

RESEARCH ARTICLE

Intelligent matching algorithm for application cases of higher vocational medical microbiology laboratory courses

Rong Xu^{1, †}, Xiuqin Wang^{2, †}, Jiachen Liu², Dun Hu², Jin Chen^{2, *}

¹School of Pharmacy, ²School of Medical Technology, Anhui Institute of Medicine, Hefei, Anhui, China.

Received: June 30, 2025; accepted: October 31, 2025.

In the context of digital education, efficient management and intelligent application of teaching resources for higher vocational medical microbiology experimental courses are crucial to improving teaching quality. Currently, platforms supporting these courses often rely on rudimentary matching that fails to mine deep semantic associations between cases or accurately identify similarities in core elements, leading to low matching efficiency. This research proposed an intelligent matching algorithm for application cases in higher vocational medical microbiology laboratory courses. Centered on a structured semantic model, the algorithm employed an "entity-relationship-entity" framework for multi-dimensional case analysis. A case library was constructed through processes including data collection and cleaning, knowledge graph mapping, and semantic enhancement. Targeting the characteristics of long-text case descriptions, a method integrating text summarization extraction and a relevance evaluation mechanism was introduced. Supervised datasets were built by annotating the relevance of text fragments based on core course elements, enabling iterative optimization of the evaluation mechanism for calculating precise case feature weights. For a target case, an information table was constructed, and attribute weights were determined using the knowledge granularity rough set principle combined with expert experience. Similarity was calculated *via* Euclidean distance measurement and subsequently converted into a similarity score. Eventually, a weighted average algorithm comprehensively evaluated similarity across multiple fields, and matching rules were formulated to achieve intelligent case matching. The results demonstrated that the proposed method performed excellently in both matching degree and speed, effectively improving overall matching efficiency. This research provided a robust technical framework for case-based teaching in vocational medical education and offered significant value to the scientific teaching community by enhancing the precision and efficiency of resource retrieval and recommendation in specialized courses.

Keywords: medical microbiology; laboratory course; case library; feature weight; similarity; intelligent matching algorithm.

*Corresponding author: Jin Chen, School of Medical Technology, Anhui Institute of Medicine, Hefei, Anhui, China. Email: chenjin@ahyz.edu.cn.

[†]These authors contributed equally to this work.

Introduction

The cultivation of practical skills is a cornerstone of higher vocational education, particularly in the medical field. Medical microbiology laboratory courses are essential components of this training, designed to equip students with the

competencies required for clinical diagnostics and research. These courses encompass a wide array of core contents including, but not limited to, pathogen detection and identification, immunological serological testing, antimicrobial susceptibility testing, molecular diagnostic techniques, and biosafety practices [1]. The

effective integration of application cases into teaching is a critical pedagogical approach for bridging theoretical knowledge and practical clinical skills.

With the rapid advancement of digital education technologies, the management and utilization of teaching resources have undergone significant transformation. Educational platforms increasingly leverage data-driven methods to enhance learning experiences. In the broader field of information matching, several algorithmic approaches have been proposed, which include text similarity-based algorithms for resuming matching [2], subgraph matching techniques within knowledge graphs to uncover entity relationships [3], similarity metrics like Cosine and Jaccard indices for matching tasks in scientific literature [4], and evolutionary algorithms like the Memetic algorithm to align educational content with learner preferences in e-learning systems [5]. Despite these advances, a significant gap remains in the context of higher vocational medical microbiology education. Existing platforms often support only rudimentary case matching, which primarily relies on superficial keyword overlaps. This approach fails to explore the deep semantic associations between complex case narratives and cannot accurately identify the similarity of core pedagogical elements such as underlying principles, experimental techniques, or clinical scenarios. Consequently, case resources that are often scattered, non-standardized in format, and poorly adapted for teaching are underutilized, which leads to low retrieval and matching efficiency, ultimately hindering the effectiveness of case-based teaching methodologies.

To address these limitations, this study proposed an intelligent matching algorithm specifically tailored for the application cases of higher vocational medical microbiology laboratory courses to enhance the precision and efficiency of case retrieval by constructing a centralized, structured, and extensible application case library and implementing a sophisticated algorithm that could understand and match cases

based on their deep semantic features and pedagogical relevance. The research employed a structured semantic model as the core framework and adopted an "entity-relationship-entity" paradigm for multi-dimensional case analysis. The results of this research would provide a robust, semantically aware framework for intelligent educational resource management in the scientific teaching community beyond traditional keyword-based matching, offer a more nuanced and accurate approach that understands the contextual and pedagogical nuances of medical cases, and improve the accessibility and utility of case-based learning materials for educators, thereby enhancing teaching efficiency and student learning outcomes in vocational medical education. The methodologies developed for semantic analysis and weighted similarity matching in this study could be adapted and applied to other specialized educational domains facing similar challenges with complex, unstructured teaching resources.

Materials and methods

Constructing an application case library

A centralized, structured, and extensible course application case library with a structured semantic model as the core and using the "entity-relationship-entity" framework to analyze the case was constructed in this study. The classification of clinical laboratory scenarios through the case type dimension was standardized, forming an application case system for higher vocational medical microbiology laboratory courses covering common teaching scenarios, which was convenient for teachers to quickly locate the contents. Based on the test feature dimension, the sample types, test methods, and other elements were integrated to standardize the description of application cases in higher vocational medical microbiology test courses. The operation association dimension was introduced to record information and strengthen the understanding of operation [5], where the dimension was annotated through

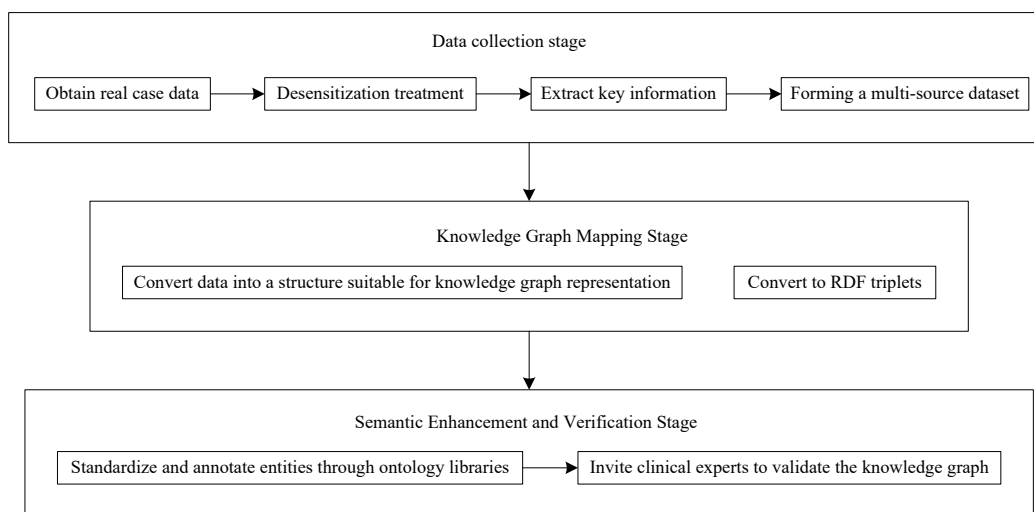


Figure 1. Process of constructing an application case library for higher vocational medical microbiology laboratory courses.

knowledge tracing to ensure the scientificity and teaching value of the case. To improve the retrieval efficiency of the case library and strengthen the knowledge correlation between cases, knowledge graph technology was introduced. Specifically, the application cases of the higher vocational medical microbiology laboratory course were converted into case data that were easy to process. Multi-source data were integrated to build an association network, and a higher vocational medical microbiology laboratory case library based on this network was constructed through data collection, data mapping, and semantic enhancement (Figure 1). In the data collection stage, the real cases were collected by removing sensitive and private information through desensitization, while extracting key information. Meanwhile, the textbooks, literatures, and teacher experience cases of higher vocational medical microbiology laboratory courses were integrated to form a multi-source dataset. In the knowledge graph mapping stage, the application case table structure of the higher vocational medical microbiology laboratory course was converted into resource description framework (RDF) triple form to achieve a structured representation, which was convenient for subsequent processing and analysis. In the semantic enhancement and verification stage, the ontology library was used

to annotate the data, while experts were invited to verify and improve the data quality [6]. During the case library construction, four-tuple model was used to formally represent the application cases of the higher vocational medical microbiology laboratory course to describe the case structure more clearly as follows.

$$C = (I, F, D, \Phi) \quad (1)$$

where I was the collection of application cases of higher vocational medical microbiology laboratory courses. F was the set of case feature attributes. D was a collection of attribute values. Φ was the information function [7]. With the help of this model, the application cases of higher vocational medical microbiology laboratory courses were transformed from raw data into structured semantic knowledge. To ensure the standardization and unification of cases, Chinese text processing technology was adopted to unify the case feature attributes to achieve case comparability and improve retrieval accuracy.

Calculation of feature weights of application cases

The weights of the characteristics of the application cases were calculated. Based on the

insight into the distribution law of long text information, the study assumed that the key information of the higher vocational medical microbiology laboratory course cases was concentrated in a few core sentences, which contained the elements required for analysis. A text summary extraction algorithm based on the characteristics of the course was then proposed to preliminarily screen the key information. In the feature weight calculation stage, a relevance evaluation mechanism based on the characteristics of course case text was introduced. The long text was divided into segments and scored using a bidirectional encoder representation from transformers (BERT) based model. High-scoring segments were more important in weight calculation. In this correlation evaluation mechanism, assuming there were n text blocks (t_1, t_2, \dots, t_n) for text block t^+ , its correlation score $s(t^+)$ was calculated as below.

$$s(t^+) = \psi \left(\sum_{p=1}^P \lambda_p \cdot \text{Focus}(\text{Embed}(t^+), \text{Embed}(d_p)) \right) \quad (2)$$

where ψ was the activation function. λ_p was the p th attention weight. *Focus* was the attention calculation function. *Embed* was the embedding function. d_p was the vector corresponding to the core content of the higher vocational medical microbiology laboratory course. The formula calculated the association between the text block and the reference vector through the attention mechanism and obtained the relevance score of the core content of the higher vocational medical microbiology laboratory course after activation function transformation [8]. The relevance of text fragments was annotated based on the core elements of the course and constructed a supervised label dataset to support iterative optimization of the evaluation mechanism and improve the accuracy of feature weight calculation. The fit between the fragment and the higher vocational medical microbiology laboratory course and the frequency of feature

items were combined to define case features. The weight $u_{q,case}$ of case feature q was defined as follows.

$$u_{q,case} = \frac{\sum_{w_i \in W_{case}} O(w_i \in g_q) \cdot \text{value}(w_i)}{\sum_{w_i \in W_{case}} \text{value}(w_i)} \quad (3)$$

where W_{case} was the set of all feature items in the application cases of higher vocational medical microbiology testing courses. $\text{value}(w_i)$ was the weight of feature items w_i . $O(w_i \in g_q)$ was the indicator function. Based on the weight $u_{q,case}$ of feature q , the final weight $w_{i,final}$ of the feature item could be expressed as follows.

$$w_{i,final} = \beta \cdot TF-IDF(w_i) + (1-\beta) \cdot \sum_{q: w_i \in g_q} u_{q,case} \quad (4)$$

where $w_{i,final}$ was the final weight of feature item w_i in the application case of the medical microbiology testing course in higher vocational education. β was the weight adjustment factor. $TF-IDF(w_i)$ was the TF-IDF value of feature term w_i in the application case of the vocational medical microbiology testing course. g_q was the set of feature terms associated with feature q . Aiming at the long text characteristics of course cases, a feature weight calculation method was proposed in this study to overcome the limitations of BERT and improve matching accuracy and efficiency.

Formulation of matching rules to achieve intelligent case matching

The feature weights of all cases in the case library were calculated to formulate scientific and reasonable matching rules to accurately and efficiently retrieve cases that were identical or most similar to the target case and provide

support for the teaching practice of higher vocational medical microbiology testing courses. In the case matching stage, a case information table was constructed and knowledge granularity rough sets with expert experience were combined to calculate attribute weights to suit teaching practice. The Euclidean distance was used to measure the similarity between the current application cases of higher vocational medical microbiology laboratory courses and the conditional attributes of the cases in the case library. Specifically, let the target application case of higher vocational medical microbiology laboratory courses be X , a case in the case library be Y_p ($p = 1, 2, K, q$). The value of target case X on attribute Z_r ($r = 1, 2, K, e$) was x_r , and the value of case Y_p on attribute Z_r was y_{pr} . Then the Euclidean distance μ_{pr} between case X and case Y_p on attribute Z_r could be expressed as below.

$$\mu_{pr} = |x_r - y_{pr}| \quad (5)$$

The smaller the Euclidean distance was, the closer the values of the two cases on this attribute were, which meant that their characteristics on this attribute were more similar. However, it was not enough to consider the similarity of a single attribute. To comprehensively consider the similarity of all attributes, the weighted Euclidean distance V_p between case X and case Y_p was further calculated. Assuming the weight of attribute Z_r was v_r , then V_p was shown as follows.

$$V_p = \sqrt{\sum_{r=1}^s v_r (x_r - y_{pr})^2} \quad (6)$$

By introducing attribute weights, the overall similarity between two cases could be more reasonably measured. The smaller the V_p , the higher the overall similarity between case X

and case Y_p . The goal of case matching was to retrieve from the case library with the case that was identical or most similar to the current applied case X of the higher vocational medical microbiology examination course [9-11]. In actual operation, when a new medical microbiology test was received, the case library was first searched to see if there was a completely matching case [12-15]. If there was one, the result was directly returned, so that the required cases could be quickly provided for teaching. If there was no completely matching case, a deep search was then performed using a search algorithm based on Euclidean distance [16, 17]. To transform the weighted Euclidean distance V_p to similarity W_p that was convenient for comparison with the similarity threshold, W_p was then expressed as follows.

$$W_p = \frac{\max_{1 \leq l \leq q} V_l}{V_p} \quad (7)$$

where W_p was the similarity between the application case X of the higher vocational medical microbiology examination course and case Y_p . $\max_{1 \leq l \leq q} V_l$ was the maximum weighted Euclidean distance between all cases in the case database and the target application case X of the higher vocational medical microbiology examination course. The distance metric was transformed into a similarity metric to make the matching result judgment more intuitive and easier to understand. In view of the situation where the case of the higher vocational medical microbiology laboratory course contained multiple feature fields, a weighted average algorithm was used to comprehensively evaluate the string similarity [18-20]. Assume that, for a certain feature field, the string similarity for case X and case Y_p was T_{pr}^{str} , the weight of this feature field was $w_{i,final}^{str}$, then the comprehensive similarity P_p^{total} could be expressed as below.

Table 1. Experimental datasets.

Category	Dataset	Sources	Amount	Descriptions
Case dataset	Medical microbiology laboratory case collection	Textbooks, clinical cases, medical papers, <i>etc.</i>	500 pieces	Case number: a unique identifier for each case, used to distinguish different cases
				Case title: a concise summary of the core content of the case
				Chapter: the chapter corresponding to the case in the course
				Related knowledge points: microbiological testing knowledge points involved in the case
				Depth of knowledge points: complexity of knowledge points
				Difficulty level: case difficulty
Student dataset	Student basic information and learning performance dataset	Academic administration system, learning platform, questionnaire survey, <i>etc.</i>	300 pieces	Student ID: unique identifier for each student, used to distinguish different students
				Learning progress: course chapters that the student has completed
				Test scores: including theoretical test scores and practical operation test scores
				Interest level: students' interest ratings for bacterial and viral testing

$$P_p^{total} = \sum_r w_{i,final}^{str} T_{pr}^{str} \quad (8)$$

In the similarity synthesis, normalization was performed first to the similarity of each feature field string T_{pr}^{str} , and then multiplied by weight to obtain comprehensive similarity. Through weighted average, the similarity between the two cases was comprehensively evaluated and then compared with the threshold. If the comprehensive similarity was greater than or equal to the threshold, the match was determined to be successful, otherwise there was no matching case.

Environment configuration and experiment design

This research was conducted on a Dell PowerEdge R740xd server (Dell Technologies, Round Rock, Texas, USA) equipped with 2 Intel Xeon Gold 6248R CPUs (24 cores/48 threads each, 2.4 GHz base frequency, 3.0 GHz turbo frequency), 512 GB DDR4 ECC REG RAM (3,200 MHz), and a storage configuration comprising 4 × 2.4 TB SAS 12 Gbps 10k RPM HDDs in a RAID 10

array alongside 2 × 1.92 TB NVMe SSDs for caching and high-speed storage. The operating system was Ubuntu 22.04 LTS (64 bit). Python 3.10 (<https://www.python.org/>) with key libraries including TensorFlow 2.12 (<https://tensorflow.google.cn/install?hl=zh-cn>), scikit-learn 1.3.0 (<https://scikit-learn.org/>), pandas 2.0.3 (<https://pandas.pydata.org/>), and NumPy 1.25.0 (<https://numpy.org/>) were employed in this research. Jupyter Notebook 7.0.0 (<https://jupyter.org/>) and PyCharm Professional 2023.3 (<https://www.jetbrains.com/pycharm/>) were applied as primary development tools. Data were stored and managed using MySQL 8.0 (<https://www.mysql.com/>) and MongoDB 6.0 (<https://www.mongodb.com/>). This research collected application cases of higher vocational medical microbiology testing courses. The collected experimental data consisted of two datasets including case dataset and student dataset (Table 1). The results of proposed method defined as method 1 were compared to that of several existing methods adopted from previous studies and defined as methods 2 [1], 3 [2], and 4 [4], respectively, to further

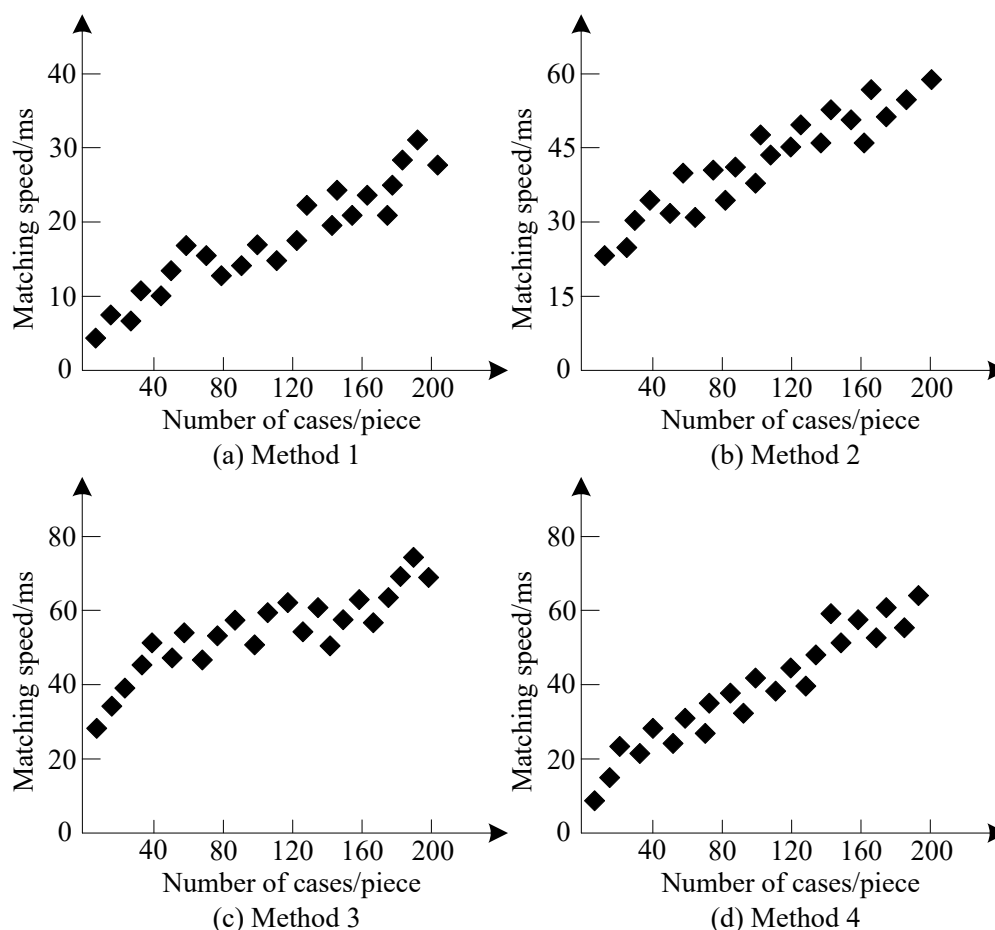


Figure 2. Comparison of matching speed of four matching algorithms in the case of higher vocational medical microbiology testing.

comprehensively evaluate the performance of each algorithm and their matching speed.

Results and discussion

The results showed that the proposed algorithm demonstrated significant advantages in matching speed. The scatter points of method 1 were concentrated in the area of fast matching and short time, indicating that it could process cases very quickly (Figure 2), which was because the algorithm design was optimized, making the proposed algorithm efficient in data structure processing to quickly parse and organize case data, build an efficient index and query mechanism, and greatly shorten the matching time. The results demonstrated that, when the

number of cases increased or the task became more complicated, the matching speed of the proposed method (method 1) remained stable, while the other methods fluctuated greatly with some methods even decreased. The results indicated that the proposed algorithm of this study had better scalability and stability when dealing with large-scale or complex tasks.

This study scientifically constructed a case library suitable for medical microbiology laboratory courses in higher vocational colleges, achieving centralized management of case resources. Through innovative methods, it accurately calculated case feature weights, formulated reasonable matching rules, quantitatively identified features, provided high-quality feature input, and optimized the case retrieval process,

avoiding the problem of scattered case resources and significantly improving the efficiency and accuracy of case matching. The results of this study provided strong support for course teaching practice. By using the "entity relationship entity" framework to analyze the core elements and combining the text summarization and relevance evaluation mechanisms, this research significantly improved the accuracy and matching efficiency of semantic association mining. Future study should explore case matching methods that integrate multimodal data fusion and introduce a dynamic weight adjustment mechanism to further optimize the real-time requirements of interdisciplinary course cases.

Acknowledgements

This work was supported by the Anhui Provincial Higher Education Research Project (Grant No. 2024AH050846, 2024AH050854), Research Foundation for Advanced Talents of Anhui Medical College (Grant No. 2023RC007), Provincial Quality Project for Higher Education Institutions in Anhui Province (Grant No. 2024ahyzjyxm08, 2024yzjc089), and Wang Jianhua Scientific Research and Innovation Team Project of Anhui Medical College (Grant No. WJH2023GGWS002).

References

- Shi Y, Shan J. 2022. Research on resume matching recommendation algorithm based on text similarity. *Comput Simul.* 39(4):441-444.
- Sun Y, Li G, Du J, Ning B, Chen H. 2022. A subgraph matching algorithm based on subgraph index for knowledge graph. *Front Comput Sci.* 16:1-18.
- Rinjeni TP, Indriawan A, Rakhmawati NA. 2024. Matching scientific article titles using cosine similarity and jaccard similarity algorithm. *Procedia Comput Sci.* 234:553-560.
- Delaramifar M, Sargazi Moghadam T. 2025. Matching course content with learner preferences in E-learning systems based on the Memetic algorithm. *Lang Res.* 15(2):125-156.
- Zhang J, Zhang T, Zhang C, Yao Y. 2022. An improved ICCP-based underwater terrain matching algorithm for large initial position error. *IEEE Sens J.* 22(16):16381-16391.
- Yuan Y, Li Z, Liu Z, Yang Y, Guan X. 2021. Double deep Q-network based distributed resource matching algorithm for D2D communication. *IEEE Trans Veh Technol.* 71(1):984-993.
- Hu Y, Jiang W, Tak-Shing PY, Zhang J. 2022. Integrating user suitability and course matching degree for online course recommendation method. *J Comput Res Dev.* 59(11):2520-2533.
- Zhou L, Zhang F, Zhang S, Xu M. 2021. Study on the personalized learning model of learner-learning resource matching. *Int J Inf Educ Technol.* 11(3):143-147.
- Clem E, Dawson V. 2024. The emergence of case matching in discontinuous DPs. *Nat Lang Linguist Theory.* 42(3):955-1002.
- Sun ZX, Yu WJ, Si ZH, Xu J, Dong ZH, Chen X. 2024. Explainable legal case matching via graph optimal transport. *IEEE Trans Knowl Data Eng.* 36(6):2461-2475.
- Mansournia MA, Poole C. 2023. Case-control matching on confounders revisited. *Eur J Epidemiol.* 38(10):1025-1034.
- Xiao Y, Li C, Song L, Yang J, Su J. 2021. A multidimensional information fusion-based matching decision method for manufacturing service resource. *IEEE Access.* 9:39839-39851.
- Solanki JD, Vohra AS, Hirani CN, Bhatt DN. 2024. Arterial stiffness is associated with prehypertension in both non-hypertensives and treated hypertensives-A matched case control study. *Indian Heart J.* 76(3):224-228.
- Kalla P, Namerow LB, Walker SA, Ruaño G, Malik S. 2023. Contrasting *ABCB1* pharmacogenetics and psychotropic responses in child and adolescent psychiatry: A case comparison. *Pharmacogenomics.* 24(3):131-139.
- Trenkler C, Blessing E, Jehn A, Karcher J, Schoefthaler C, Schmidt A, et al. 2024. Retrospective case control matched comparison of the antegrade versus retrograde strategy after antegrade recanalisation failure in complex *de novo* femoropopliteal occlusive lesions. *Eur J Vasc Endovasc Surg.* 67(5):799-808.
- Mosier BR, Bantis LE. 2024. Combining multiple biomarkers linearly to minimize the Euclidean distance of the closest point on the receiver operating characteristic surface to the perfection corner in trichotomous settings. *Stat Methods Med Res.* 33(4):647-668.
- Sha CM, Wang J, Dokholyan NV. 2024. Predicting 3D RNA structure from the nucleotide sequence using Euclidean neural networks. *Biophys J.* 123(17):2671-2681.
- Shao S, Li D. 2025. Retracted: Application of fuzzy prediction control model based on neural network in teaching resource recommendation and matching. *J Intell Fuzzy Syst.* 48(1_suppl):29-44.
- Sharma R, Kumar Mahanti G, Panda G, Singh A. 2024. Thyroid nodules classification using weighted average ensemble and D-CRITIC based TOPSIS methods for ultrasound images. *Curr Med Imaging.* 20(2):e050423215446.
- Wang P, Zhao T, Cao J, Li P. 2024. Soft sensor modeling of self-organizing interval type-2 fuzzy neural network based on adaptive quantum-behaved particle swarm optimization algorithm. *Int J Fuzzy Syst.* 26(5):1716-1729.