Nystagmus Detection Method Based on Gating Mechanism and Attention Mechanism

Maolin Hou
School of Computer Science and Engineering
Xi'an Technological University
Xi'an, China
E-mail: 1298191511@qq.com

***Abstract*—In this paper, a new model based on the combination of improved LSTM and self-attention mechanism is studied for the detection of nystagmus caused by vestibular illusion in pilots during flight. An efficient and robust nystagmus detection method was proposed by constructing experimental simulation scenarios and collecting and analyzing pilot eye movement data. The improved LSTM model enhances the ability of capturing the medium and long term dependence of the ocular shock sequence by adding a gating unit, and the introduction of self-attention mechanism further improves the analytical accuracy of the model for complex eye movement sequences. The experimental results show that the model has excellent performance in accuracy, recall rate and F1 score, which is significantly better than the traditional model, providing a new technical means for the detection of vestibular illusion.The LSTM-GRU-Attention model has been experimentally verified to perform best in accuracy, recall, and F1 score, reaching 095, 0.91, and 0.93 respectively, indicating that the outperforms the other two models in overall classification performance, positive sample recognition ability, and balance between accuracy and recall.**

Keywords-LSTM; Self-Attention; Nystagmus

I. INTRODUCTION

Vestibular system is the main organ of the human body to perceive the changes of body position and environment, plays a key role in the human body's own sense of balance and spatial sense, is an important part of the balance system, and is closely related to spatial disorientation and movement disease. If the vestibular function is abnormal, it will directly affect the pilot's operation quality and work efficiency, health status and flight safety. Therefore, vestibular function examination has become an important part of the pilot recruitment physical examination [1]. In recent years, studies on the interaction between eye movement and vestibular system function mainly stimulate the vestibular system to obtain relevant eye movement, so as to verify the close coupling relationship between eye movement and vestibular system. Nystagmmus one of the most obvious and important signs of various vestibular reactions in clinical practice.

Wang et al. proposed a pupil location and iris segmentation method based on the full convolutional network, and used the shape and structure information of pupil center, iris region and its inner and outer boundaries to achieve pupil location and iris segmentation at the same time. The human eye pupil detection method based on deep learning and appearance texture features [3] has received more and more attention, and its effectiveness and robustness have also promoted practical applications related to eye tracking. On the other hand, as the amount of data increases, the differences between different individuals also increase, and the data distribution becomes more diverse, which decreases the detection ability based on texture features. At the same time, massive data requires a lot of manpower to manually label. How to design a more robust and effective model using a small number of limited samples is the main problem to be solved for human eye pupil detection based on appearance texture features.

The method based on context information mainly uses the eye region and its context face structure and texture information to realize the accurate positioning of the pupil of the human eye.

DOI: 10.2478/ijanmc-2024-0041

Based on the idea of coarse to fine, multi-scale nonlinear [5] feature mapping is proposed based on the supervised descent method [4] to achieve accurate pupil detection. Inspired by the face key point detection method. A large number of flight practice studies have shown that pilots are prone to flight illusion during flight, and flight illusion is the most representative of Spatial Disorientation (SD) and one of the important factors causing serious flight accidents [2].

The vestibular illusion detection method studied in this paper is mainly based on the illusion of tilt shape in flight space disorientation, which is based on the fact that tilt illusion accounts for the largest proportion in flight illusion manifestations [7], and the detection of nystagmus [6] by computer vision technology is the main method of this paper.

II. Type Style and Fonts

This paper mainly focuses on the application of machine learning in vestibular [8] illusion detection, focusing on the spatial disorientation pilots may encounter during flight, with a special focus on tilt illusion. Therefore, the construction of experimental simulation scenes, how to induce the generation of nystagmus or illusion, and data collection and analysis have become the main research contents.

Specifically, the research includes:Simulation and data collection of the experimental scene: Design the experimental scene of the illusion of tilt shape to simulate the possible situation in flight. The eye movement data of pilots under different conditions are collected and data sets are built for training and validation of machine learning models.

Establishment of vestibular [8] illusion detection method: Through machine learning technology, the vestibular illusion detection model is constructed, mainly focusing on the illusion of oblique morphology, and the model will be trained based on the rotating motion of various angles and directions that the pilot may experience.

Application of computer vision technology in nystagmus detection: The use of computer vision technology, with special attention to the occurrence of nystagmus, through the analysis and processing of video [10] data, improve the accuracy of eye tracking, so as to more accurately capture the eye movement changes caused by vestibular illusion.

III. Network model

The overall architecture of the algorithm consists of three main parts: eye position recognition (RITNet), the Embedding layer, and the improved LSTM model combined with Transformer self-attention mechanism (Figure 1).

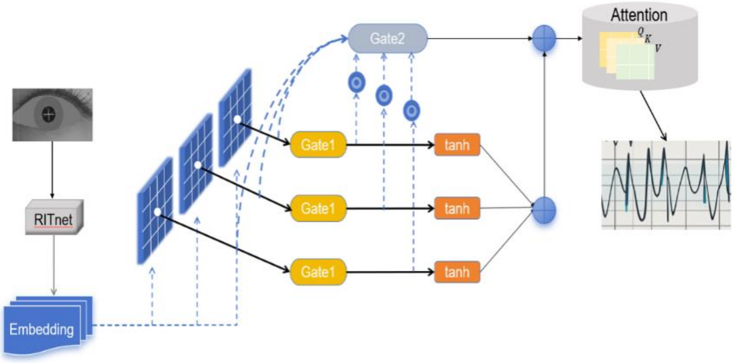

Figure 1. Network model

A. Eye Position Recognition (RITNet)

Accurate recognition of eye position is an important prerequisite for detecting vestibular illusion, especially when simulating the spatial disorientation [9] of pilots in actual flight (such as tilt illusion), capturing eye movement data is a key link to understand the physiological response of pilots. Therefore, it is particularly important to choose an efficient and accurate eye tracking method.

In simulated flight environments where pilots are confronted with rapidly changing visual and spatial cues, tiny movements of the pilot's eyes are critical to the creation and response to the illusion. In order to detect the pilot's eye movement response, this paper collects the pilot's pupil movement data with high-precision eye tracking equipment, and adopts RITNet model to process the data. RITNet can efficiently handle eye-tracking tasks in complex scenarios such as different lighting, occlusion, and pilot blinking, ensuring continuity and reliability of pupil position information.

The core of RITNet is its use of convolutional neural networks (CNNS) to extract multiple layers of features from input images, combined with contextual information to enhance the robustness of the model. Specifically, RITNet includes the following key steps:

Pupil detection and iris segmentation: RITNet uses a full convolutional network to simultaneously locate the pupil center and segment the iris region. This process combines information about the shape and structure of the pupil and iris to pinpoint the location of the pupil.

Multi-scale feature extraction: The model can extract the context information of the area around the pupil from different scales. By introducing multi-scale convolution kernel, RITNet can capture features of different sizes, so as to adapt to pupil changes under different conditions. The model can recognize the changes of the pilot's eye attitude, rotational movement, and pupil changes under different lighting conditions during flight.

Case segmentation: RITNet is based on case segmentation technology, which enables it to not only accurately detect the eye position of a single pilot, but also separate the eye information of different individuals in multiple scenarios. This is particularly important for the acquisition of eye movement data in the experimental scene of multi-person flight simulator.

Continuous tracking of time series: By processing the input continuous image frames, RITNet can generate a continuous sequence of pupil positions. This sequence data not only reflects the change of pupil position, but also provides time information for subsequent nystagmus detection. Especially in vestibular delusion-induced experiments, the pilot's pupil movement can change rapidly, and RITNet can ensure that no critical information is lost by continuously tracking these changes.

B. Embedding layer

In the vestibular illusion detection task, the pupil position sequence output by RITNet contains rich timing information. However, the length of these sequences may vary depending on the pilot's experimental process and actual eye movement reaction time. In order to be able to convert these input data of different lengths into a fixed format that the deep learning model can handle, the Embedding layer is introduced to play the key role of "information compression" and "semantic transformation".

In this paper, the main task of the Embedding layer is to convert the continuous pupil position information output by RITNet into embedded vectors of fixed dimensions. This process is similar to the word vector embedding in natural language processing, which can compress the original position information into a vector space with semantic characteristics, which is convenient for model processing and understanding.

Because the pupil position information is output in the form of sequence, the reaction time of different pilots under different experimental conditions may lead to the difference in the length of pupil position sequence. The introduction of the Embedding layer can effectively solve this problem, so that sequences of different lengths can be mapped to the same dimension. Through this transformation, the subsequent LSTM and Transformer layers of the model can efficiently process this data without bias due to differences in input length.

The core of the Embedding layer is to map the high-dimensional pupil position information sequence to the low-dimensional vector space while preserving the most important position information features in the sequence. Specifically, the pupil position sequence output by RITNet is a time series containing information about the specific position of the pilot's eyes at each point in time. The Embedding layer learns the important features of the location information sequence and converts it into a fixed-length embedding vector.

C. Improved LSTM model

The LSTM unit is used to learn long-term dependencies in the ocular shock wave sequence in the task of prediction.

The improved LSTM model has two new gating units: Gate1 and Gate2(Figure 2).

Gate1: Control the incoming and outgoing information according to the ocular shock vector output on the Embedding layer and the LSTM unit status at the last moment. It helps the model better understand the influence of historical ocular shock states on the current state.

Gate2: The output of the model is further adjusted according to the current LSTM unit status and the context information of the ocular shock sequence. It enhances the model's understanding of the overall context of the ocular shock sequence.

In the processing of the pilot's ocular shock wave sequence, it is a typical time series signal, which contains the physiological response of the pilot in the face of spatial disorientation. It usually exhibits a certain rhythm and reflects the collaborative work between the pilot's vestibular system and the visual system. Traditional deep learning models may not be effective at capturing time dependencies in data. The LSTM model, with its special "memory unit" design, can well retain and utilize the earlier time step information in the sequence to deal with long distance dependence, which makes it very suitable for processing time series data such as eye shock wave.

In order to accurately predict the ocular shock waves of pilots in vestibular illusion (especially tilt illusion), the model must have the ability to capture long-term dependencies in the sequence data. The improved LSTM model proposed in this paper has made a key enhancement on the basis of the traditional LSTM model, especially introducing two gate control units: Gate1 and Gate2. These improvements help the model to better handle complex sequences of eye tremors and improve its ability to predict vestibular illusions.

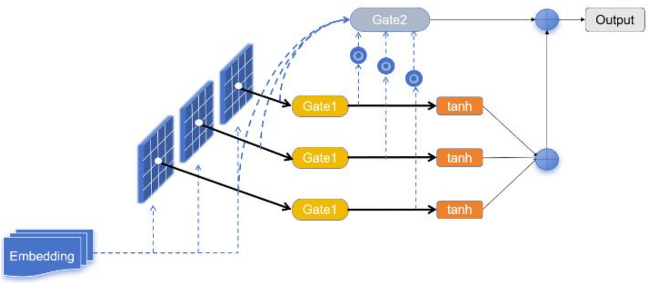

Figure 2. improved LSTM model

Gate1 is designed to help the model better understand the relationship between the current moment and the historical state of the eye shock by introducing additional gating units. The core function of Gate1 is to dynamically control the inflow and outflow of information according to the Embedding vector of the pupil position output and the previous state of the LSTM unit on the embedding layer, so as to determine which information should be retained and which information should be forgotten. This design solves the gradient disappearance problem common to LSTM in long series data, ensuring that the model can extract useful information from distant historical states.

The introduction of Gate2 further enhances the ability of the model to understand the context information of the whole ocular shock sequence. Gate2 adjusts the output of the model according to the current LSTM unit status and context information to ensure that the model can capture the global dependencies closely related to the current prediction task.

Gate1: This gating unit is used to control the inflow and outflow of information, in particular to help the model better understand the influence of the historical state of the eye shock on the current state. Gate1 computes the gating value based on the current input (pupil position information processed by the Embedding layer) and the LSTM unit status at the previous time to determine which information should be forgotten and which information should be retained (Figure 3).

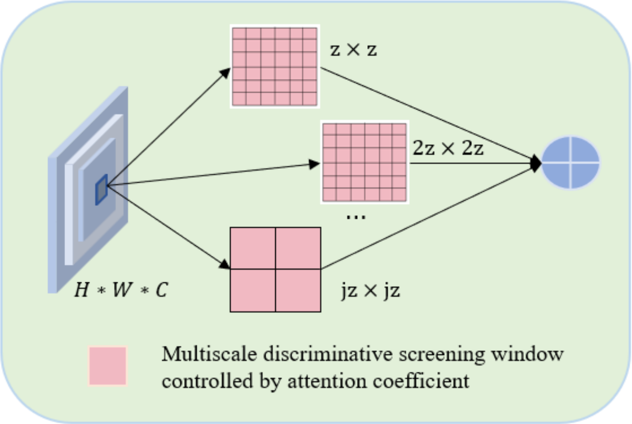
z × z
2z × 2z
...
H * W * C
jz × jz
Multiscale discriminative screening window controlled by attention coefficient

Figure 3. Gate1

The three Gate1 branches are fed into the three parallel Gate1 branches through Embedding vectors processed on the embedding layer. Each Gate1 branch calculates the similarity score α_(t,i) according to formulas (1) and (2), which is used to control the filtering of information:

$$\alpha_{t,i}=\frac{\exp(e_{t,i})}{\sum_{j=1}^{n}\exp(e_{t,i})},i=1.2...n \tag{1}$$

The formula for e_(t,i) is as follows (2):

$$e_{t,i}=g(h_{t-1},x_i;\theta) \tag{2}$$

α(t,i), based on the similarity of h(t-1) and xi, obtained by the softmax function, represents the correlation between the hidden state and the external input.

According to α(t,i), each Gate1 branch divides the embedded vector using partition Windows of different sizes (zj×zj) to capture information at different scales. This allows the model to consider both global and local features, improving the efficiency of information extraction.

Assume that the input Gate1 is H∈RH×W×C.The feature map, in the NTH branch of j parallel branches, is sized by controlling the multi-scale partition window.Divide H into sizes of (zj×zj,C) tensor. Represents grid for each non-overlapping slice of size zj×zj. This allows larger partitions to capture more external input and visual errors, and smaller partitions to extract information on finer areas to preserve the relationship between them.

Gate2: The gate control unit further adjusts the output of the model based on the LSTM unit status at the current moment and the context information of the ocular shock sequence. Gate2 enhances the model's understanding of the overall context of the ocular shock sequence, making the prediction result more a ccurate.

The output from all three Gate1 branches is passed into a shared Gate2.

Gate2, as a motion decision screening gate, further screens the output of the model with the current LSTM unit state and the context information of the ocular shock sequence. Formula (3) describes the calculation process of Gate2, where g(i,j)m is a vector representing feature selection among the aggregation vectors of redundant information, external input and ocular shock wave sequence information.

$$\begin{cases} g_{j,i}^{m}=\sigma(W^{m}[r_{i,j}^{t,l},h_{j}^{t,l}]+b^{m}) \\ r_{j,i}^{t}=\varphi_{r}[x_{i}^{t}-x_{j}^{t},y_{i}^{t}-y_{j}^{t};W^{r}] \\ h_{j}^{t,l}=\tanh(h_{t-1},x_{i})+(1-\alpha_{t,i})\odot h_{t-1} \end{cases} \tag{3}$$

D. Transformer self-attention mechanism

Based on the improved LSTM model, the self-attention mechanism of Transformer is introduced (Figure 4).

The self-attention mechanism generates an attention weight matrix by calculating the correlation score between any two positions in the input sequence, and then weights the input sequence.

This helps the model to capture the long-distance dependence in the sequence of pupil position, and enhances the model's prediction of tilt illusion.

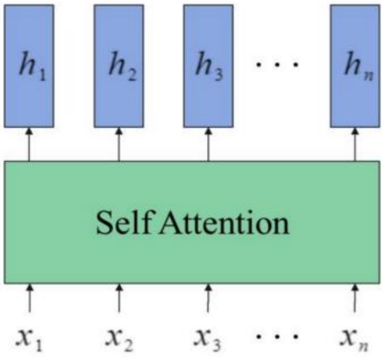

Figure 4. self-attention structure

IV. Experiment

A. Preparation Data

The experimental data set contains pupil position sequence data collected during simulated flight missions, captured by high-precision eye tracking equipment, and pre-processed steps (such as filtering, normalization, interpolation processing) to ensure the quality of the data. The dataset size covers thousands of samples, each containing a continuous sequence of pupil positions over a period of time and their corresponding ophthalmogram labels. Before data preprocessing can begin, we need to data label the raw eye movement data. As shown in Figure 5, the slow-phase nystagmus region is marked yellow. The marking process is as follows : ① Draw a line chart with the above eye movement data. ② Select the area where the difference between the maximum and minimum values of the ordinate is greater than 1 and the slope is slow as slow-phase nystagmus, and mark all frames in this area as 1. Repeat the above process until all the points contained in the slow-phase nystagmus region have been correctly labeled, and the remaining points are labeled as 0.

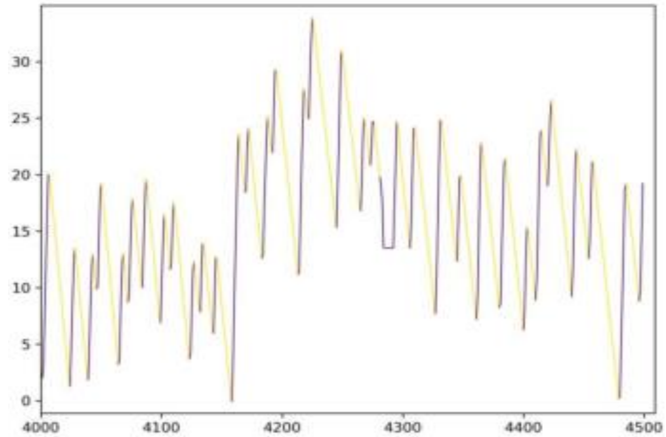

Figure 5. Data tag example

B. Training process

Data loading and partitioning:

The pre-processed data set was divided into training set, verification set and test set according to the ratio of 7:1.5:1.5.

- Hyperparameter Settings:Learning rate: The initial learning rate is 0.001 and the learning rate attenuation strategy is adopted.
- Batch size: Select the appropriate batch size based on the data set and GPU video memory size.
- Training rounds: The initial training rounds are set to 50 or 100, and the early stop mechanism is used to stop the training in advance.
- Early stop mechanism:When the validation set performance does not improve in several consecutive rounds, the early stop mechanism is triggered to stop the training and save the current optimal model.

C. Evaluation indicators

- Accuracy: Measures the proportion of samples the model correctly classifies. For class imbalance problems such as tilt illusion detection, the accuracy may be affected by a high proportion of negative class samples, so the accuracy should be evaluated in combination with other indicators to obtain a comprehensive performance analysis.
- Recall rate: Measures the ability of the model to recognize the tilt illusion, i.e. the proportion of true cases (TPS) that are correctly identified. In tilt illusion detection missions, recall rates are critical because undetected illusions can lead to a potential risk of pilot illusion. The high recall rate indicates that the model has a strong sensitivity in detecting the actual illusion, which helps to avoid the case of missing detection.
- F1 score: The F1 score is a harmonic average of accuracy and recall rates, and is particularly suitable for class imbalance problems. By considering both the model's Precision (that is, the proportion of samples that correctly predict a positive class) and the recall rate. The introduction of F1 scores balances the relationship between recall and accuracy, ensuring that the model does not miss important positive samples while maintaining a low false positive rate.
- ROC curve and AUC value: By plotting the ROC curve and calculating the AUC value, we can evaluate the performance of the model under different thresholds. The value of AUC can directly reflect the ability of the model to separate the complex ocular shock sequences, and provide a strong evaluation basis for the vestibular illusion detection task.

D. Experimental results and analysis

The chart below shows the changes of each index of the model under different training rounds.

As the training progressed, the model's performance continued to improve, especially in terms of accuracy, recall and F1 scores, showing significant improvements. These charts not only intuitively reflect the gradual enhancement of the model's ability to identify positive samples, but also demonstrate its optimization effect in reducing misjudgments. With these results, we were able to gain a clearer understanding of the performance and potential of the improved LSTM-GRU-Attention model in the task of ocular shock pattern recognition (Figure 6, Figure 7).

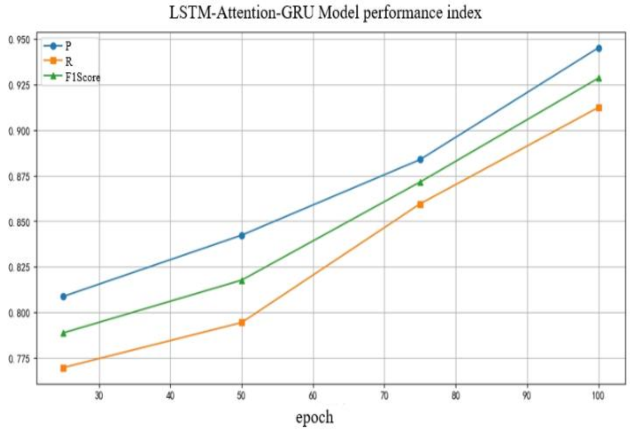

Figure 6. Training indexes of each round of the model

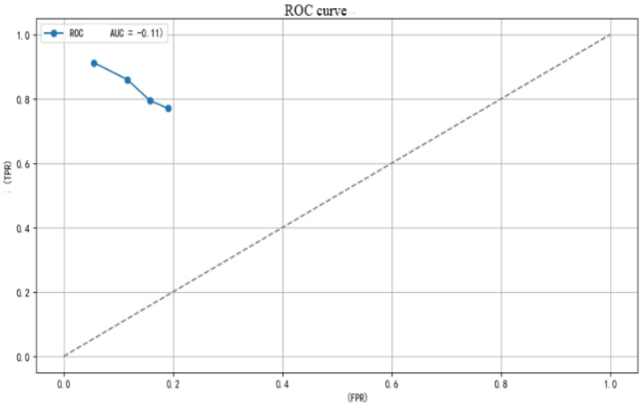

Figure 7. ROC curve

These indicators show that with the increase of training rounds, the performance of the model is gradually improved. Specifically, the accuracy rate increased with the increase of training rounds, and finally reached 0.945234. The recall rate also increased with the increase of training rounds, indicating that the model's ability to recognize positive examples was gradually enhanced. F1 scores also increased in most cases with more training rounds, reaching a maximum of 0.928573. The true positive rate showed that the ability of the model to correctly identify positive cases increased from 0.769667 to 0.912489. The false-positive rate indicates that the frequency of the model mistakenly identifying negative cases as positive cases gradually decreases to 0.054766, which shows the optimization effect of the model.

In conclusion, with the increase of training rounds, the LSTM-GRU-Attention waveform recognition network has shown better performance improvement and optimization in the slow-direction eye shock waveform recognition task.

E. Comparative experiment

This comparison experiment aims to verify the performance of the proposed "LSTM-GRU-Attention" model (hereinafter referred to as "My model") on the tilt illusion detection task and compare it with existing models. These include the "LSTM-Transformer" model, the "LSTM-Attention" model, and the ARIMA model.

In order to ensure the comprehensiveness and fairness of the comparison experiment, this paper selected two representative models to compare with the model designed in this paper:

LSTM-attention model: This model introduces the Attention mechanism based on the classical LSTM, so that it can weight the important time steps in the sequence. Through the attention mechanism, the model can dynamically adjust the focus, capture the key information in the eye shock sequence more effectively, and improve the detection ability of tilt illusion.

ARIMA model: As a traditional time series analysis model, ARIMA model models linear time series by means of autoregression and moving average. Although it performs well when dealing with simple sequence data, its performance can be limited when faced with complex nonlinear nystomograms. Therefore, the introduction of ARIMA models helps to demonstrate the advantages of deep learning models in the processing of complex time series data.

Experimental results as follow (Figure 8):

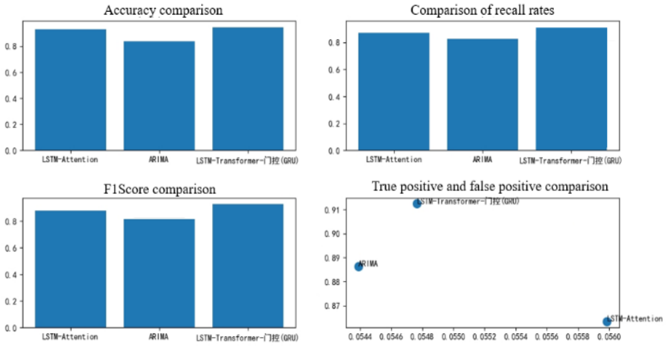

Figure 8. Data results of 100 rounds of training for each model

- Accuracy: The accuracy of LSTM-GRU-Attention model is the highest, reaching 0.945234, indicating that this model is superior to the other two models in overall classification performance.
- Recall rate: The recall rate of LSTM-GRU-Attention model is also the highest, which is 0.912489, indicating that this model has a good performance in identifying positive samples.
- F1 score: LSTM-GRU-Attention model has the highest F1 score, reaching 0.928573, indicating that the model has achieved a good balance between accuracy and recall rate.

True positive rate and false positive rate: LSTM-GRU-Attention model has the highest true positive rate and relatively low false positive rate, which further proves the advantages of this model in waveform recognition tasks.

The LSTM-GRU-Attention model shows excellent performance in waveform recognition tasks. This is mainly due to the fact that the model combines three different network structures, LSTM, self-attention and GRU, which can capture the long-term dependence relationship of data, and improve the recognition ability of the model by using the attention mechanism and the gating mechanism. In contrast, ARIMA model, as a traditional model based on time series analysis, is powerless to deal with complex tasks such as waveform recognition. Although the LSTM-Attention model also uses deep learning technology, it is still inferior to the LSTM-GRU-Attention model in some indicators. The experimental results show that the LSTM-GRU-Attention model has achieved the best performance in accuracy, recall rate, F1 score, true positive rate and false positive rate, and is the optimal model in waveform recognition task.

V. CONCLUSIONS

In this paper, a detection method based on machine learning is proposed for the vestibular illusion that pilots may encounter during flight, especially the tilt illusion. By constructing experimental simulation scenarios and collecting eye movement data, a detection model based on RITNet, improved LSTM model and Transformer self-attention mechanism is designed and implemented. In the research process, computer vision technology and embedding layer processing are used to realize the efficient recognition of complex eye movement sequences.

The experimental results show that the LSTM-GRU-Attention model proposed in this paper is superior to the traditional model in many indexes, demonstrating strong detection ability and robustness. This suggests that by introducing gated units and self-attention mechanisms, features related to vestibular illusions can be captured more effectively, thereby improving the overall performance of the model. Compared with other existing methods, this model has excellent performance in accuracy rate, recall rate, F1 score and so on, and has achieved a good application effect.In the future, we will continue to optimize this model by collecting more diverse eye movement data, covering different flight phases and conditions, as well the performance of different groups of pilots, to further improve its detection accuracy and generalization ability. Meanwhile, we will explore the possibility of combining deep learning with other technologies to develop a more intelligent, efficient, and reliable vestibular illusion detection system, providing a solid guarantee for flight safety.

REFERENCES

[1] Kumar, Ravi, et al. "Imitation Learning with Human Eye Gaze via Multi-Objective Prediction." arXiv preprint arXiv:2102.13008, 2023.

[2] Lewkowicz R, Biernacki M P. A survey of spatialdisorientation incidence in Polish military pilots[J]. Int J Oc-cup Med Env,2020, 33 (6): 791-810.

[3] Zhong, Shanshan, et al. "Switchable Self - attention Module." Computer Vision and Pattern Recognition 2022.

[4] Agarwal, Rohit, et al. "packetLSTM: Dynamic LSTM Framework for Streaming Data with Varying Feature Space." arXiv preprint arXiv:2410.17394, 2024.

[5] Lee, Yerin, et al. "Pupil Detection and Segmentation for Diagnosis of Nystagmus with U-Net." 2022 International Conference on Electronics, Information, and Communication (ICEIC) 2022.

[6] Wu Xiang ， Yu Shen ， Shen Shuang ， et a1 ． Quantitative analysis of the biomechanical response of semicircular canals and nystagmus under different head positions[J] ． Hearing Research ， 2021，407：108282.

[7] Wang C, Guo D, Jia H, et al. Simulation and verification of avestibular perception model[C]. //Proceedings of the Interna-tional Conference on Man-Machine-Environment System Engi-neering. Berin: Springer， 2020.149-156.

[8] Sungho Kim, May Jorella Lazaro & Yohan Kang (2023) Galvanic vestibular stimulation to counteract leans illusion: comparing step and ramped waveforms, Ergonomics, 66:4, 432-442, DOI: 10.1080/00140139.2022.2093403

[9] Newman RL, Rupert AH. The magnitude of the spatial disorientation problem in transport airplanes. Aerosp Med Hum Perform. 2020; 91(2):65-70.

[10] Mouelhi, Aymen, et al. "Sparse classification of discriminant nystagmus features using combined video-oculography tests and pupil tracking for common vestibular disorder recognition." Computer Methods in Biomechanics and Biomedical Engineering 24.4 (2020): 400-418.