Scholar Publishing Group
International Journal of World Medicine
https://doi.org/10.38007/IJWM.2025.060103
ISSN 2789-7133 Vol. 6, Issue 1: 19-27

Multimodal Medical Data Intelligent Classification Method and System Implementation Based on Improved SVM and Similarity Learning Algorithm

Jun Ye[1,a*]

[1]*Electrical and computer engineering, Carnegie Mellon University, Pittsburgh, Pennsylvania, 15213, United States*

[a]*jun.peter.ye@gmail.com*

[*]*corresponding author*

Keywords: Intelligent diagnosis and treatment, multimodal electronic medical records, patient similarity classification, GPT model, XGBoost method

Abstract: Intelligent diagnosis and treatment play a crucial role in the era of medical big data, especially in processing multimodal electronic medical record data. This article focuses on the patient similarity classification task and proposes two effective classification methods to address challenges such as poor data integrity and diverse expressions. A patient similarity learning method based on multimodal summary extraction is proposed by utilizing the GPT model for TLDR summary extraction and combining it with an improved Support Vector Machine (SVM). A patient similarity learning method based on unstructured text approximate matching is proposed using Jieba segmentation, bag of words model, and improved Jaccard function, combined with XGBoost method. Based on the above method, a patient similarity classification system was designed and implemented, providing doctors with a tool for rapid diagnosis. These methods still have shortcomings, and future work will explore lightweight network architectures, design algorithms that consider dynamically adding patient features and time series, and integrate image classification and text classification to promote further development of intelligent diagnosis and treatment.

1 Introduction

Intelligent diagnosis and treatment are pivotal to advancing medical information systems. By harnessing artificial intelligence technology, particularly deep learning and data mining techniques, it enables the comprehensive analysis of medical text records. This facilitates the provision of efficient and accurate auxiliary diagnosis and treatment plans for healthcare professionals. With the explosive growth of medical data, traditional manual processing methods are no longer able to meet the dual demands of efficiency and accuracy in modern healthcare. Electronic Health Records (EHR), as the core component of medical information technology, comprehensively record patients' medical information and play an irreplaceable role in medical decision-making, disease monitoring, and clinical research. The unstructured nature, complexity, and abundance of specialized terminology and specific semantic backgrounds in medical texts pose significant challenges for automated processing and analysis. Patient similarity classification, as a fundamental problem in medical informatics, has profound value in predicting patients' future health status, guiding precise medical decisions, and developing personalized treatment plans. Although the widespread application of machine learning methods in natural language processing, computer vision, and other fields has promoted significant progress in patient similarity classification research in recent years, existing methods still have many limitations in processing medical data, such as insufficient utilization of structured and semantic information, and neglect of patient attribute information. How to effectively mine and integrate this information to improve the accuracy of patient similarity calculation has become the focus of current research. In the field of text similarity calculation, traditional methods such as edit distance, bag of words model, and topic model have limited effectiveness in processing complex texts. With the rise of deep learning technology, text similarity calculation methods based on word vectors, syntactic structures, semantic role labeling, and deep learning models are gradually emerging. These methods can capture deep semantic information of text and significantly improve the accuracy of similarity calculation. Especially in the medical field, designing efficient deep learning models based on the uniqueness and practical needs of medical data is crucial for achieving precision medicine and smart healthcare services. The successful implementation of this system will provide solid technical support for clinical decision support systems and accelerate the process of medical informatization and intelligence

2 Correlation theory

Recently, the application of machine learning technology in the medical field has become increasingly widespread, demonstrating strong potential. Although a study exploring the reform of undergraduate course teaching in universities, based on machine learning and improved support vector machine (SVM) algorithm, has been withdrawn, it has not hindered the in-depth exploration of SVM in the medical field. The comparative study of SVM and Naive Bayes prediction accuracy in predicting heart disease provides a new perspective for clinical decision-making, which helps doctors to more accurately assess the condition. The perioperative cardiovascular management guidelines released by authoritative institutions such as the American College of Cardiology also emphasize the clinical application value of SVM in non-cardiac surgery, further promoting the application of SVM in medical practice. In the field of ophthalmic disease diagnosis, the combination of Resnet rescaling and support vector machine (Resnet RS SVM) method has achieved significant results in a hospital in a certain country, improving the diagnostic accuracy of retinal injury and providing strong support for early detection and treatment of ophthalmic diseases. The research on entity mapping methods for Chinese electronic medical records has also made significant progress. By integrating similarity algorithms with pre trained models, the information processing flow of electronic medical records has been optimized, and the utilization efficiency of medical data has been improved. In terms of motion target detection, methods based on structural similarity and robust principal component analysis have achieved accurate detection of motion targets, providing effective means for recognition and tracking of motion targets. New solutions have also emerged in the field of multimodal medical data analysis, and the MMGPL method based on graphical prompt learning provides new ideas for processing complex medical data, which helps doctors to have a more comprehensive understanding of patient conditions. The TNF method achieves effective classification of multimodal medical data through three branch neural fusion technology, providing a new tool for the analysis and utilization of medical data. The research on interpretable artificial intelligence methods for multimodal and longitudinal data coordination in medical imaging also provides new ideas for improving the intelligent processing level of medical imaging, which helps doctors to more accurately interpret medical images and improve the accuracy and efficiency of diagnosis. Research on interpretable AI for multimodal medical imaging data offers new ways to improve image processing, aiding doctors and enhancing diagnosis accuracy and efficiency. Machine learning in healthcare is evolving, vitalizing the medical industry.

3 Research method

3.1 TLDR and Advanced Language Models

Similarity classification is a common task in natural language processing and machine learning, widely used in fields such as information retrieval and recommendation systems. However, redundant data may affect the efficiency and accuracy of classification models. To address this issue, researchers have proposed a new method that utilizes TLDR tasks to generate concise text summaries to remove redundancy. This method uses a multi-layer Transformer decoder as the model and trains parameters by maximizing the likelihood function to predict words in a given context. After model training, pre-trained models can be used for label prediction by adjusting parameters to adapt to the supervised target task. Using language modeling as an auxiliary target for fine-tuning can help improve the generalization ability of supervised models. GPT-2 and GPT-3 are two advanced language models, among which GPT-2 adopts a unidirectional Transformer model and conducts unsupervised pre training through a wide and rich corpus, aiming to train high-capacity language models with high-quality training data to complete downstream multitasking unsupervised. The model is shown in Figure 1

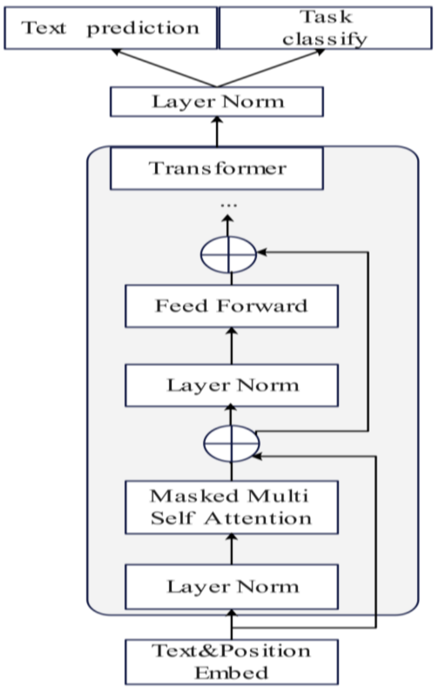

Figure 1 GPT-2 Model Structure

GPT-3 adopts in context learning for training, which can complete other instances in the task by making simple predictions based on given task examples or explanations. In terms of unstructured text vectorization, text vector averaging, as a simple and effective text representation method, has demonstrated excellent performance in many natural language processing tasks. This method segments the text and converts it into corresponding word vectors, then averages all word vectors to obtain the average vector representation of the text, thus preserving the semantic information of the text and suitable for various types of text data

3.2 SVM optimization for multimodal summary similarity learning

Single hot encoding is a commonly used encoding method, especially when processing datasets containing classification features. Due to the non-numeric nature of classification features, directly using these labels as numerical inputs may lead to the model misinterpreting their numerical relationships. Single hot encoding eliminates this partial order relationship by assigning an independent binary vector to each category, ensuring that the model does not make predictions based on incorrect numerical assumptions. This encoding method is simple and intuitive, widely used in data representation and processing in various fields and is often used in conjunction with other data preprocessing techniques such as missing value processing, feature scaling, and feature selection. In the field of machine learning, support vector machines, as a powerful classifier, have the ability to handle high-dimensional data, non-linear separable data, small sample datasets, and flexibly adjust hyperparameters to adapt to different datasets and problems. SVM separates sample points of different categories by finding a hyperplane and maximizes the distance from the hyperplane to the nearest sample point. For multi class classification problems, the One vs One strategy can be adopted, which trains a binary classifier for each class pair and classifies the sample points as the class with the highest number of votes based on the number of votes obtained for each class during prediction. This method can effectively classify and recognize text when dealing with similarity learning tasks in multimodal summary extraction, improving the accuracy and efficiency of the model.

3.3 Multimodal data preprocessing

In the disease dataset, we are faced with subject data containing different numbers and types of biomarkers, aimed at determining whether the subjects have a specific disease. In order to improve the accuracy and reliability of the model, we implemented a multimodal data preprocessing strategy. To address missing values, we deleted rows to ensure data integrity. We simplified data with secondary classification and merged synonyms/duplicates in orthopedic records, excluding those with excessive data loss. For a large disease dataset, we integrated various data types and fixed errors. Techniques like one hot encoding, OCR, and self-supervised learning were used. Datasets were classified by disease severity and split into training/testing sets. Our SVM-G model, validated by accuracy, recall, precision, and F1 score, outperformed others on multi-disease datasets, especially after summary extraction and duplicate removal. These results strongly demonstrate the effectiveness of multimodal data preprocessing in disease datasets, and our proposed similarity learning method shows good applicability on different datasets. In the performance comparison of various methods, our method has consistently demonstrated strong performance. In a set of experiments, our method achieved an accuracy of 89.2% and an F1 score of 88.5%, outperforming SVM (87.7% accuracy, 85.4% F1 score), KNN (65.8% accuracy, 63.6% F1 score), KNN-G (slightly improved over KNN), MCE and MCE-G (F1 score between 50.3% and 53.8%), LRSFS and LRSFS-G (F1 score between 60.1% and 67.2%), as well as SNB and SNB-G (with different performance, SNB-G had a higher F1 score of 65.3%). In another set of experiments, our method achieved an accuracy of 82.5% and an F1 score of 71%, once again outperforming other methods such as SVM (70.2% accuracy, 70% F1 score), KNN (71% accuracy, 60% F1 score), etc. In addition, in previous comparisons, our method achieved an accuracy of 72.9% and an F1 score of 63.1%, demonstrating its consistent advantage in different experimental and evaluation metrics. Overall, in both sets of experiments, our method outperformed other methods in most evaluation metrics, fully demonstrating the powerful potential of multimodal data preprocessing and similarity learning methods in disease dataset analysis.

4 Results and discussion

4.1 Medical text classification model based on approximate matching of unstructured text

In comparative experiments, we comprehensively evaluated the performance of multiple classifiers on specific datasets, covering key metrics such as accuracy, recall, precision, and F1 score. The method proposed in this chapter achieved an accuracy of 96.4%, a significant improvement compared to the baseline classifier XGBoost's 92.8%. It also showed a clear advantage in accuracy, reaching 74.3%. Although the recall rate of the method in this chapter is slightly lower at 61.7% than that of XGBoost at 43.2%, considering F1 67.4%, the method in this chapter is still superior to XGBoost at 52.7%. Traditional classifiers such as SVM, KNN, MCE, LRSFS, and SNB were compared in terms of their performance when using an improved Jaccard similarity calculation marked as "- J". The results are shown in Figure 2.

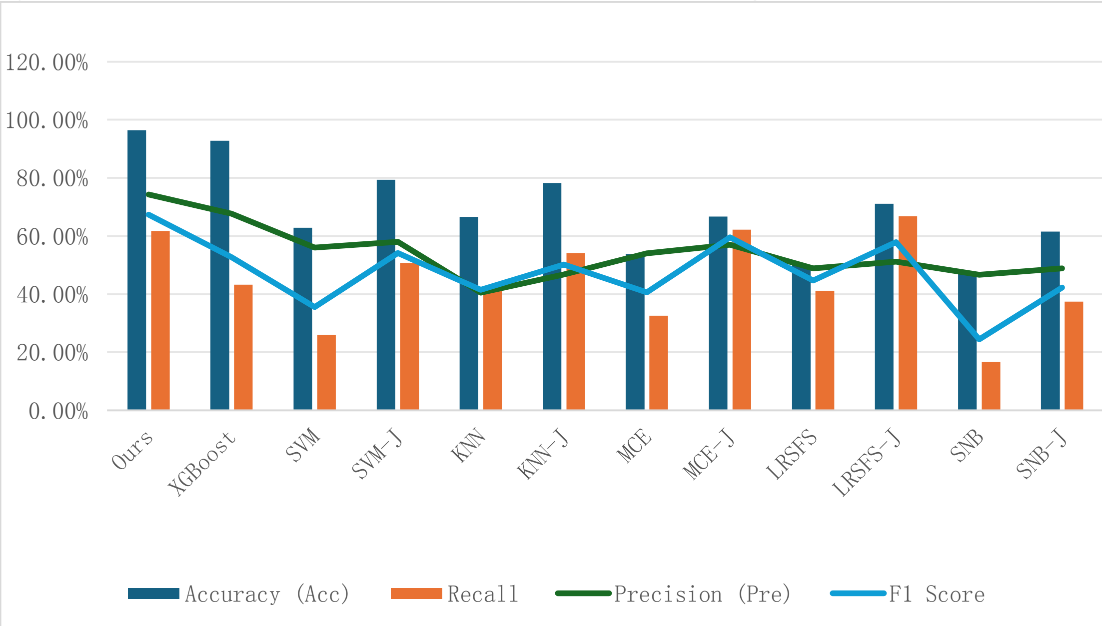

Figure 2 Comparison of performance indicators for different methods

The results showed that SVM-J, KNN-J, MCE-J, LRSFS-J, and SNB-J had significant improvements in multiple indicators, especially the accuracy of LRSFS-J increased by about 21%, reaching 71.1%, which further verified the effectiveness of the improved Jaccard similarity calculation. In experiments on orthopedic datasets and pneumonia (including COVID-19 related) datasets, we used XGBoost-G model based on XGBoost and KNN-G model based on K-Nearest Neighbors for text similarity calculation. The XGBoost-G model has achieved significant improvements in all indicators. Compared with the best performing KNN-G model, the accuracy on the traditional Chinese medicine orthopedic dataset has increased by 7.8%, the recall has increased by 11.9% (another similar indicator such as accuracy has also increased by 25.6%), and the F1 score has increased by 19.4%; On the COVID-19 dataset, despite a decrease in recall and F1 score, mainly due to sample imbalance, accuracy and another accuracy metric still improved by 5.5% and 6.7%, respectively. This further validates the effectiveness of the XGBoost-G model in text similarity calculation. Meanwhile, after unstructured processing of the data, the performance indicators of all algorithms have been improved, once again proving the effectiveness of the text similarity calculation method proposed in this chapter. Our algorithm performs well in dataset classification, improving the Jaccard similarity of unstructured text and increasing accuracy without additional time, providing strong support for patient similarity methods that integrate structured and unstructured data.

4.2 Design and development of patient similarity classification system based on python

With the rapid development of medical information technology, multimodal electronic medical record data has become an important resource in clinical medicine, providing doctors with comprehensive patient information. The classification of patient types plays a crucial role in daily life and medical practice. However, algorithm models in the theoretical stage are difficult to directly solve problems in practical applications. In order to facilitate and quickly diagnose patients, this study designed and developed a patient similarity classification system based on Python. The system adopts sklearn22.2.1+python3.7 as the machine learning framework, combined with Django framework and SimpleUI for system development. The development environment includes Python 3.7 as the development language, PyCharm 2022.2.1 as the development tool, Django as the development framework, and a hardware environment consisting of Inter i7-13700KF CPU, NVIDIA GeForce RTX 2080s GPU, and 16.0G memory. The system is based on Django's Model Template View (MTV) framework pattern, which redirects HTTP requests to the corresponding view processing function through a URL mapper to achieve request processing and response. The model defines the data structure of the application and provides database management functionality, while templates are used to dynamically generate and display page content. This system has been developed and tested in a development environment, and a Python virtual environment was used for configuration. After starting the project, the system page can be accessed by accessing the specified website, providing strong support for practical applications in the medical field.

4.3 Comparative analysis of evaluation effects

This study designed and implemented a machine learning based patient similarity classification system, which mainly includes three functional modules: medical record import, abstract extraction, and classification processing. The medical record import module allows medical staff to upload various patient data to the system and provides the function of updating data, so that patients can update their medical record information in real time after receiving new examinations and tests. The abstract extraction module utilizes the TLDR API to extract abstracts from patients' unstructured text records, reducing redundant information and facilitating doctors to quickly analyze the condition. The classification processing module performs preliminary classification based on the patient's examination results, dividing the patient into three categories: "normal", "severe", and "critical", and extracting key information to display in the "intelligent diagnosis" column to help doctors determine the condition. The system interface design is intuitive and easy to use, and medical staff can complete tasks such as importing medical records, extracting abstracts, and classifying through simple operations. In addition, the system also supports doctors to improve and modify preliminary diagnostic results, ensuring the accuracy and completeness of diagnostic results. The implementation of this system fully utilizes the convenience of human-computer interaction, which can automatically establish a patient similarity network, improve the efficiency of doctor diagnosis, and benefit both patients and hospitals.

5 Conclusion

Intelligent diagnosis and treatment play a crucial role in the era of medical big data, especially in the processing and analysis of multimodal electronic medical record data. This article focuses on the task of patient similarity classification, proposes two effective classification methods to address the challenges of multimodal electronic medical record data, and designs and implements a patient similarity classification system. This paper proposes a patient similarity learning method based on multimodal summary extraction to address issues such as poor data integrity, diverse expressions, and high repetition. The method utilizes a GPT model to implement TLDR summary extraction and combines an improved SVM support vector machine for classification, achieving high classification accuracy. In order to fully utilize the potential relationships between unstructured data and data elements and improve the accuracy and efficiency of text similarity calculation, this paper proposes a patient similarity learning method based on approximate matching of unstructured text. It adopts Jieba segmentation, bag of words model, and improved Jaccard function, combined with XGBoost method for classification, further improving the classification effect. Based on the above methods, this article also designs and implements a patient similarity classification system, providing doctors with a convenient and fast tool for patient similarity classification, which helps to achieve rapid diagnosis. These methods still have some shortcomings. Future work will focus on the following aspects: for the patient similarity learning method based on multimodal summary extraction, we will explore the construction of a lightweight network architecture to shorten the classification training time and maintain ideal classification performance. We'll design an algorithm for patient similarity learning that considers dynamic features and time series, integrating image and text classification. The system will align with clinical needs, gather user feedback, expand functions, and advance intelligent diagnosis and treatment.

References

*[1] Gu, Yiting. "Practical Approaches to Develo**High-performance Web Applications Based on React." Frontiers in Science and Engineering 5.2 (2025): 99-105.*

[2] Shi, Chongwei. "Ovarian Hereditary Diseases: Progress in Prevention and Treatment and Research on Prenatal Diagnosis." Scientific Journal of Technology 7.2 (2025): 125-131.

[3] Shi, Chongwei. "Research on Gene Identification Algorithms Based on Signal Processing Techniques." 2024 6th International Conference on Artificial Intelligence and Computer Applications (ICAICA). IEEE, 2024.

[4] Liu, Boyang. "Study on the Frequency of Computer Language Use Based on Big Data Analysis." Academic Journal of Computing & Information Science 7.10 (2024).

[5] Yang J. Research on the Strategy of MedKGGPT Model in Improving the Interpretability and Security of Large Language Models in the Medical Field[J]. Academic Journal of Medicine & Health Sciences, 5(9): 40-45.

[6] Chen, H., Wang, Z., & Han, A. (2024). Guiding Ultrasound Breast Tumor Classification with Human-Specified Regions of Interest: A Differentiable Class Activation Map Approach. In 2024 IEEE Ultrasonics, Ferroelectrics, and Frequency Control Joint Symposium (UFFC-JS) (pp. 1-4). IEEE.

[7] Varatharajah, Y., Chen, H., Trotter, A., & Iyer, R. K. (2020). A Dynamic Human-in-the-loop Recommender System for Evidence-based Clinical Staging of COVID-19. In HealthRecSys@ RecSys (pp. 21-22).

[8] Cao, Y., Cao, P., Chen, H., Kochendorfer, K. M., Trotter, A. B., Galanter, W. L., ... & Iyer, R. K. (2022). Predicting ICU admissions for hospitalized COVID-19 patients with a factor graph-based model. In Multimodal AI in healthcare: A paradigm shift in health intelligence (pp. 245-256). Cham: Springer International Publishing.

[9] Chen, H., Zuo, J., Zhu, Y., Kabir, M. R., & Han, A. (2024). Generalizable Deep Learning for Pulse-echo Speed of Sound Imaging via Time-shift Maps. In 2024 IEEE Ultrasonics, Ferroelectrics, and Frequency Control Joint Symposium (UFFC-JS) (pp. 1-4). IEEE.

[10] Zhao F. Application and Performance Improvement of K-Means Algorithm in Collaborative[C]//2025 International Conference on Intelligent Systems and Computational Networks (ICISCN). IEEE, 2025: 1-6.

[11] Fan, Sunjia, et al. "Defense methods against multi-language and multi-intent LLM attacks." International Conference on Algorithms, High Performance Computing, and Artificial Intelligence (AHPCAI 2024). Vol. 13403. SPIE, 2024.

[12] Maqsood S, Damaševičius R, Maskeliūnas R. Multi-modal brain tumor detection using deep neural network and multiclass SVM[J]. Medicina, 2022, 58(8): 1090.

[13] Zhu P. Construction and Experimental Verification[C]//Cyber Security Intelligence and Analytics: The 6th International Conference on Cyber Security Intelligence and Analytics (CSIA 2024), Volume 1. Springer Nature, 2025, 1351: 391.

[14] Dong P. Research on the Application of Knowledge Graph-Driven Two-Dimensional Convolutional Embedding Methods in Recommender Systems[C]//2025 International Conference on Intelligent Systems and Computational Networks (ICISCN). IEEE, 2025: 1-6.

[15] Amal S, Safarnejad L, Omiye J A, et al. Use of multi-modal data and machine learning to improve cardiovascular disease care[J]. Frontiers in cardiovascular medicine, 2022, 9: 840262.

[16] Abid M H, Ashraf R, Mahmood T, et al. Multi-modal medical image classification using deep residual network and genetic algorithm[J]. Plos one, 2023, 18(6): e0287786.

[17] Navaz A N, T. El-Kassabi H, Serhani M A, et al. A novel patient similarity network (PSN) framework based on multi-model deep learning for precision medicine[J]. Journal of Personalized Medicine, 2022, 12(5): 768.

[18] Kumar S, Rani S, Sharma S, et al. Multimodality Fusion Aspects of Medical Diagnosis: A Comprehensive Review[J]. Bioengineering, 2024, 11(12): 1233.

[19] Chen, H., Varatharajah, Y., de Ramirez, S. S., Arnold, P., Frankenberger, C., Hota, B., & Iyer, R. (2020). A retrospective longitudinal study of COVID-19 as seen by a large urban hospital in Chicago. medRxiv, 2020-11.

[20] Chen, H., Zhu, Y., Zuo, J., Kabir, M. R., & Han, A. (2024). TranSpeed: Transformer-based Generative Adversarial Network for Speed-of-sound Reconstruction in Pulse-echo Mode. In 2024 IEEE Ultrasonics, Ferroelectrics, and Frequency Control Joint Symposium (UFFC-JS) (pp. 1-4). IEEE.

*[21] Yang, **zhu. "Application of Multi-model Fusion Deep NLP System in Classification of Brain Tumor Follow-Up Image Reports." Cyber Security Intelligence and Analytics: The 6th International Conference on Cyber Security Intelligence and Analytics (CSIA 2024), Volume 1. Vol. 1351. Springer Nature, 2025.*

[22] Chen, H., Ma, K., & Shen, J. (2024). Interpretable Machine Learning Facilitates Disease Prognosis: Applications on COVID-19 and Onward. International Journal of Computer Science and Information Technology, 3(3), 428-436.

[23] Zhang, Yiru. "Design and Implementation of a Computer Network Log Analysis System Based on Big Data Analytics." Advances in Computer, Signals and Systems,(2024) 8(6),40-46.

[24] Liu, Yu. "Build an Audit Framework for Data Privacy Protection in Cloud Environment." Procedia Computer Science 247 (2024): 166-175.

[25] Liu, Boyang. "Design and Application of Experimental Data Management System Integrating Remote Monitoring and Historical Data Analysis." Journal of Electronics and Information Science 9.3 (2024): 160-167.

[26] Yang, Jinzhu "Integrated Application of LLM Model and Knowledge Graph in Medical Text Mining and Knowledge Extraction."Social Medicine and Health Management (2024), 5(2): 56-62

[27] Zhao, Fengyi. "Risk Assessment Model and Empirical Study of in Vitro Diagnostic Reagent Project Based on Analytic Hierarchy Process." International Journal of New Developments in Engineering and Society 8.5 (2024), 76-82

[28] Wang, Yuxin "Research on Intelligent Macro Image Recognition Algorithm of Oil Pipe Failure Based on Deep Learning." Journal of Image Processing Theory and Applications (2025), 8(1): 1-7

[29] Xu, Y. (2025). Research on Maiustream Web Database Development Technclogy. Journal of Computer Science and Artificial Intelligence, 2(2),29-32

[30] Zhao, Fengyi "Development Design and Signal Processing Algorithm Optimization of Traditional Chinese Medicine Pulse Acquisition System Based on CP301 Sensor." Advances in Computer, Signals and Systems (2024), 8(6): 106-111

[31] Shi C. Research on Deep Learning Algorithms for Predicting DNA-Binding Proteins Based on Sequence Information[C]//2024 IEEE 2nd International Conference on Electrical, Automation and Computer Engineering (ICEACE). IEEE, 2024: 1566-1570.

[32] Xu, Yue. "Research on Graph Network Social Recommendation Algorithm Based on AGRU-GNN." 2024 IEEE 4th International Conference on Data Science and Computer Application (ICDSCA). IEEE, 2024.

[33] Shi, C. (2024). Research on the Application of Computer Technology in Biostatistics. Journal of Computing and Electronic Information Management.Vol. 14, No. 3, 2024,12-15

[34] Wang Y. Design and Implementation of a General Data Collection System Architecture Based on Relational Database Technology[C]//The International Conference on Cyber Security Intelligence and Analytics. Cham: Springer Nature Switzerland, 2024: 561-572.

[35] Zhang, Jinshuo "Research on Real Time Condition Monitoring and Fault Warning System for Construction Machinery under Multi Source Heterogeneous Data Fusion." Journal of Engineering Mechanics and Machinery (2024), 9(2): 139-144‘

*[36] Yang, **zhu. "Research on the Application of Medical Text Matching Technology Combined with Twin Network and Knowledge Distillation in Online Consultation." Frontiers in Medical Science Research, 2024, 6(11), 25-29*