Proceedings of CONF-CIAP 2026 Symposium: Applied Mathematics and Statistics
DOI: 10.54254/2753-8818/2026.CH30688

Research on Precise Positioning and Prediction of Intelligent Biopsy Devices Based on Machine Learning Models

Yifan Wang[1*], Na Lin[1], He Sun[1], Chenyu Liu[1], Dongjun He[1]

[1]School of Computer Science and Technology, Harbin University of Science and Technology, Harbin, China

**Corresponding Author. Email: 3526602705@qq.com*

Abstract. In the field of clinical diagnosis, biopsy, as a core technology for obtaining tissue samples for pathological analysis, its positioning accuracy directly affects the accuracy of diagnostic results and the rationality of treatment plans. In view of the deficiencies of the existing biopsy localization techniques, this paper proposes the HLOA-CNN-BIGRU regression algorithm, aiming to achieve precise localization of intelligent biopsy devices. During the research process, correlation analysis and violin graph analysis were first conducted, and then the performance of the proposed model was verified through comparative experiments with various machine learning algorithms. The experimental results show that the comprehensive prediction performance of the HLOA-CNN-BIGRU model significantly surpasses that of other comparison models. In terms of error evaluation, the mean square error and root mean square error of this model reached 0.037 and 0.193 respectively, both being the minimum values among all the test models, demonstrating the optimal square error control ability and overall deviation suppression effect. The average absolute error is 0.154, only slightly higher than the 0.151 of the AdaBoost model, which is at the top level and demonstrates an outstanding ability to control the average error. The average absolute percentage error was 7.836, which was lower than all other comparison models except the AdaBoost model, demonstrating outstanding ability to suppress relative errors. In terms of model fitting, the coefficient of determination of this model reaches 0.889, which is not only higher than that of ensemble learning models such as Random Forest, AdaBoost, GBDT, CatBoost, and XGBoost, but also far exceeds decision tree and linear regression models. It can explain the data variation pattern to the greatest extent and has the best fitting effect. The proposal of this algorithm provides an effective solution for improving the positioning accuracy of intelligent biopsy devices and has significant practical significance for optimizing the clinical diagnosis process and perfecting the formulation of treatment plans.

Keywords: Precise Positioning, Intelligent Biopsy Devices, Machine Learning Models

1. Introduction

In the field of clinical diagnosis, biopsy, as the core method for obtaining tissue samples for pathological analysis, its positioning accuracy directly determines the diagnostic accuracy and the

DOI: 10.54254/2753-8818/2026.CH30688

rationality of treatment plan formulation [1]. Traditional biopsy operations are highly dependent on the clinical experience of physicians and are easily affected by factors such as the complexity of human anatomical structures, dynamic displacements of tissues and organs, and limitations of the operating field of view. Positioning deviations often occur, leading to problems such as failed sample collection and repeated punctures. This not only increases the patient's suffering and the risk of complications but may also delay the diagnosis and treatment of diseases [2]. With the popularization of the concept of precision medicine, minimally invasive and intelligent technologies have become the core trends in the development of medical equipment. Intelligent biopsy devices, leveraging their technical advantages such as integrated image guidance and mechanical control, have gradually entered clinical applications. However, existing devices still face technical bottlenecks such as insufficient dynamic positioning accuracy and weak anti-interference ability in complex physiological environments. It is difficult to meet the precise biopsy requirements of early-stage tumors and other diseases. Therefore, improving the positioning accuracy of intelligent biopsy devices has become a research hotspot in the current field of medical engineering [3].

Machine learning algorithms provide key technical support for the precise positioning of intelligent biopsy devices, promoting the transformation of devices from experience-driven to data-driven [4]. During the positioning process, machine learning algorithms can conduct in-depth mining of multi-source perception data and achieve feature extraction and pattern recognition of complex physiological information by constructing data models. For instance, supervised learning algorithms based on image data can achieve precise segmentation and positioning marking of lesion areas, providing target coordinate references for devices. Reinforcement learning algorithms can optimize the puncture path planning through dynamic interactive training and adjust the motion parameters of the robotic arm in real time to adapt to changes in tissue displacement. Unsupervised learning algorithms can mine potential patterns in unlabeled data and enhance the adaptability of the device to the anatomical differences of different patients. However, existing single machine learning models have limitations when dealing with spatio-temporal correlation information of multimodal data, making it difficult to balance positioning accuracy and real-time performance, and thus unable to fully cope with the dynamic complexity of clinical scenarios [5].

In view of the deficiencies of the existing technology, this paper proposes the HLOA-CNN-BIGRU regression algorithm for the precise positioning task of intelligent biopsy devices. This algorithm integrates the technical advantages of convolutional neural networks and bidirectional gated recurrent units. It efficiently extracts spatial features from the data through the CNN module and captures temporal dynamic information during the puncture process using the BiGRU module, achieving a deep fusion of spatio-temporal dimension features. Meanwhile, the hybrid lion swarm optimization algorithm is introduced to globally optimize the model parameters, solving the problem that traditional optimization methods are prone to fall into local optimal solutions, and improving the convergence speed and prediction accuracy of the model. This algorithm aims to break through the positioning bottleneck of existing models in complex physiological environments, enhance the robustness of intelligent biopsy devices against dynamic interference factors, further improve the accuracy and stability of puncture positioning, provide more reliable technical support for clinical precise biopsy, and promote the in-depth application of intelligent medical equipment in the field of pathological diagnosis.

2. Data from data analysis

This study focuses on the precise positioning of intelligent biopsy devices, and the data contains a total of 728 sample records. This dataset contains 7 independent variables with clinical practical

DOI: 10.54254/2753-8818/2026.CH30688

significance and 1 core predictor variable. The independent variables include the distance between the target tissue and the probe, the Angle between the target tissue and the probe axis, tissue hardness, probe working frequency, operating pressure, patient body fat percentage and lesion size. The predictor variable is the positioning error. The dataset can be directly used for the training and validation of machine learning regression models, providing support for the prediction research on the positioning accuracy of intelligent biopsy devices.

Select some data for display, as shown in Table 1.

Table 1. Some of the data

distance	angle	tissue_hardness	probe_frequency	operation_pressure	body_fat	lesion_size	localization_error
2.19	18.1	37.6	11.5	7.6	17.1	0.74	2.22
4.78	14.7	37.1	14.4	9.7	37.3	0.8	3
3.79	14.8	18.6	13.7	3	24	0.85	2.73
3.19	73.3	36.4	12.7	7	24	1.34	2.22
1.2	59.9	25.8	12.9	7.7	32.8	0.44	2.23

Output the heat map of the correlation between each variable and the positioning error, as shown in Figure 1. From the analysis, it can be seen that distance, operation_pressure and body_fat have a weak positive correlation with localization_error. On the other hand, angle, tissue_hardness, probe_frequency, lesion_size are weakly negatively correlated with localization_error. Therefore, as these variables increase, the positioning error in millimeters usually decreases.

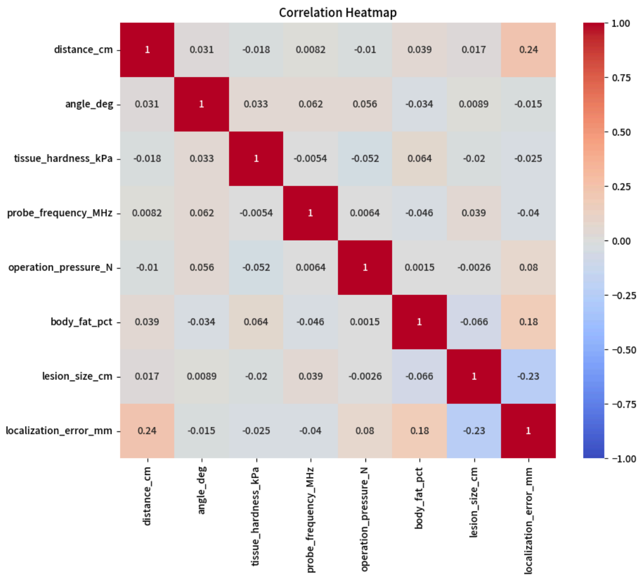


Figure 1. Correlation heat map

Output the violin plots of each variable to observe the distribution of the data, as shown in Figure 2.

DOI: 10.54254/2753-8818/2026.CH30688

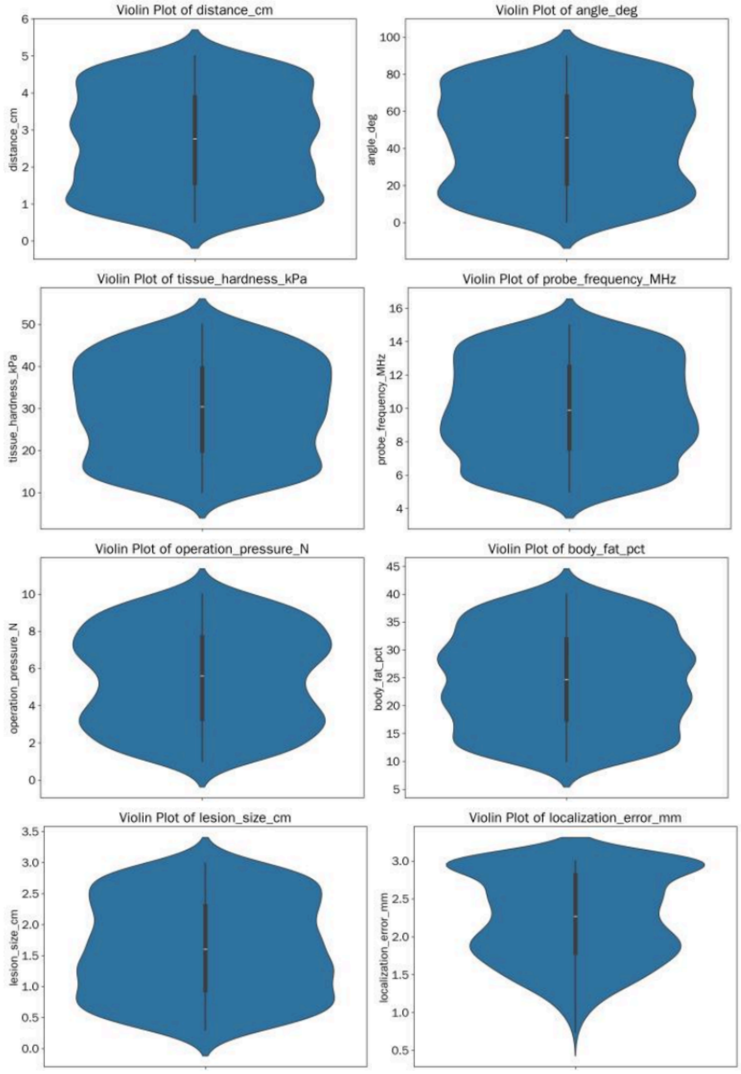


Figure 2. Violin plots of key indicators

3. Method

3.1. HLOA

Hybrid Lion Pride Optimization Algorithm (HLOA) is a new type of intelligent optimization algorithm improved based on the traditional lion pride optimization algorithm. It achieves performance improvement by simulating the social behaviors of lion pride such as hunting and breeding and integrating the core mechanisms of other optimization algorithms [6]. When solving complex optimization problems, the traditional lion swarm algorithm is prone to fall into local optimal solutions and has a slow convergence speed. HLOA has made dual improvements to address this defect: on the one hand, it introduces the global search strategy of the particle swarm optimization algorithm to expand the search range through information sharing among individuals in the population; On the other hand, optimize the design of the fitness function to enhance the algorithm's ability to explore complex solution Spaces [7].

DOI: 10.54254/2753-8818/2026.CH30688

3.2. CNN

Convolutional Neural Network (CNN) is a deep learning model specifically designed for processing grid-structured data. With three core mechanisms - local receptive fields, weight sharing, and pooling operations - it has demonstrated outstanding performance in feature extraction tasks [8]. Its network structure is usually composed of an input layer, a convolutional layer, a pooling layer and a fully connected layer. The convolutional layer performs convolution operations with the local area of the input data through the convolution kernel to generate a feature map containing spatial features. The pooling layer downsamples the feature map, reducing the data dimension and computational load while retaining key information. The fully connected layer maps the extracted features to specific output results [9]. The network structure of CNN is shown in Figure 3.

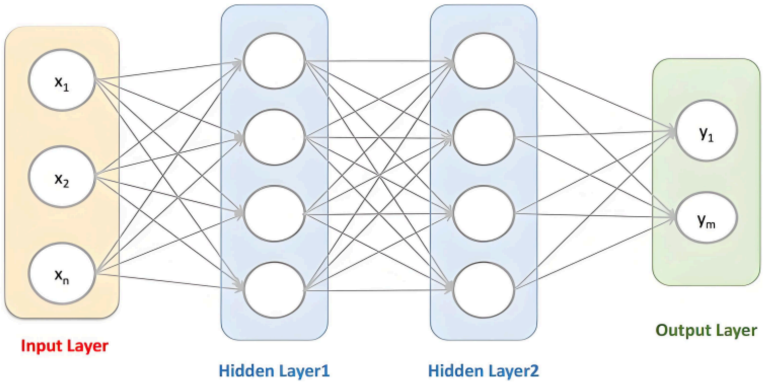


Figure 3. The network structure of CNN

3.3. BIGRU

The Bidirectional Gated Recurrent Unit (BiGRU) is an extended time series data processing model based on the gated recurrent Unit (GRU). By introducing a bidirectional propagation mechanism, it effectively solves the problems of vanishing gradients and incomplete capture of time series information in traditional recurrent neural networks (RNNS). The core structure of BiGRU includes an update gate and a reset gate. The update gate controls the retention of historical information and the integration of new information, while the reset gate determines whether to ignore historical information. Through the collaborative effect of the two gates, precise capture of key temporal features is achieved. Unlike unidirectional GRUs that can only utilize past time series information, BiGRU consists of a forward GRU and a backward GRU. The forward GRU processes information from the beginning of the sequence to the current moment, while the backward GRU processes information from the current moment to the end of the sequence. Eventually, the bidirectional features are fused and output [10]. The network structure of BIGRU is shown in Figure 4.

DOI: 10.54254/2753-8818/2026.CH30688

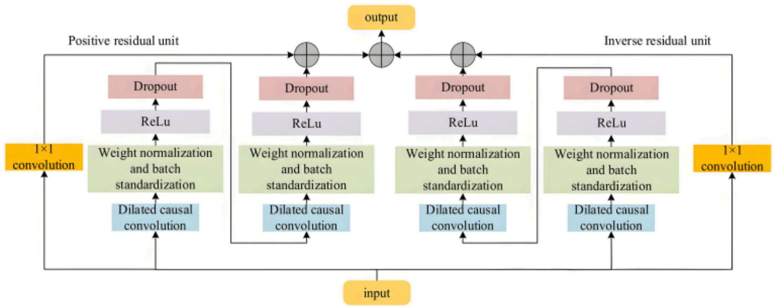


Figure 4. The network structure of BIGRU

3.4. HLOA-CNN-BIGRU

The HLOA-CNN-BIGRU algorithm is a hybrid deep learning algorithm designed for the precise positioning task of intelligent biopsy. By organically integrating the technical advantages of HLOA, CNN and BiGRU, an integrated model with the capabilities of spatial feature extraction, time series information capture and parameter optimization is constructed. The core architecture of this algorithm is divided into three major modules: Firstly, the CNN module is utilized to perform multi-scale convolution operations on the data; Secondly, the dynamic time series data during the puncture process is processed through the BiGRU module to capture the temporal dimension change patterns such as tissue displacement and equipment movement, achieving a deep integration of spatio-temporal features. Finally, the HLOA algorithm is introduced to globally optimize the network parameters of CNN and BiGRU. By simulating the search behavior of lion pride, key parameters such as weights and learning rates are adjusted to solve the problem that traditional parameter optimization methods are prone to fall into local optimum.

4. Result

In this project, the parameters of the HLOA algorithm are set as follows: the number of search agents 8, the maximum number of iterations 6, the optimization dimension 3, the learning rate range 1e-4 to 1e-2, the number of BiGRU neurons range 10 to 30, and the L2 regularization coefficient range 1e-4 to 1e-1. The model training adopts the Adam optimizer, with a maximum of 500 training rounds. The learning rate is scheduled in a segmented manner, the attenuation factor is 0.1, and the attenuation period is 400. The L2 regularization coefficient is obtained through HLOA optimization. The proportion of the training set in the data division is 0.7, and the activation function adopts ReLU.

In this paper, Decision tree, Random Forest, AdaBoost, GBDT, CatBoost, XGBoost and Linear regression are used as comparison models to output the experimental results, as shown in Table 1. From the perspective of various performance indicators, the HLOA-CNN-BIGRU model proposed in this paper significantly outperforms other comparison models in terms of comprehensive prediction performance. In the error evaluation dimension, the MSE and RMSE of this model reached 0.037 and 0.193 respectively, both being the lowest values among all models, indicating that it performs best in controlling the square error between the predicted value and the true value and the overall deviation degree. The MAE is 0.154, only slightly higher than the 0.151 of the AdaBoost model, which is at the top level among all comparison models, reflecting its excellent mean absolute error control ability. The MAPE was 7.836, lower than that of all models except AdaBoost, demonstrating a strong ability to suppress relative errors. In terms of model fitting, the R^2 value of

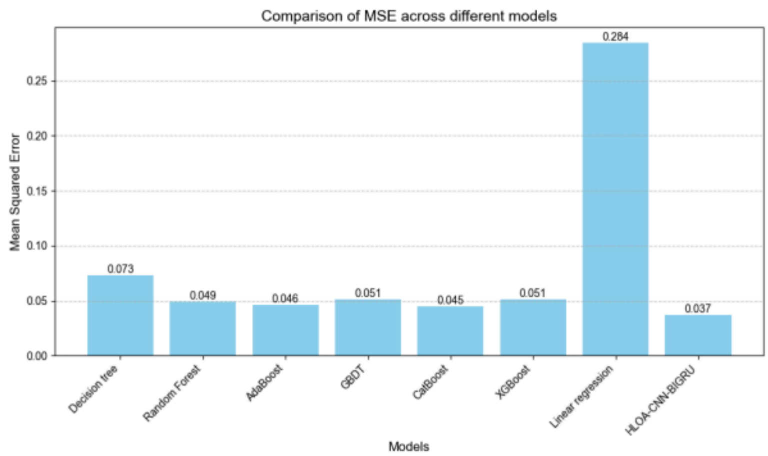
DOI: 10.54254/2753-8818/2026.CH30688

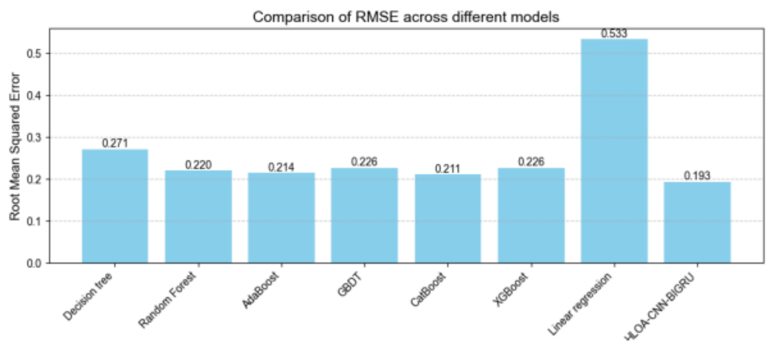
HLOA-CNN-BIGRU reaches 0.889, which is higher than that of all ensemble learning models including Random Forest, AdaBoost, GBDT, CatBoost, and XGBoost, and far exceeds decision tree and linear regression models. It indicates that this model can explain the variation pattern in the data to the greatest extent and has the best fitting effect.

Table 2. The results of each comparative experiment

Model	MSE	RMSE	MAE	MAPE	R^2
Decision tree	0.073	0.271	0.205	10.312	0.792
Random Forest	0.049	0.22	0.172	8.518	0.843
AdaBoost	0.046	0.214	0.151	7.478	0.871
GBDT	0.051	0.226	0.176	8.235	0.844
CatBoost	0.045	0.211	0.175	8.48	0.867
XGBoost	0.051	0.226	0.17	8.473	0.856
Linear regression	0.284	0.533	0.455	20.528	0.123
HLOA-CNN-BIGRU	0.037	0.193	0.154	7.836	0.889

Output the bar chart comparison results of the indicators for each model, as shown in Figure 5.





DOI: 10.54254/2753-8818/2026.CH30688

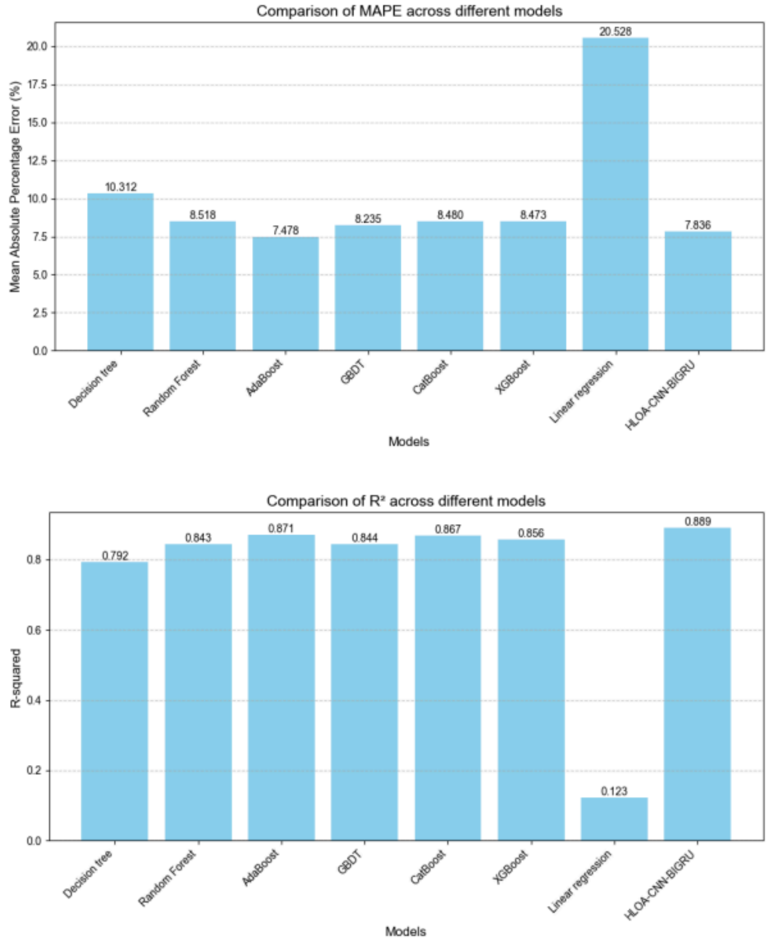


Figure 5. The bar chart comparison results of the indicators for each model

Output the line graph of the predicted values - actual values of the test set of the HLOA-CNN-BIGRU model proposed in this paper, as shown in Figure 6.

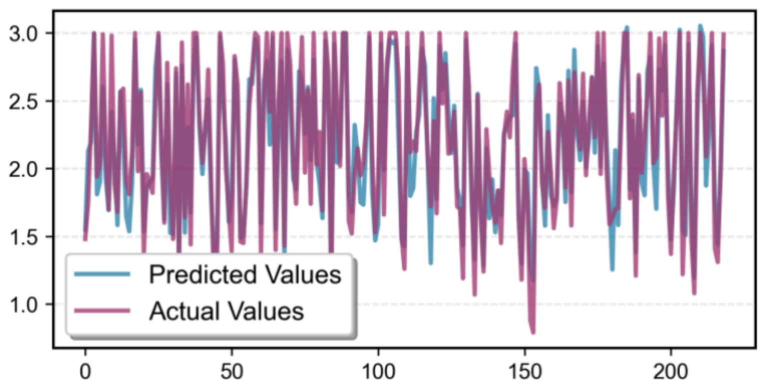


Figure 6. The line graph of the predicted values - actual values of the test set of the HLOA-CNN-BIGRU model

DOI: 10.54254/2753-8818/2026.CH30688

5. Conclusion

In clinical diagnosis, biopsy is a key method for obtaining tissue samples for pathological analysis. The accuracy of its positioning plays a decisive role in the accuracy of the diagnostic results and the rationality of the treatment plan. In view of the shortcomings of the existing technology, this study designed the HLOA-CNN-BIGRU regression algorithm to achieve the precise positioning of the intelligent biopsy device. The research first conducted correlation analysis and violin graph analysis, and then carried out comparative experiments through multiple machine learning algorithms. In terms of performance indicators, the comprehensive predictive ability of this model is significantly superior to other comparison models: In terms of error performance, its MSE and RMSE are as low as 0.037 and 0.193 respectively, the lowest among all models, and it has the best control effect on square error and overall deviation. The MAE reached 0.154, only slightly higher than the 0.151 of the AdaBoost model, which is at the top level, and the average absolute error control ability is excellent. The MAPE is 7.836, which is superior to all other models except AdaBoost, and has a strong ability to suppress relative errors. In terms of fitting degree, the R^2 value of this model reaches 0.889, which is higher than that of various ensemble learning models, decision trees and linear regression models. It can explain the data variation pattern to the greatest extent and has the best fitting effect. This research achievement provides reliable algorithmic support for the precise upgrade of intelligent biopsy devices and has significant practical significance for improving the efficiency of clinical diagnosis and the scientific nature of treatment plans.

6. Summary and future work

In future work, first of all, by incorporating samples from different pathological types, patient groups and clinical institutions, the scale and diversity of clinical datasets can be expanded, which will help improve the generalization ability of the model and its adaptability to complex real-world scenarios. Secondly, optimizing the algorithm structure, such as integrating the attention mechanism to enhance the extraction of key spatio-temporal features, can further improve the accuracy and robustness of location prediction. Thirdly, conducting in-depth clinical trials, integrating the proposed algorithm with actual biopsy equipment, and testing the stability, real-time performance and operational safety of the algorithm in clinical practice is the key to promoting the transformation of the algorithm from theoretical research to clinical application.

References

[1] Salvi, Utkarsh , et al. "Decoding Mental Health: Leveraging Machine Learning for Social Media Analysis." #i{Doctoral Symposium on Human Centered Computing} Springer, Singapore, 2025.

[2] Purohit, Shardha , et al. "Analyzing the Impact of Social Media Usage on Mental Health: A Machine Learning Approach." #i{2024 15th International Conference on Computing Communication and Networking Technologies (ICCCNT)} (2024): 1-6.

[3] Azzolina, Danila , et al. "Editorial: Machine learning approaches for monitoring mental health and substance abuse using social media data." #i{Frontiers in Public Health} (2025).

[4] Pachava, Vengalarao , et al. "Machine Learning Analysis of Social Media's Impact on Mental Health of Indian Youth." #i{International Research Journal of Multidisciplinary Scope} 5.2(2024): 623-635.

[5] Kim, Jina. , D. Lee , and E. Park . "Authors' Reply to: Bibliometric Studies and the Discipline of Social Media Mental Health Research Comment on "Machine Learning for Mental Health in Social Media: Bibliometric Study"." #i{Journal of medical Internet research} 23.6(2021): e29549.

[6] Resnik, Philip , et al. "Bibliometric Studies and the Discipline of Social Media Mental Health Research. Comment on "Machine Learning for Mental Health in Social Media: Bibliometric Study" (Preprint)." #i{Journal of medical Internet research} (2021).

DOI: 10.54254/2753-8818/2026.CH30688

[7] Gao, Yan. "Research on the Application of Chinese Language Teaching System based on Artificial Intelligence Technology." 2024 Third International Conference on Distributed Computing and Electrical Circuits and Electronics (ICDCECE). IEEE, 2024.
[8] Sharan, Roneel V. , et al. "Macro-Sleep Staging With ECG-Derived Instantaneous Heart Rate and Respiration Signals and Multi-Input 1-D CNN& #x2013; BiGRU." #i{IEEE Transactions on Instrumentation and Measurement} (2024): 73.
[9] Tang, Yikai , et al. "Degradation Prediction of Proton Exchange Membrane Fuel Cell Based on Multi-Head Attention Neural Network and Transformer Model." #i{Energies (19961073)} 18.12(2025).
[10] Wang, Yuxiao , C. Suo , and Y. Zhao . "Multi-head attention-based variational autoencoders ensemble for remaining useful life prediction of aero-engines." #i{IOP Publishing Ltd} (2024).