insights for developing a screening and assessment model for pilot competency.

4. Discussion

Despite the numerous advantages of physiological measures, such as uninterrupted and objective properties, in comparison to direct subjective assessments, there have been limited studies investigating the assessment of SA in pilotage tasks using physiological measurement techniques. In this study, we explored an SA recognition method for pilots based on eye-tracking data gathered during a bridge simulation experiment. The SA recognition model based on the
RF-SVM algorithm demonstrates effective SA recognition using four principal components derived from eye-tracking data (Table 5). In this study, eye-tracking data were collected from 25 pilots during low-visibility scenarios to evaluate their competency. The significance of the correlations between visual behaviors (represented by thermograms and scan paths) and SA level (determined by SART) was quantified through a permutation test. Relevant eye-tracking features were derived from pertinent metrics and divided into 5-second segments following interpolation and filtration to prevent overfitting. The selection of the top 6 significant features was based on their RE and RMSE, and feature combinations in the RF model were determined through PCA. The next step was precise RF-SVM-based identification of the at-risk cognitive state, that is, a low-SA level. As revealed by the comparison among the 3 performance metrics, the performance of our model was superior to that of the remaining 2 models that were devoid of feature optimization.

The findings of the current work demonstrate that effective identification of the at-risk cognitive state is possible by utilizing eye-tracking features within a well-designed recognition framework that offers high accuracy, sensitivity, and specificity. Figure 7 presents the recognition results obtained from RF-SVM, RF, and SVM utilizing different types of features. It is evident that the classification accuracy significantly improved to 93.43% (RF-SVM) when we used a combination of the top 6 most significant features, in contrast to 86.79% for RF without feature optimization. When the same feature data were employed as inputs, a comparison between RF and SVM revealed that SVM outperformed RF in terms of ACC, TPR, and TNR. This highlights the suitability of SVM when dealing with a small-sample size of nonlinear eye-tracking data in a low-visibility environment. Hence, an effective RF-SVM model for detecting the at-risk cognitive state needs to comprise a strong classifier (i.e., SVM module), as well as the input data optimized with a combination of salient eye-tracking features (i.e., modified RF module).

It should be noted that a precise comparison of results with other studies that utilize eye-tracking is challenging due to variations in the simulation approach, eye-tracking feature selection, and classification of SA groups across different studies. In addition, a limitation of this study is the performance issue with the eye-tracking devices. In situations where pilots exhibit rapid movements, data collection may be hindered by the slower acquisition speed of the eye-tracking device, leading to a reduction in local data. While data interpolation can assist in addressing this concern, it may inevitably affect the significance of correlation results. Moreover, in terms of feature extraction, the accuracy of recognition is influenced by the duration of the data segments (epochs) and the correlation metrics of eye-tracking data in conditions of poor visibility. Moreover, the present study provides evidence that ship pilots with high levels of SA tend to exhibit more active visual behaviors, as indicated by nine eye-tracking features with epoch lengths of 5 s. However, as previously demonstrated in other studies [42–44], SA levels are not only associated with eye movement features but also with other physiological measurements. Therefore, recognizing at-risk cognitive states is a complex process that warrants further investigation, considering the fusion metrics including True Positive Rate

Table 6: Comparative analysis of performance metrics.

<table>
<thead>
<tr>
<th>Classification algorithm</th>
<th>Accuracy</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>AUC</th>
</tr>
</thead>
<tbody>
<tr>
<td>RF</td>
<td>0.869</td>
<td>0.883</td>
<td>0.837</td>
<td>0.894</td>
</tr>
<tr>
<td>SVM</td>
<td>0.895</td>
<td>0.895</td>
<td>0.877</td>
<td>0.907</td>
</tr>
<tr>
<td>RF-SVM</td>
<td>0.934</td>
<td>0.940</td>
<td>0.876</td>
<td>0.942</td>
</tr>
</tbody>
</table>

Figure 8: ROC curve of the three methods.
electroencephalogram (EEG) and heart rate variability (HRV) signals with varying epoch lengths, as well as the classification of SA groups using diverse approaches [29]. With the development of sensing technology, the optimization of portable devices provides the possibility of data acquisition and filtering for the pilots during the actual tasks [45]. The current research results are a prior exploratory study of using physiological indicators to monitor the pilotage process, and subsequent research on behavioral pattern recognition by fusing EEG can provide technical support for auxiliary decision-making of intelligent navigation.

5. Conclusion

Situation awareness (SA) plays a crucial role in marine safety, as the lack of SA can contribute to approximately 75% of maritime accidents caused by human error. In contrast to direct SA assessment methods, physiological measurement techniques, such as eye-tracking, have the potential to provide objective and continuous SA evaluation in pilotage tasks. Nevertheless, it remains unclear how to infer SA using eye movement features. This study conducted a bridge simulation experiment for ship piloting to examine the relationship between eye-tracking features and SA, as assessed by a SART. In addition, we developed an RF-SVM model to identify pilots’ SA based on eye-tracking metrics.

The results obtained with our RF-SVM algorithm provide evidence for its effectiveness in recognizing at-risk cognitive states, specifically low levels of SA, based on eye-tracking features. Our identification model, which includes modified RF, SVM, and verification modules, was applied to eye-tracking data collected from 25 ship pilots during a bridge simulation experiment conducted under poor visibility conditions. By employing permutation simulations and PCA for RMSE and RE rectification, we determined that the optimal eye-tracking features consist of four principal components. The performance of our proposed feature combinations surpasses that of RF and SVM without feature optimization, revealing the potential of our approach in computer-assisted screening of cognitive competency for ship pilots. The outcomes of this study could establish a theoretical foundation for the utilization of physiological measurement techniques in assessing situation awareness (SA) and, in turn, aid in the monitoring and evaluation of pilots’ competencies. Moving forward, it will be important to validate and apply our recognition model in various emergency scenarios, such as ship departure, anchoring, and encounters, using multiple fusion metrics such as HRV and EEG. By doing so, we can not only benefit from immediate improvements in cognitive state surveillance and prevention of unsafe behavior but also pave the way for developing comprehensive systems to assess the physical and mental competencies of ship pilots.

Data Availability

Data are available from the corresponding author upon request.

Ethical Approval

All experimental procedures involving human subjects were implemented following the ethical standards of the institutional and/or national research committee(s), as well as the 1964 Helsinki Declaration and its later revisions or analogous ethical standards.

Consent

Every individual subject in the present work, including their statutory guardians, signed an informed consent form.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported by the Jinhua Municipal Project for Public Welfare Technology Application Research (Grant No. 2023-4-033) and the National Natural Science Foundation of China (Grant No. 71503166).

References


