ROYAL
AERONAUTICAL
SOCIETY

REGULAR PAPER

A methodology to detect pilot perception of warning information by eye movement data and deep residual shrinkage networks

C.-Q. Yan[1] , Y.-C. Sun[1,*] , X. Zhang[1] , H.-Y. Mao[2] and J.-Y. Jiang[1]

[1]College of Civil Aviation, Nanjing University of Aeronautics & Astronautics, 211106 Nanjing, China and [2]China Aeronautical Radio Electronics Research Institute, 200241, Shanghai, China
*Correspondence author. Email: sunyc@nuaa.edu.cn

Abstract

This paper studied the use of eye movement data to form criteria for judging whether pilots perceive emergency information such as cockpit warnings. In the experiment, 12 subjects randomly encountered different warning information while flying a simulated helicopter, and their eye movement data were collected synchronously. Firstly, the importance of the eye movement features was calculated by ANOVA (analysis of variance). According to the sorting of the importance and the Euclidean distance of each eye movement feature, the warning information samples with different eye movement features were obtained. Secondly, the residual shrinkage network modules were added to CNN (convolutional neural network) to construct a DRSN (deep residual shrinkage networks) model. Finally, the processed warning information samples were used to train and test the DRSN model. In order to verify the superiority of this method, the DRSN model was compared with three machine learning models, namely SVM (support vector machine), RF (radom forest) and BPNN (backpropagation neural network). Among the four models, the DRSN model performed the best. When all eye movement features were selected, this model detected pilot perception of warning information with an average accuracy of 90.4%, of which the highest detection accuracy reached 96.4%. Experiments showed that the DRSN model had advantages in detecting pilot perception of warning information.

Nomenclature

ANOVA	analysis of variance
BPNN	backpropagation neural network
CNN	convolutional neural network
DNN	deep neural network
DRN	deep residual network
DRSN	deep residual shrinking network
LR	logistic regression
ReLU	rectified linear unit
RF	random forest
ROC	receiver operating characteristic
SD	standard deviation
SVM	support vector machine

1.0 Introduction

With the development of technologies, the systems used for human-computer interaction, environmental perception, weapon use, countermeasures and flight control are becoming increasingly complex. The continuous advancement of airborne systems, while providing unprecedented useful information for pilots, seriously challenges their limited cognitive abilities by bringing about "information overload" in the brain. This has a negative impact on pilots' operational behaviours and situational awareness, which may result in the inability to perceive the existence of warning information, thus endangering flight safety and even causing accidents. Studies have shown that over 88% of aviation accidents involving human error were attributed to problems with situation awareness [1, 2]. Therefore, accurately identifying the pilot perception of warning information is of great significance for realising intelligent flight control, improving pilots' control behaviours and maintaining flight safety. This study proposed a method for detecting pilot perception of warning information using eye movement data based on machine learning models.

1.1 Display of warning information in an aircraft cockpit

Pilots rely on external information and equipment information to judge the current state of the aircraft and evaluate the subsequent flight control. Cockpit devices displayed the attitude, speed, altitude, warnings, and other information [3]. The ability of pilots to interpret the visual-spatial information and determine the orientation of the aircraft played an important role in flight safety [4]. In the process of perception of aircraft cockpit information, most unintentional behaviours can be attributed to the result of inappropriate cockpit human-computer interaction [5]. In the research of warning information display, it was found that there was a correlation between the warning information display of the equipment and the optimisation of the cockpit design. The significance of the equipment's warning information seemed to affect flight performance and attention distribution [6]. For example, different warning information designs had certain effects on the pilots' first gaze and reaction time [7]. At the same time, the pilots' rapid perception of warning information in the cockpit played a key role in maintaining situational awareness [8]. Li compared visual data from two different designs of crew warning systems interacting with pilots and determined that the different designs of crew warning systems affect pilot situational awareness [9]. Warning information was designed to help operators understand the situation and predict future states [10]. Current research mainly focused on the optimisation of cockpit design, and there were few comparative studies on different coded early warning designs. The main content of this paper was to compare the accuracy of different algorithm models to detect the pilot perception of warning information and the influence of different warning information designs on the accuracy of the model.

1.2 Application of eye movement data

The sources of information that pilots relied on were mainly the visual channel and the auditory channel, of which about 80–90% of the information came from the visual channel [11]. With the continuous development of aviation technology, the information that pilots need to deal with is also changing. For example, pilots no longer focus on visual information outside the cockpit as in the early days of aviation, but instead have to pull information from multiple sources (more instruments in the cockpit) to manage the flight. The manual processing of visual data remained one of the key elements of aviation safety and effectiveness [12]. Eye-tracking has been widely used in the ergonomics research of aviation cockpit. Eye tracking has been used to study potential factors affecting pilot attention and situational awareness in the cockpit since it provides pilot eye movement data [13]. When evaluating the relationship between humans and equipment through eye-tracking devices, eye movement data were used to analyse human behaviours. This hypothesis was confirmed by many previous researchers [14], such as mental fatigue [15], cognitive load [16], information acquisition [17], situational awareness [18],

scanning behaviours [19], attention [20], physiological measurements [21], workload [22, 23], interface evaluation and human-computer interaction [24, 25].Research on eye movement data that predicted pilots' behaviour included how different role assignments influenced decisions in high-risk environments, and how to predict pilots' decisions by paying attention to relevant information about pilots' choices [26]. When using eye movement to predict dangerous situations, Francisco [27] found that saccade, gaze, blink and gaze dispersion in horizontal and vertical dimensions in eye movement features were more likely to predict the occurrence of dangerous situations.

1.3 Application of machine learning model in eye movement data

Eye movement data can directly reflect pilots' situational awareness, cognitive load, attention distribution and decision prediction. At the same time, the eye movement data acquisition technology has been mature, and more and more algorithms and models based on eye movement data have been studied. In previous studies, eye movement data were used to train and test SVM and logistic regression (LR) models to detect pilot distraction. The results showed that the SVM model was superior to the traditional LR model [28]. Aiming at the multi-classification model of eye movement feature training, the classification models established by Decision Tree, K-Nearest Neighbour, Bayesian Network and SVM were compared, and the optimal model finally obtained was the SVM model with linear kernel [29]. In the user classification for the user's gaze data, using cross-validation validation and feature selection based on a hierarchical tree, the classification accuracy of the RF classifier was $0.88 \pm (0.11)$ [30]. A classification model of cognitive distraction assessment based on RF was constructed, and the highest accuracy was obtained by using gaze data, glance data and noise-related features in the model [31]. Cheng established a data set of eye movement to train CNN, SVM and BPNN for eye movement recognition. The results showed that CNN had the highest recognition rate [32].

In this paper, we built DRSNs-based model to identify pilot perception of warning information. A DRSN can be formed by stacking many basic modules. Each basic module had a sub network for automatic learning to obtain a set of soft thresholds of the feature map. Thus, each sample had its own unique set of thresholds. DRSN was used to process the data with a large amount of noise, and high accuracy was obtained, which verifies the effectiveness of this method [33]. Currently, DRSN was not used in the field of eye movement data analysis. The eye movement data collected in experiments were easily disturbed by the external environment, such as light, head shaking, etc., resulting in the data often containing noises, uncertain factors and incomplete information, so it was more suitable to use the DRSN model for analysis.

2.0 Methods

2.1 Subjects

Twelve male subjects were recruited to participate in this study. Their mean age was 38.2 years (SD = 4.1). All subjects had extensive experience in simulated helicopter flying, and the average simulated flight time was 3,160h (SD = 1126.4). Their binocular vision was normal, and they agreed to record their eye movement data during the experiment. Subjects were told they could stop the experiment at any time. The experiment was reviewed and approved by the Nanjing University of Aeronautics & Astronautics institutional review board. The experiments were carried out according to the Declaration of Helsinki. All participants provided written informed consent.

2.2 Equipment

The layout of the helicopter cockpit simulation platform was shown in Fig. 1. Flight control and display devices in the helicopter cockpit simulation platform included a cyclic control stick, a collective pitch
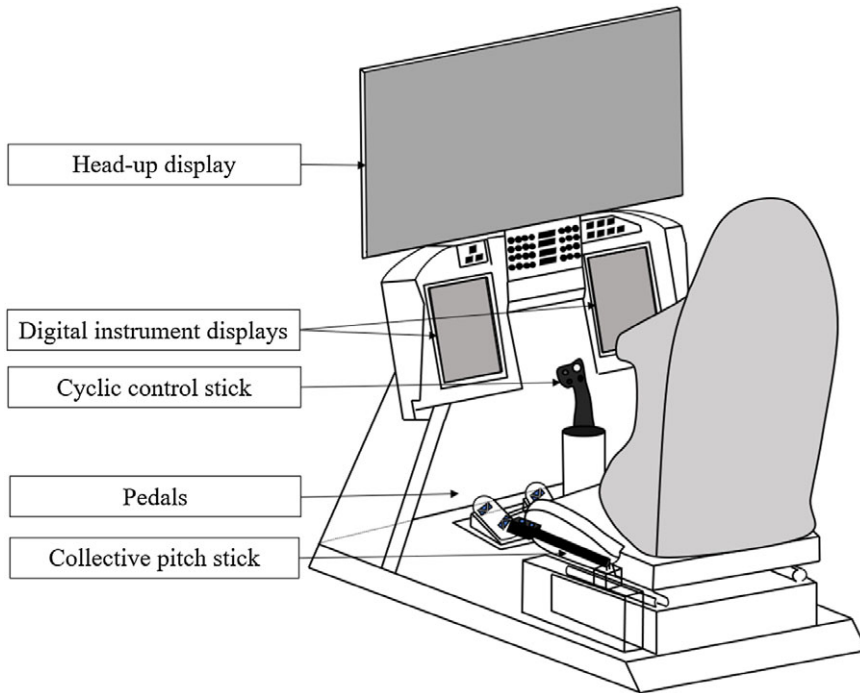
Figure 1. *Overall schematic diagram of helicopter cockpit simulation platform.*

stick, pedals, a head-up display and two digital instrument displays. The screen sizes of the visual display and the digital instrument display were 40 inches and 14 inches respectively, the resolutions were $1,920 \times 1,080$ and $1,366 \times 768$, and the refresh rates were both 60Hz. In the experiment, the helicopter model we selected from the helicopter cockpit simulation platform was Kamov Ka-52.

Eye-tracking used the SMI ETG 2W glasses-type eye-tracking devices. We used eye-tracking devices that allowed subjects to freely turn their heads and tracked a greater range than desktop fixed eye-tracking devices. Eye-tracking was accurate to 0.5° with a range of 80° horizontally and 60° vertically. Begaze SMI software (version 3.5) was used to analyse eye movement data.

2.3 Experimental design

Participants were required to fly the helicopter on the cockpit simulation platform to complete the flight task of the designated route, as shown in Fig. 2(a). Start at waypoint A, climb to 2,000 feet to start cruising and then descend to 1,000 feet at waypoint B. Then climb to 3,000 feet to start cruising and descend to 1,000 feet at waypoint C. Finally, climb to 2,500 feet to start cruising, descend before reaching waypoint A and land at waypoint A. It took about 15min to complete an airline flight. The English warning information in red font appeared in the lower right corner of the screen on the right side of the digital instrument display, with the content of "missile approaching", as shown in Fig. 2(b), and the duration was 15s. The warning information were divided into the presence or absence of flickering (bright 1s, dark 0.5s) in terms of vision, and the presence or absence of audible warnings (deep-deep-deep-deep, ring 0.35s, off 0.35s) in terms of hearing. The displays of warning information were divided into four encoding forms, as shown in Table 1. During the flight, each coded warning information appeared four times randomly. When participants noticed the warning information, they were required to report the occurrence of the warning information to the note-takers.

Table 1. *Details of the four warning coding types*

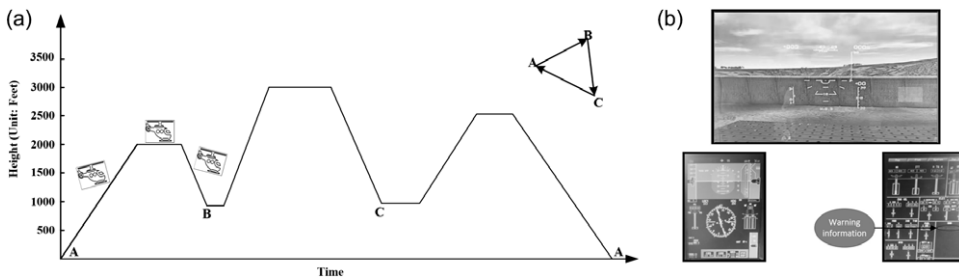| Characteristics of the warning information	Warning type number			
	A1	A2	A3	A4
Static warning information	*	*	*	*
Sound		*		*
Flicker			*	*

Figure 2. *Schematic diagram of the flight route and warning location.*

2.4 Procedure

Participants were carefully introduced to the research content and experimental environment. The participants then wore and calibrated eye-tracking devices and started flight tasks in sequence. Eye movement data were recorded by an eye-tracking device. Recording was started when participants started the flight task and stopped when all tasks were completed. Due to the limited number of personnel with sufficient flight experience, in order to collect sufficient experimental data, each participant repeated four flight tasks, and each flight had a rest time of 10min. But it should be noted that the type and appearance time of the warning information in each flight task were random. A total of 48 sets of experiments were conducted and 752 sets of valid eye movement data based on warning information were collected. Participants were paid after completing all flight tasks.

2.5 Data analysis

After sorting and processing the collected eye movement data, 25 eye movement features were obtained. According to the meaning of eye movement features, eye movement features were divided into eye features and visual features, as shown in Table 2. To compare the effect of warning information on various eye movement feature data, an analysis of variance was used to determine the level of importance of eye movement features. The ANOVA method ranked the importance of features based on the F value that is the ratio of the mean square between and the mean square within for each feature. The calculation formula of F was shown in Equation (1).

$$F = \frac{\left[n_1\left(\bar{x}_1 - \bar{x}\right) + n_2\left(\bar{x}_2 - \bar{x}\right) + \cdots + n_k\left(\bar{x}_k - \bar{x}\right)\right] \big/ (k-1)}{\left[\sum\left(\bar{x}_{1i} - \bar{x}\right)^2 + \sum\left(\bar{x}_{2i} - \bar{x}\right)^2 + \cdots + \sum\left(\bar{x}_{ki} - \bar{x}\right)^2\right] \big/ (n-k)} \tag{1}$$

The eye movement features were sorted according to their importance, and each eye movement feature was classified by Euclidean distance. The above ANOVA and classification were performed using IBM SPSS Statistics 22 software.

About correlation calculation and heat map generation, we used the Python software. By inputting eye movement data and running relevant programs, the relevant thermodynamic diagrams of each feature of eye movement data were obtained.

Table 2. *The serial number, abbreviation and meaning of each eye movement feature*

Categories	Feature No.	Abbreviations	Meanings
Eye features	1	CB	Category Binocular
	2	PSRX	Pupil Size Right X
	3	PSRY	Pupil Size Right Y
	4	PDR	Pupil Diameter Right
	5	PSLX	Pupil Size Left X
	6	PSLY	Pupil Size Left Y
	7	PDL	Pupil Diameter Left
	8	EPRX	Eye Position Right X
	9	EPRY	Eye Position Right Y
	10	EPRZ	Eye Position Right Z
	11	EPLX	Eye Position Left X
	12	EPLY	Eye Position Left Y
	13	EPLZ	Eye Position Left Z
Visual features	14	PRBX	Point of Regard Binocular X
	15	PRBY	Point of Regard Binocular Y
	16	PRRX	Point of Regard Right X
	17	PRLY	Point of Regard Right Y
	18	PRLX	Point of Regard Left X
	19	PRLY	Point of Regard Left Y
	20	GVRX	Gaze Vector Right X
	21	GVRY	Gaze Vector Right Y
	22	GVRZ	Gaze Vector Right Z
	23	GVLX	Gaze Vector Left X
	24	GVLY	Gaze Vector Left Y
	25	GVLZ	Gaze Vector Left Z

For a traditional deep learning model, the more layers of the network, the stronger the corresponding nonlinear expression ability and the more features learned by the model. However, with the increase of the number of network layers, it was difficult for the nonlinear expression of the traditional multi-layer network structure to represent the identity mapping, so the model may suffer from network degradation problem. Noise interference was common in eye movement data, which affected the accuracy of the model to identify the pilot perception of warning information. To address the above problems, this paper proposed a method of apply DRSN model to eye movement data. The DRSN model was able to overcome the difficulty that traditional learning models cannot achieve identity mapping on nonlinear transformations when training data samples in deep networks. At the same time, the interference of noise data samples and redundant data samples on feature threshold extraction was suppressed.

When DRSN was based on back-propagation for model training, its loss was not only back-propagated layer-by-layer through convolutional layers, etc., but also back-propagated more conveniently through the identity mapping of residual terms. Then, soft threshold was used to denoise the data, and a better model was obtained [33].

It was assumed that the required solution mapping was $H(x^l)$, and this problem was transformed into the residual mapping function $F(x^l)=H(x^l)-x^l$ for solving the network. Compared with the ReLU function, the soft threshold was more flexible to set the eigenvalue interval. In the residual shrinkage network, the threshold was automatically adjusted according to the situation of the sample itself through the attention mechanism. A Part of the DRSN model was shown in Fig. 3, where the size of the input x^l was C × N, and after passing through the hidden layer 1, the ReLU function was used to obtain x^{l+1} as the input of the hidden layer 2. In the hidden layer 2, a small sub-network was constructed to learn a set
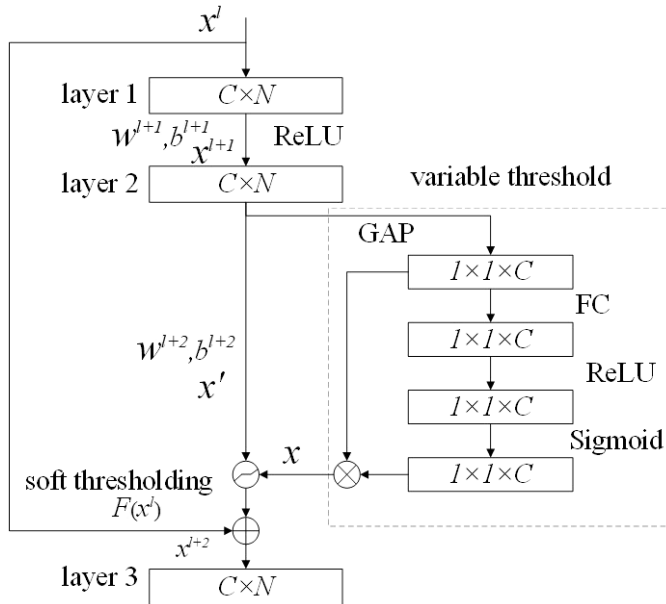
Figure 3. *Local frame diagram of DRSN model.*

of thresholds γ between 0 and 1, and then the soft thresholding of the features was performed and the residual term $F(x^l)$ was added to obtain the output x^{l+2}. The output of each layer was as follows:

$$x^{l+1} = \mathrm{Re}LU\left(w^{l+1}x^l + b^{l+1}\right)$$
$$x' = w^{l+2}\mathrm{Re}LU\left(w^{l+1}x^l + b^{l+1}\right) + b^{l+2} \tag{2}$$

In Equation (2), w and b were the weight vector and the bias vector, respectively.

Equation (3) was the soft thresholding result obtained by comparing x' each dimension with the corresponding threshold γ.

$$x' = \begin{cases} x' - \gamma & x' > \gamma \\ 0 & -\gamma \le x' \le \gamma \\ x' + \gamma & x' < -\gamma \end{cases}$$
$$x^{l+2} = x' + F\left(x^l\right) \tag{3}$$

At the same time, to compare the superiority of DRSN algorithm model, according to the reading and induction of literatures, three machine learning algorithm models widely used in eye movement data were selected for comparison, namely SVM, RF and BPNN. For the four machine learning algorithm models, when selecting the eye movement data features, they were classified according to the importance of the eye movement features, and the eye movement features of different importance levels were selected as the data input of the model. Finally, the best eye movement feature suitable for the model was selected. The specific architecture and parameters of the four algorithm models were shown in Tables 3 and 4.

3.0 Results

3.1 Importance of eye movement features

To obtain the importance of different eye movement features in warning, the F of 25 eye movement features were calculated by ANOVA. The importance of 25 eye movement features were ranked according to the importance, and the ranking was shown in Fig. 4. According to the classification function in

Table 3. *The architecture of the DRSN model*

Number	Layer	Filter numbers	Kernel size	Activation
1	Conv2D	25	2	ReLU
2	Soft thresholding	–	–	–
3	Dropout	Rate = 0.25	–	–
4	Conv2D	25	2	ReLU
5	Conv2D	25	2	ReLU
6	Soft thresholding	–	–	–
7	AveragePooling2D	Size = 2	–	–
8	BatchNormalisation	–	–	–
9	Dropout	Rate = 0.25	–	–
10	Dense	–	2	Sigmoid

Table 4. *Specific parameters of SVM, RF and BPNN models*

Model	Number	Parameter	Value
SVM	1	C	1
	2	Kernel	Sigmoid
	3	Probability	True
RF	1	n_estimators=	101
	2	Max_depth	17
	3	Min_samples_split	10
	4	Min_samples_leaf	10
	5	Max_characteristics	8
	6	Criterion	Gini
BPNN	1	Activation	tanh
	2	Hidden_layer_sizes	(64, 32, 32)
	3	Max_iter	1,000
	4	Random_state	1
	5	Solver	sgd

SPSS software, 25 features were classified according to the Euclidean distance of importance. The classification obtained five levels of eye movement features. There was a large difference between the importance of adjacent features indicated by the grey dashed lines. Since the first level only contains one eye movement feature, it was not feasible in the subsequent algorithm model analysis. After comprehensive consideration, the first level and the second level were combined to get the four-level eye movement features. As shown in Fig. 4, different colours were used to represent different levels of eye movement features.

3.2 Correlation among eye movement features

We explored the correlation of eye movement feature data and used the Pearson correlation coefficient method to calculate the correlation between features. Generally speaking, when the correlation coefficient was regarded as an absolute value, the between 0–0.09 indicated no correlation, 0.1–0.3 indicated weak correlation, 0.3–0.5 indicated moderate correlation, and 0.5–1.0 indicated strong correlation [34].

As shown in Fig. 5, we found a strong correlation between PSRX, PSRY and PDR in the pupil data of the right eye, and similar findings were also found in the pupil data of the left eye. These findings were consistent with the actual situation and verifies the validity of the data. Through the correlation heatmap,
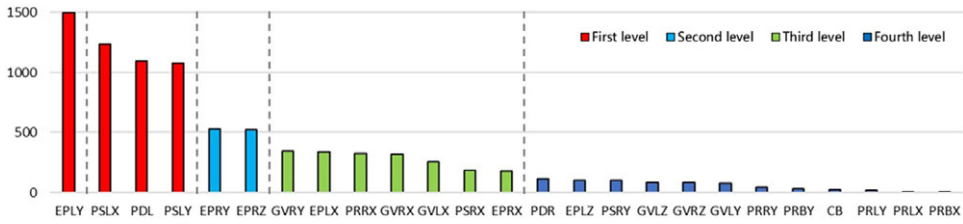
Figure 4. *The eye movement features were sorted according to the importance, and the Euclidean distance between each eye movement features was calculated to stratify them into four levels, which were distinguished by different colours.*
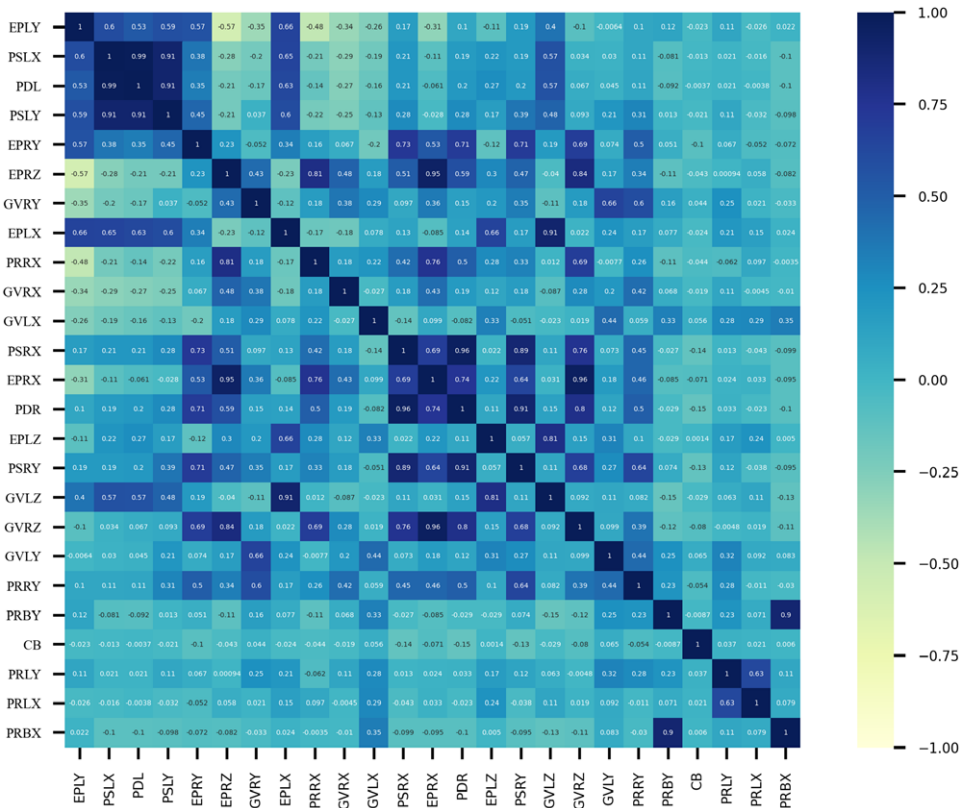

Figure 5. *Heat map of correlation between eye movement features.*

there was a strong correlation between the left eye pupil data and PSLX and PSLY corresponding to the relatively close importance of PSLX, PSLY and PDL in Fig. 4. Figure 5 provided us with the correlation between different eye movement features and verified the importance ranking of eye movement features.

3.3 Machine learning models

When four machine learning algorithm models were fed with different eye movement feature data, the recognition accuracy of the models for different coded warnings was studied. The eye movement features with different importance were divided to obtain four levels of eye movement features, as shown in Fig. 4. Part I was defined as the eye movement features of the first level, Part II was the eye movement features

Table 5. *Accuracies of the four models under different eye movement features*

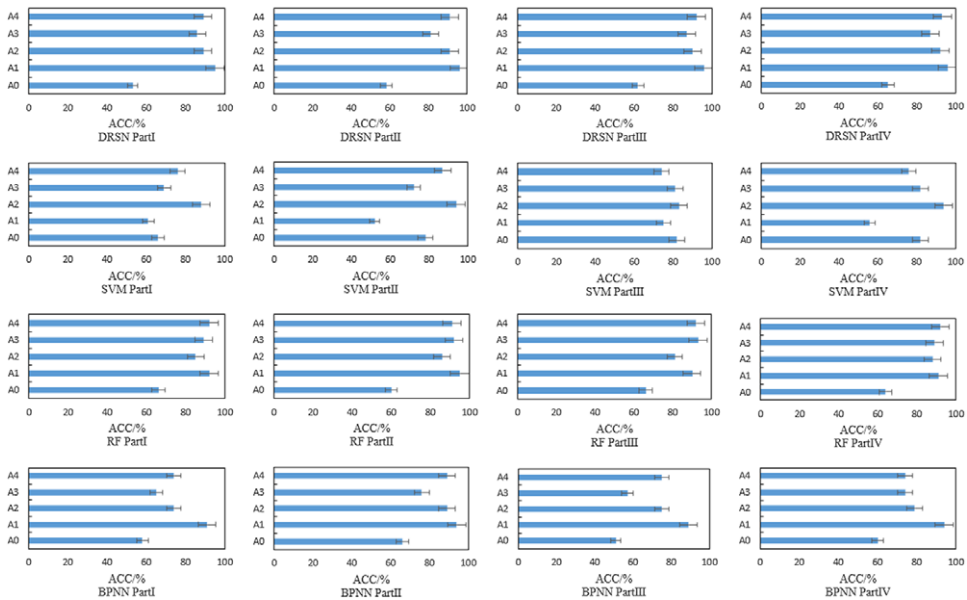| Features	DRSN	SVM	RF	BPNN
Part I	89.2%	79.6%	85.3%	82.9%
Part II	88.9%	81.3%	85.7%	88.1%
Part III	89.3%	80.7%	84.6%	79.8%
Part IV	90.4%	77.5%	85.4%	84.5%

Figure 6. *The recognition accuracy of the four models for different coded warning information under different eye movement features, where the horizontal axis is the accuracy, and the vertical axis is the warning information (A1, A2, A3 and A4) in different coding forms and information without warning (A0).*

of the first two levels, Part III was the eye movement features of the first three levels and the Part IV was the eye movement features of all levels. Four machine learning models, RF, SVM, BPNN and DRSN, were used to detect pilot perception of warning information through data corresponding to different types of eye movement features. It should be noted that the data included non-warning data and warning data, in which the category of non-warning data was A0, and the categories of warning data were A1, A2, A3 and A4.

When the selected eye movement features were different, the accuracy of the four models for different encoded warnings was shown in Fig. 6. Except for the SVM model, the other three models had high accuracy in identifying A1 warning information. However, the SVM model was more accurate than the other three models in recognising A0. Both the DRSN model and the RF model performed well in the identification process for different coding types of warning information, and there was no situation where the identification accuracy of one or some types of warning information was significantly lower. When the BPNN model recognised the A3 warning information, the accuracy rate was low no matter which Part of the eye movement feature was selected.

ROC curves of the four modes when different eye movement features were selected were shown in Fig. 7. It was found that the ROC curve of the DRSN model of the four models could reach the optimum
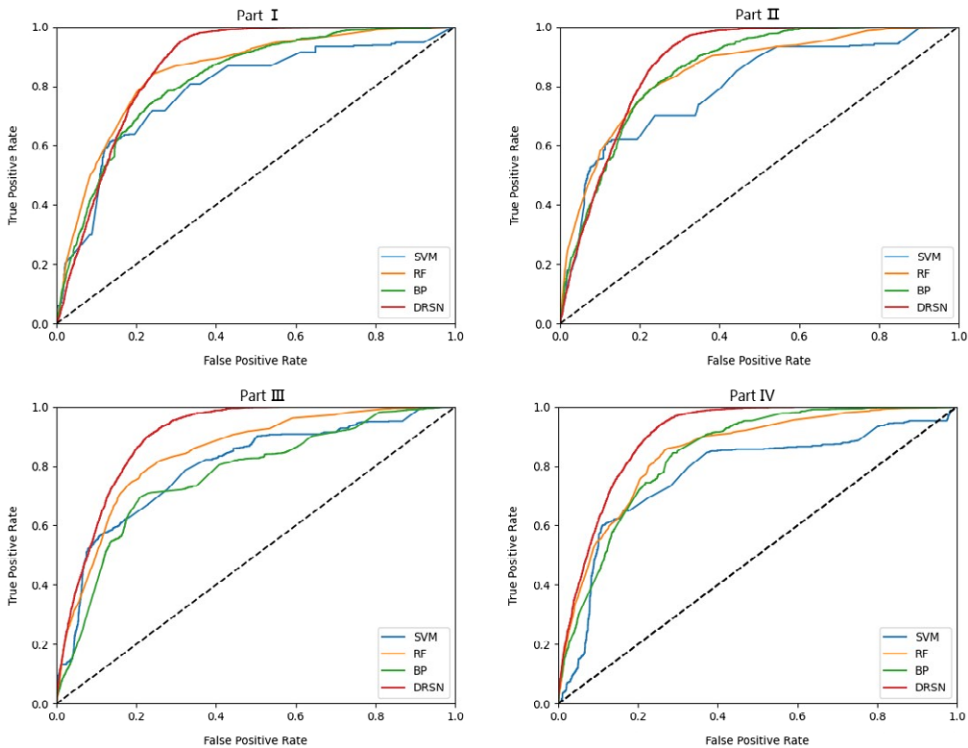
Figure 7. *ROC curves of the four models under different eye movement features.*

in the end. However, at the beginning of model training, choosing different eye movement features had a significant impact on the ROC curve of the DRSN model.

The accuracies of the four models when selecting different eye movement characteristics were shown in Table 5. We found that the best and worst performing models were the DRSN model and the SVM model, respectively. The selection of different eye movement features did not have a great impact on the DRSN, SVM and RF models, but had a greater impact on BPNN.

4.0 Discussion

In this study, we found that the pilot perception of warning information during flight experiments was better reflected in the data of eye features. When ANOVA was performed on eye movement features, the importance of each feature was obtained, and the importance of the feature was ranked and classified according to the size of the value F. Results showed that the importance of eye features was large and ranked at the top. Because the features of Part I (EPLY, PSLX, PDL, PSLY) and the Part II (Part I + EPRY, EPRZ) were both eye features, only the visual features were found in the features of the Part III (Part II + GVRY, EPLX, PRRX, GVRX, GVLX, PSRX, EPRX) and Part IV (Part III + PDR, EPLZ, PSRY, GVLZ, GVRZ, GVLY, PRRY, PRBY, CB, PRLY, PRLX, PRBX).

This study used machine learning model to analyse eye movement data to determine whether pilots were aware of warning information, and to identify different encoded forms of perception warnings through eye movement data. Four algorithm models were used to analyse eye movement data with different eye movement features, and the feasibility of method was judged by the accuracy of the models. The recognition accuracy of the four algorithm models when using eye movement data with different eye movement feature was shown in Fig. 6. Among the features of the Part IV, the SVM model had higher recognition accuracy for the warning information in the form of A2 encoding. In contrast, the SVM

model had the lowest warning recognition accuracy for the A1 coding form, and the warning information recognition accuracy for the A2 and A3 coding forms was about 80%. When the Part II and Part IV features were selected, the model of the SVM algorithm had the highest recognition accuracy for the warning information in the form of A2 encoding, reaching 94.3%. When selecting Part I and Part II features, there was a problem that the SVM model had low recognition accuracy for other types of warning information. However, this problem was improved when selecting Part III and Part IV features.

According to the information in Fig. 6, it was concluded that the recognition accuracy of RF model was significantly better than that of SVM model. The recognition accuracy of BPNN model in selecting Part II feature was higher than that of selecting Part III feature, which was contrary to the result of SVM model. The recognition accuracy of the BPNN algorithm model and the RF algorithm model for the A1 coding form was higher than that of the other three coding forms. These two models had the highest accuracy in selecting the features of the Part II, with 94.2% and 95.0%, respectively. However, the two models were easy to misidentify A0 data without warning information as warning information. Compared with the other three encoding forms, when the DRSN model selected the features of the Part IV, the model had a good performance in the recognition accuracy of the warning information in different coding forms. The DRSN model had higher warning information recognition accuracy for the A1 encoding form. When selecting Part IV features, the DRSN model had the highest recognition accuracy, reaching 96.4%. However, the model was also prone to misidentify A0 data as warning information. The warning information in the form of an A1 code was static warning information displayed on the screen, but there was no flashing and warning sound. A possible explanation was that A1 coded warning information requires more visual attention from the pilot. This was consistent with previous research finding that degraded, blurred, dark or other harder-to-see stimuli would require more viewing time [35].

From the accuracy of the four models in Table 5 and the ROC curves of the four models in Fig. 7, we found that when the SVM model and BPNN model selected the Part II features, the two models had the highest recognition accuracy for warning information. The accuracy rates were 81.3% and 88.1%, respectively. Increasing the number of features did not constantly enhance the accuracy of the model. Adding more features from the Part III to the SVM model and BPNN model, compared with using the Part II features, the accuracy of these two models decreased by about 0.6% and 8.3%. Furthermore, adding all features to the SVM algorithm model and BPNN algorithm model, compared with using the Part II features, the accuracy of these two models decreased by about 3.8% and 3.6%. This phenomenon had also been confirmed by Lou Y [36] and Destyanto [37]. These results may have been caused by the following reasons. One possible explanation was due to a phenomenon called the peak or "Hughes effect", which meant that the accuracy of a model did not increase continuously with the number of features. The Hughes effect was that, given a fixed dataset size, recognition accuracy initially increased with the number of features, but decreased when the number of features was higher than optimal [38, 39]. Therefore, a good model could be built using correlated features determined by model accuracy. Another possible explanation was that the Part II features were all eye features, which were more important in detecting pilot perception of warning information using SVM and BPNN models. In the RF model, when the model selects the Part IV feature, the recognition accuracy of the warning information was the highest, reaching 85.4%. At the same time, we found that the selection of different eye movement features had no significant effect on the accuracy of the RF model, and the accuracy rates were all between 84% and 86%. Previous studies had demonstrated differences in the importance of different eye movement features in machine learning models [40]. Thus, the possible reason why the accuracy of RF models did not change significantly was that RF algorithm models retained important features and deleted unimportant features.

In the DRSN model, when the Part IV was selected, the recognition accuracy of warning information was the highest, reaching 90.4%, which was better than SVM, RF and BPNN models. When the features of the Part I, Part II and Part III were selected, the accuracy of the model to identify the warning information reached 89%. In previous studies, the research methods for judging human perception based on physiological features include statistical analysis and machine learning, and the accuracy ranges from 50% to 82% [41–44]. In contrast, the method proposed in this paper had a certain degree of improvement in accuracy. It was found from Table 5 that the more features of eye movement data in the DRSN model,

the higher the corresponding accuracy. Because DRSN model was a new upgraded version of DRN model, which was a distributed reference network that deeply integrates attention mechanisms and soft thresholding. In the algorithm, the attention mechanism was able to notice unimportant features and then set them to zero through a soft threshold, and important features were noticed and kept by an attention mechanism. The ability of DRN model to extract useful information from noisy data has been enhanced. In this way, the higher the number of features in the eye-tracking data, the more useful features or feature combinations were extracted. To the best of our knowledge, our proposed model to assess pilot warning information perception based on eye movement data was the first to integrate and discuss all eye movement features, and it was also the first time that a DRSN model was used to predict the risk of insufficient information perception in a flight task.

5.0 Conclusions

This paper proposed an analysis method of eye movement data based on DRSNs. The effectiveness of the method was verified by conducting experiments, collecting and processing data, and training and testing models. In terms of experiments, different coding types of warning information were designed, including the presence or absence of flickering and the presence or absence of sound. The subjects piloted the helicopter in the simulation cockpit to complete the designated route. During the flight, warning information appeared randomly, and eye movement data of the subjects were collected. In terms of data processing, the collected eye movement data was divided according to the time when the warning information appeared, and an eye movement data sample based on the warning information were constructed. According to ANOVA, the importance of the eye movement features was calculated. The eye movement features were sorted and classified according to the importance. Thereby, the eye movement features contained in the eye movement data samples were further studied. In terms of models, a residual network was built and added to the CNN to build a DRSN model. The addition of the residual network solved the degradation problem of CNN, and the soft threshold realised the noise reduction of eye movement data samples. Eye movement data samples containing different eye movement features were used to train and test the DRSN model. In terms of model superiority, the DRSN model was compared with three machine learning models, namely SVM, RF and BPNN. Among the four models, DRSN performed the best, followed by RF and BPNN, and SVM performed the worst. When all eye movement features were selected, the DRSN model detected an average accuracy of 90.4% for pilots' perception of warning information, and the detection accuracy for A1-coded type of warning information reached 96.4%. Experiments showed that the DRSN model had advantages in detecting pilot perception of warning information.

Further exploration will be required in future work. For example, flight scenarios will be expanded for testing the proposed model in high-fidelity flight simulators. It is also necessary to determine that how realistic the use of eye trackers will be on long-haul flights, and whether the technology can be integrated into future intelligent cockpits.

Declaration of Competing Interest. The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] Wang, H., Jiang, N., Pan, T., Si, H. and Zou, W. Cognitive load identification of pilots based on physiological-psychological characteristics in complex environments, *J Adv Transport*, 2020. (l), pp 1–16. https://doi.org/10.1155/2020/5640784.

[2] Endsley, M.R. Design and evaluation for situation awareness enhancement, Proc Human Factors Soc Ann Meet, 1988, **32**, (2), pp 97–101. https://doi.org/10.1177/154193128803200221.

[3] Sterling, R.D. Defense, security, and cockpit displays, Proceedings of SPIE - The International Society for Optical Engineering, 2004, p 5443.

[4] Liggett, K.K. and Gallimore, J.J. An analysis of control reversal errors during unusual attitude recoveries using helmet-mounted display symbology, *Aviat Space Environ Med*, 2002, **73**, (2), pp 102–111.

[5] Sherry, L., Fennell, K., Feary, M. and Polson, P. Human-computer interaction analysis of flight management system messages, *J Aircr*, 2012, **43**, (5), pp 1372–1376.

[6] Su, H., Liu, Z., Cao, L., Dai, S. and Lou, Z. Effects of visuospatial working memory capacity on tasks of fighter aircraft instrument reading. *Space Med Med Eng*, 2012, **25**, (06), pp 412–416.

[7] Li, W.C., Cao, J., Lin, J.H., Braithwaite, G. and Greaves, M. The evaluation of pilot's first fixation and response time to different design of alerting messages, 14th International Conference on Engineering Psychology and Cognitive Ergonomics, 2017, pp 21–31.

[8] Endsley, M.R. and Garland, D.J. Situation Awareness Analysis and Measurement. CRC Press, Boca Raton, USA, 2000.

[9] Li, W.C., Zhang, J., Minh, T.L., Cao, J. and Wang, L. Visual scan patterns reflect to human-computer interactions on processing different types of messages in the flight deck, *Int J Ind Ergon*, 2019, **72**, pp 54–60. https://doi.org/10.1016/j.ergon.2019.04.003.

[10] Kearney, P., Li, W.C. and Lin, J. The impact of alerting design on air traffic controllers' response to conflict detection and resolution, *Int J Ind Ergon*, 2016, **56**, pp 51–58. https://doi.org/10.1016/j.ergon.2016.09.002.

[11] Wang, W.H., Hou, S.Y. and Jiang, XB. Application evaluation of self-explaining intersections based on visual information, International Conference on Transportation and Development 2020: Transportation Safety, 2020, pp 83–94. https://doi.org/10.1061/9780784483145.008

[12] Vidulich, M.A., Wickens, C.D., Tsang, P.S. and Flach, J.M. Information processing in aviation. In E. Salas and D. Maurino (Eds.), *Human Factors in Aviation*, London, UK: Elsevier, 2010, pp 175–207. https://doi.org/10.1016/B978-0-12-374518-7.00007-9.

[13] Robinski, M. and Stein, M. Tracking visual scanning techniques in training simulation for helicopter landing, *Journal of Eye Movement Research*, 2013, **6**, (2).

[14] Vsevolod, I.D., Lefrançois, O., Dehais, F. and Causse, M. The neuroergonomics of aircraft cockpits: the four stages of eye-tracking integration to enhance flight safety, *Safety*, 2018, **4**, (1), p 8. https://doi.org/10.3390/safety4010008.

[15] Zhang, L., Zhou, Q., Yin, Q. and Liu, Z. Assessment of pilots mental fatigue status with the eye movement features, International Conference on Applied Human Factors and Ergonomics, 2019, pp 146–155. https://doi.org/10.1007/978-3-319-94334-3_16.

[16] Mohan, D.B., Jeevitha, S., Prabhakar, G., Saluja, K.S. and Biswas, P. Estimating pilots' cognitive load from ocular parameters through simulation and in-flight studies, *J Eye Movem Res*, 2019, **12**, (3).

[17] Feng, S.A., Tl, B., Xw, A., Zl, C., Yz, A. and Xl, A. The influence of pilot's attention allocation on instrument reading during take-off: the mediating effect of attention span. *Appl Ergon*, 2020, **90**, p 103245. https://doi.org/10.1016/j.apergo.2020.103245.

[18] De Winter, J.C.F., Eisma, Y.B., Cabrall, C.D.D., et al. Situation awareness based on eye movements in relation to the task environment. *Cogn Tech Work*, 2019, **21**, pp 99–111. https://doi.org/10.1007/s10111-018-0527-6.

[19] Allsop, J. and Gray, R. Flying under pressure: effects of anxiety on attention and gaze behavior in aviation, *J Appl Res Mem Cognit*, 2015, **3**, (2), pp 63–71. https://doi.org/10.1016/j.jarmac.2014.04.010.

[20] Bałaj, B., Lewkowicz, R., Francuz, P., et al. Spatial disorientation cue effects on gaze behaviour in pilots and non-pilots, *Cogn Tech Work*, 2019, **21**, pp 473–486. https://doi.org/10.1007/s10111-018-0534-7.

[21] Tichon, J.G., Wallis, G., Riek, S., et al. Physiological measurement of anxiety to evaluate performance in simulation training, *Cogn Tech Work*, 2014, **16**, pp 203–210. https://doi.org/10.1007/s10111-013-0257-8.

[22] Korek, W.T., Mendez, A., Asad, H.U., Li, W.C. and Lone, M. Understanding human behaviour in flight operation using eye-tracking technology. *Eng Psychol Cognit Ergon*, 2020, pp 304–320.

[23] Friedrich, M., Lee, S.Y., Bates, P., et al. The influence of training level on manual flight in connection to performance, scan pattern, and task load, *Cogn Tech Work*, 2021, **23**, pp 715–730. https://doi.org/10.1007/s10111-020-00663-8.

[24] Shen, Z., Zhang, L., Li, R., Hou, J., Liu, C. and Hu, W. The effects of color combinations, luminance contrast, and area ratio on icon visual search performance - sciencedirect. *Displays*, 2021, **67**. https://doi.org/10.1016/j.displa.2021.101999.

[25] Yu, C., Wang, E.M., Li, W. and Braithwaite, G. Pilots' visual scan patterns and situation awareness in flight operations, *Aviat Space Environ Med*, 2014, **85**, (7), pp 708–714. https://doi.org/10.3357/asem.3847.2014.

[26] Behrend, J. and Dehais, F. How role assignment impacts decision-making in high-risk environments: evidence from eye-tracking in aviation, *Saf Sci*, 2020, **127**, p 104738. https://doi.org/10.1016/j.ssci.2020.104738.

[27] Costela, F.M. and Castro-Torres, J.J. Risk prediction model using eye movements during simulated driving with logistic regressions and neural networks, *Transport Res F: Traf Psychol Behav*, 2020, **74**, pp 511–521. https://doi.org/10.1016/j.trf.2020.09.003.

[28] Liang, Y., Reyes, M.L. and Lee, J.D. Real-time detection of driver cognitive distraction using support vector machines, *IEEE Trans Intell Transp Syst*, 2007, **8**, (2), pp 340–350.

[29] Nguyen, V.C., Vu, D., Lam, S.T.H. Mel-frequency cepstral coefficients for eye movement identification, Proceedings of the 2012 IEEE 24th International Conference on Tools with Artificial Intelligence (ICTAI), vol. 1, 2012, pp 253–260. https://doi.org/10.1109/ICTAI.2012.42.

[30] Frutos-Pascual, M. and Garcia-Zapirain, B. Assessing visual attention using eye tracking sensors in intelligent cognitive therapies based on serious games, *Sensors*, 2021, **15**, (5), pp 11092–11117. https://doi.org/10.3390/s150511092.

[31] Taku, H., Hirotoshi, I., Akira, Y. and Hiroaki, K. Detecting cognitive distraction using random forest by considering eye movement type, *Int J Cognit Inform Nat Intell*, 2017, **11**, (1), pp 16–28. https://doi.org/10.4018/IJCINI.2017010102.

[32] Cheng, B., Zhang, C., Ding, X. and Wu, X. Convolutional neural network implementation for eye movement recognition based on video, 2018 33rd Youth Academic Annual Conference of Chinese Association of Automation, pp 179–184. https://doi.org/10.1109/YAC.2018.8406368.

[33] Zhao, M., Zhong, S., Fu, X., Tang, B. and Pecht, M. Deep residual shrinkage networks for fault diagnosis, *IEEE Trans Ind Inf*, 2020, **16**, (7), pp 4681–4690. https://doi.org/10.1109/TII.2019.2943898.

[34] Cohen, J. Statistical power analysis for the behavioral sciences, *Comput Environ Urban Syst*, 1988, **14**, (1), p 71. https://doi.org/10.4324/9780203771587.

[35] Henderson, J.M. an dLuke, S.G. Stable individual differences in saccadic eye movements during reading, pseudoreading, scene viewing, and scene search, *J Exp Psychol Human Percep Perform*, 2014, **40**, (4), pp 1390–1400. https://doi.org/10.1037/a0036330.

[36] Lou, Y., Liu, Y., Kaakinen, J.K., et al. Using support vector machines to identify literacy skills: Evidence from eye movements. *Behav Res*, 2017, **49**, pp 887–895. https://doi.org/10.3758/s13428-016-0748-7.

[37] Destyanto, T. and Lin, R.F. Detecting computer activities using eye-movement features, *J Ambient Intell Hum Comput*, 2020, **4**, pp 1–11. https://doi.org/10.1007/s12652-020-02683-8.

[38] Bruzzone, L. and Serpico, S.B. A technique for feature selection in multiclass problems, *Int J Remote Sens*, 2000, **21**, (3), pp 549–563. https://doi.org/10.1080/014311600210740.

[39] Hughes, G. On the mean accuracy of statistical pattern recognizers, *IEEE Trans Inf Theory*, 1968, **14**, (1), pp 55–63. https://doi.org/10.1109/TIT.1968.1054102.

[40] Liao, H., Dong, WH., Huang, HS., et al. Inferring user tasks in pedestrian navigation from eye movement data in real-world environments, *Int J Geograph Inform Sci*, 2019, **33**, (4), pp 739–763. https://doi.org/10.1080/13658816.2018.1482554.

[41] Kim, N., Kim, J. and Ahn, C.R. Predicting workers' inattentiveness to struck-by hazards by monitoring biosignals during a construction task: A virtual reality experiment, *Adv Eng Inf*, 2021, **49**, p 101359. https://doi.org/10.1016/j.aei.2021.101359.

[42] Charlton, S.G., Starkey, N.J., et al. What's the risk? A comparison of actual and perceived driving risk, *Transport Res F: Traf Psychol Behav*, 2014, **25**, Part A, pp 50–64. https://doi.org/10.1016/j.trf.2014.05.003.

[43] Daley, M.S., Gever, D., Posada-Quintero, H.F., et al. Machine learning models for the classification of sleep deprivation induced performance impairment during a psychomotor vigilance task using indices of eye and face tracking, *Front Artif Intell*, 2020, **3**, p 17. https://doi.org/10.3389/frai.2020.00017.

[44] Abdelrahman, Y., Khan, A.A., Newn, J., et al. Classifying attention types with thermal imaging and eye tracking, *Proc ACM Interact Mob Wearable Ubiquitous Technol*, 2019, **3**, (3), pp 1–27. https://doi.org/10.1145/3351227.