

A Review of Deep Learning-Based Classification of Diabetic Retinopathy

Linying Xu

⁴Computer Technology, Zhejiang University, Hangzhou, Zhejiang, China

Abstract: *Diabetic retinopathy (DR) is a leading cause of blindness in diabetic patients, and early accurate classification is critical for disease management. Leveraging automated feature extraction, deep learning has become a core tool for DR classification. This review outlines commonly used datasets and evaluation metrics, then systematically surveys the application of classical deep learning models—particularly convolutional neural networks—in DR severity grading, comparing their performance and discussing current challenges and future directions.*

Keywords: Diabetic retinopathy; Deep learning; Convolutional neural networks; Image classification; Computer-aided diagnosis.

1. INTRODUCTION

In recent years, computer-aided diagnosis (CAD) systems grounded in deep learning methodologies have achieved remarkable success across diverse medical domains, including epilepsy detection from electroencephalograms, cardiovascular disease risk stratification from electrocardiograms and cardiac imaging, and—most notably—diabetic retinopathy (DR) screening from retinal fundus photographs. DR, a leading cause of preventable blindness among working-age adults, affects approximately one-third of the 537 million adults living with diabetes worldwide, as reported by the International Diabetes Federation, and early detection is critical to halting disease progression before irreversible vision loss occurs. Deep learning excels in this context by learning hierarchical feature representations directly from large-scale annotated datasets—such as the publicly available EyePACS and APTOS 2019 Kaggle collections, which together comprise over 120,000 high-resolution fundus images—and extracting high-level abstract pathological signatures through successive nonlinear transformations, thereby enabling highly accurate DR classification and severity grading. The hallmark pathological manifestations of DR evolve across progressive stages, encompassing microaneurysms (MA), intraretinal hemorrhages (HE), cotton-wool spots (CW), hard exudates (EX), and soft exudates (SE); convolutional neural networks (CNNs), particularly variants like ResNet, Inception, and EfficientNet, have demonstrated exceptional capacity to discern these subtle lesions, often matching or surpassing human expert performance in controlled benchmarks. For instance, the widely cited DeepMind study (2016) reported a sensitivity of 96.8% and specificity of 99.1% for DR detection on the EyePACS-1 dataset, while a 2021 prospective clinical validation of the IDx-DR system—the first FDA-authorized autonomous AI diagnostic device—achieved 87.2% sensitivity and 90.7% specificity for identifying more-than-mild DR, meeting primary efficacy endpoints in a primary care setting. More recently, a 2023 multi-center study employing a hybrid ResNet-squeeze-and-excitation attention model attained an area under the receiver operating characteristic curve (AUC) of 0.985 on a validation cohort of over 50,000 images, and demonstrated robust generalization across diverse ethnicities and camera manufacturers. Beyond DR, analogous deep learning architectures have been successfully adapted to detect epileptiform discharges in electroencephalograms with over 90% accuracy and to classify coronary artery disease from cardiac MRI with F1-scores exceeding 0.88, underscoring the transferable power of CNN-based feature extraction. Despite these advances, persistent challenges remain—including class imbalance, inter-observer variability in ground-truth labeling, limited interpretability of black-box models, and domain shifts across imaging devices—which have spurred active research into explainable AI, generative data augmentation, and federated learning to preserve patient privacy while pooling multi-institutional data. Nevertheless, accumulating empirical evidence, reinforced by regulatory approvals (e.g., FDA, CE marking) and real-world deployment in national screening programs across India, Singapore, and the United Kingdom, firmly establishes deep learning-driven CAD as a transformative tool for DR management, offering scalable, cost-effective, and timely triage that can alleviate the burden on overstretched ophthalmology services and ultimately reduce global blindness rates.[1-3] Through operations such as convolution and pooling, CNN model can capture local features in images, and abstract and reason through layers of stacked structures, so as to realize classification and recognition of complex images[4-6]. CNN has become an important pillar in the field of medical diagnostics and plays a vital role.

Machine learning is a classical predictive model, which has been widely used in the risk identification and risk factor research of chronic diseases. This is a course that studies how machines adapt to their environment and learn from examples, and how to learn from them. Knowledge to solve complex problems by imitating human learning processes[7]. Inspired by the medical nervous system, machine learning algorithms have been applied and developed on computer systems, especially for big data analysis and the study of complex relationships between multiple variables. At present, there are many machine learning methods, including Bagging, Boosting, support vector machines (SVM), artificial neural network (ANN) and so on. Applying machine learning methods to prediction of diabetes-related aspects can improve clinical treatment outcomes. According to the American diabetes association (ADA), diabetes can be diagnosed based on blood glucose criteria, namely: Patients with glycated hemoglobin ($\geq 6.5\%$), fasting blood glucose (≥ 7.0 mmol /L), OGTT (2-hour value ≥ 11.1 mmol /L) and typical hyperglycemia or critical symptoms of hyperglycemia had random blood glucose ≥ 11.1 mmol /L [8]. This traditional ≥ 11.1 mmol /L diagnostic method is time-consuming and costly, limiting its practical application in low-income countries. Compared to traditional methods, machine learning methods take less time and have almost no predicted costs. The use of machine learning algorithms to build predictive models on the basis of existing data can not only provide efficient predictive methods for the medical field, but also directly intervene in the prognostic factors of disease, in order to reduce the harm of disease and reduce the burden on individuals and society.

2. DIABETES MODEL CLASSIFICATION

The predictive model can better integrate the research results of epidemiology and health statistics with clinical practice, promote the tertiary prevention of diseases through high-risk screening and clinical evidence-based guidelines, and reduce the morbidity and mortality of diseases. Depending on who is being used, diabetes-related prediction models can be divided into the following three categories:

1) Diabetes risk assessment model: It is a kind of diabetes risk prediction for healthy people (non-diabetic patients), mainly used to identify high-risk groups of diabetes, and intervention on the risk factors of diabetes in these groups, so as to eliminate the disease in the bud, avoid the occurrence of diabetes, belongs to the category of primary prevention. This kind of model is simple to operate, low cost, can detect high-risk groups of diabetes as early as possible, formulate prevention and control measures in time, effectively reduce the incidence of diabetes or slow down the occurrence of diabetes, and reduce the health burden to a certain extent, which has great significance in public health.

Table 1: Diabetes prediction model

DKD	stratification	value
AGE	39-49	0
	50-59	3.0
	60-75	6.0
BMI(kg/m ²)	<24.00	0
	24.00-27.99	1.5
	≥ 28.00	3.0
Smoking history	no	0
	yes	4.0
diabetes.	0	0
	1	3.0
HBA1c(%)	<7.0	0
	7.0-7.9	1.5
	8.0-8.9	3.0
	≥ 9.0	4.5

2) Diabetes Screening model: The model is a tool to screen community populations for diabetes, guide community diabetes screening, and then conduct in-depth diabetes testing for patients with diabetes detected by the model. The purpose of this model is to identify patients from the community population, so as to achieve the purpose of early detection and early treatment, which belongs to the category of secondary prevention.

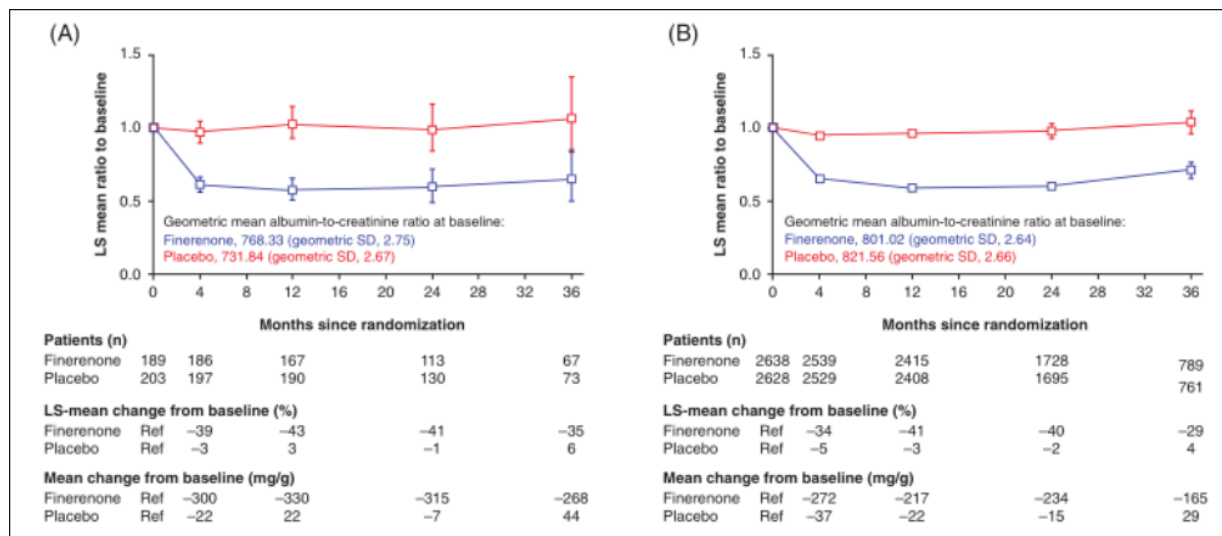


Figure 1: Diabetes screening model

3) Diabetes complication prediction model: also known as diabetes clinical model, mainly used to predict the possibility of diabetes patients in the next few years or even a longer period of time to develop a certain or several complications, and even predict the order of complications, which belongs to the category of secondary prevention and tertiary prevention. This study was carried out earlier in foreign countries, and some related models have been programmed to form predictive model software, which can be applied to mobile terminals. The most famous one is the UK Prospective diabetes Study Model (UK prospective diabetes study model, UKPDS model and Archimedes Model).

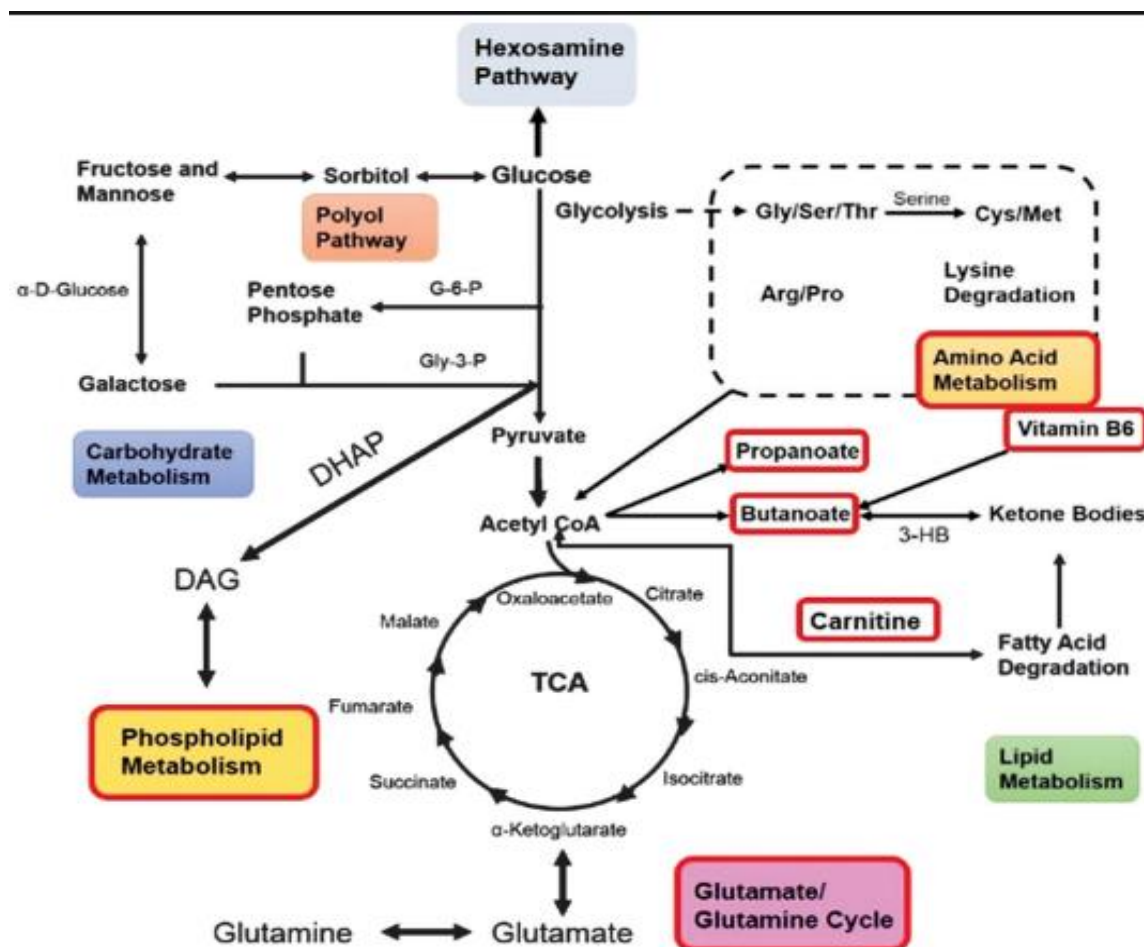


Figure 2: Diabetic complication model

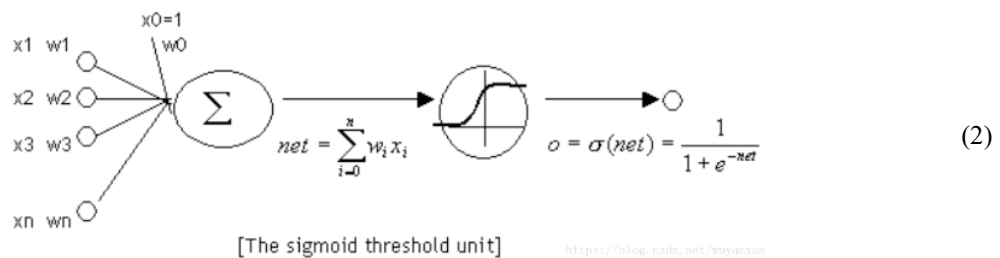
3. MACHINE LEARNING

Machine learning was first applied to the field of artificial intelligence, which belongs to the core of artificial intelligence and is a method to achieve artificial intelligence. Machine learning is a method of "training" by "eating a lot of data" and learning to complete tasks through various algorithmic rules. Traditional algorithms of machine learning include decision tree, clustering, Bayesian classification, support vector machine, Adaboost and so on. In terms of learning methods, machine learning algorithms can be divided into supervised learning (such as discriminant classification problems), unsupervised learning (such as clustering problems), semi-supervised learning, reinforcement learning, ensemble learning and deep learning.

1) There is supervised learning, under which a set of examples or training modules are provided with the correct output, and based on these training sets, the algorithm learns how to respond more accurately by comparing its output with the output as input. Supervised learning is also known as learning by example or from example. SUPERVISOR. The learning task can be divided into classification task and regression task. In classification task, the output is discrete, while in regression task, the output is continuous. Common examples of supervised learning include naive BayesNB, decision tree (DT), k-nearest neighbor, and so on. KNN), artificial neural network (ANN), support vector machine (SVM), randomforest (RF), logistic regression (logistic regression, LR), etc.

$$P(Y|X) = \frac{P(X|Y)P(Y)}{P(X)} \quad (1)$$

The formula is understood as $P(X, Y) = P(Y, X) \Leftrightarrow P(XY)P(Y) = P(YIX)P(X)$, that is, the probability that X and Y occur at the same time is the same as the probability that Y and X occur at the same time.



The input and output of a typical M-P neuron model are the same as Logistic regression, but here Sigmoid exists as an activation function. In other words, Sigmoid represents the output of only one neuron, not the output of the entire ANN. A diagram graphically represents the MP neuron.

2) Reinforcement learning, from never understanding to constantly learning, summarizing the rules, and eventually learning. Process is reinforcement learning. Reinforcement learning is considered an intermediate type of learning where the algorithm provides only one response that tells whether the output is correct. The algorithm must explore and rule out various possibilities in order to get the correct output. Reinforcement learning is thought of as learning with critics, and its algorithms do not make any suggestions or solutions to problems. Reinforcement learning algorithms fall into two main categories: value-based algorithms and policy-based algorithms. There are two common application scenarios for reinforcement learning: dynamic systems and robot control.

3) deep learning (DL) is a kind of machine learning technology that uses multi-layer nonlinear information processing for supervised or unsupervised feature extraction and transformation, as well as pattern analysis and classification [17]. Deep learning is mainly implemented using artificial neural networks (ANN), a mathematical model inspired by the way biological nervous systems work. Artificial neural network is composed of artificial neurons and their connections. The working logic of artificial neurons can be defined using a variety of linear and nonlinear mathematical functions

Action parameters can be optimized by learning from training data. Common deep neural networks (DNN) include multi-layer perceptron (MLP), convolutional neural network, CNN), recurrent neural network (RNN), deepautoencoder (DAE).

4. METHODOLOGY

The field of detection of diabetic retinopathy involves the study of fundus images. To objectively evaluate methods and verify the feasibility of algorithms, researchers can use a variety of open source datasets.

4.1 Evaluation Metrics

The commonly used indicators for diagnosing diabetic retinopathy include Accuracy, sensitivity, Specificity, and AUC. Accuracy is the ratio between the number of correctly classified samples and the total number of samples, and it is the most common classification evaluation index. Sensitivity represents the proportion of the total number of correctly classified positive cases to actual positive results. Specificity represents the proportion of the total number of correctly classified negative cases to the actual negative results. Combining sensitivity and specificity can provide comprehensive information about the classifier's performance across different classes, helping to select the appropriate model and threshold. The AUC represents the area under the ROC curve (which plots the relationship between sensitivity and specificity), and the value is between 0 and 1. The closer the AUC value is to 1, the better the performance of the classifier.

4.2 Deep Learning Method of DR Binary Classification

The research goal of the deep learning method of DR Binary classification is to classify the input retinal image by binary, that is, normal or DR. The characteristic indicators of DR include neovascularization, bleeding, exudate, etc. The following is the comparison image of normal retina and DR Proposed in the literature. This section introduces the latest progress in the application of deep learning in DR Binary classification.

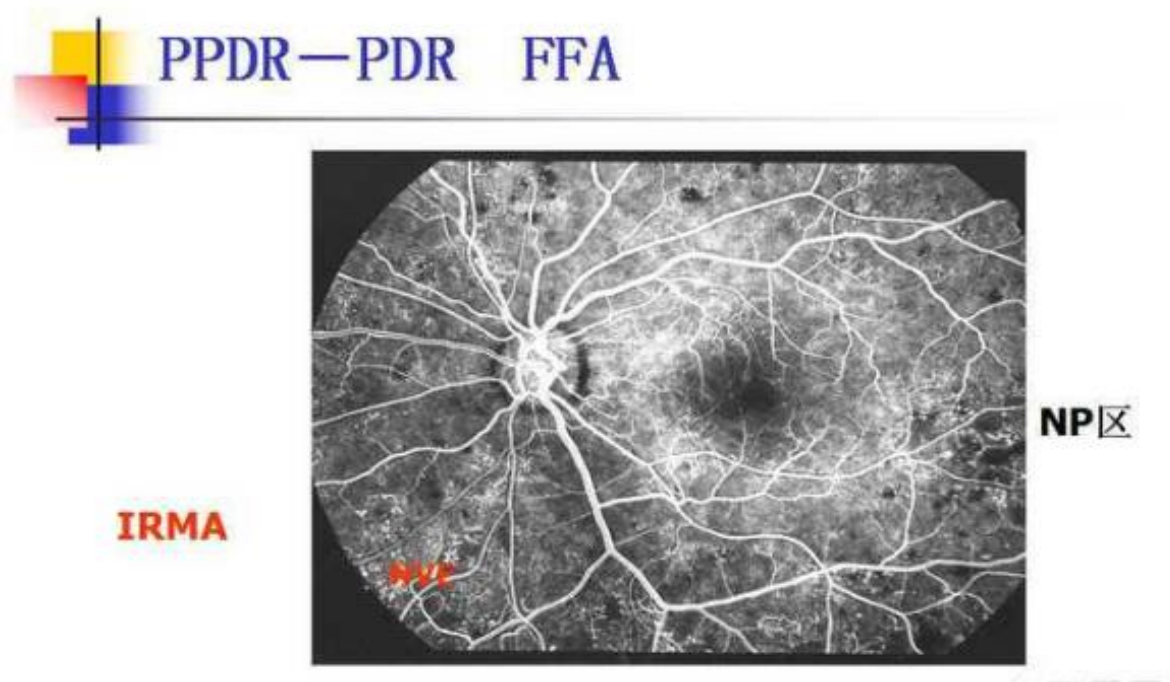


Figure 3: Comparison of normal retinal images and diabetic retinopathy images

Deep learning originated from Artificial Neural Networks (ANNs). Among the classical ANN algorithms, the most famous and frequently used architecture is the feedforward Multilayer Perceptron (MLP). Its basic structure includes input layer, output layer and hidden layer. The paper proposes the application of multi-layer perceptron network (MLP) to classify DR, and the accuracy rate is 94.11%, which proves the feasibility of ANN research in this aspect. On this basis, Harun et al. [23] used Levenberg Marquardt learning algorithm combined with MLP, model for data classification, extracted 19 features from fundus images and classified them as inputs to the neural network. The main advantage of this algorithm is that it can quickly update model parameters combined with MLP.

5. CONCLUSION

We found that accuracy was reported in all of the models mentioned above, and some other performance indicators were selected for reporting based on research needs. Due to the heterogeneity of the reported indicators, performance comparison between models is challenging. It is hoped that future researchers can calculate at least three parameters (sensitivity, specificity, and AUC) from the confusion matrix and AUC(ROC)[9-11]. Che et al. significantly contribute to the field of biomedical text document classification, particularly in Cancer Doc Classification. It takes on the substantial challenge of classifying extensive research papers on cancer, a departure from previous works that mainly focus on shorter abstracts. The wide adoption of Random Forest in machine learning and data science further emphasizes its effectiveness in handling intricate classification tasks, making it a valuable tool in addressing complex challenges in the classification of biomedical texts related to medical engineering[12]. Ideally, five (accuracy, sensitivity, specificity, accuracy, and F1 score) or more parameters (AUC, etc.) should be reported for easy comparison. If one of them is missing, it can be estimated from the other parameters. At present, although the prediction method based on machine learning has good prediction effect in terms of disease occurrence and prognosis, it also has shortcomings such as large amount of data, continuous data within a certain period of time, weak generalization ability, overfitting into local minimum, insensitive to random and unstable data, and unsatisfactory prediction effect of unbalanced data[12-13]. Unbalanced data classification is an extremely important research area, which needs to be paid attention to. When the influence of unbalanced data is fully considered in modeling, a model with high predictive performance can be obtained, so as to provide scientific methods and basis for the prevention and control of diabetes in our population. Future medical research on methodology will pay more attention to the above shortcomings, so as to establish better machine learning models, which will play a greater role in the field of population chronic disease prevention[14].

To sum up, it can be seen that deep learning technology has made remarkable achievements in the field of DR Diagnosis and early prediction in recent years. Although the improved cuff network model has been continuously improved in various performance indicators, it still faces many challenges[15-20].

- (1) DR Cases are scarce and the cost of doctor labeling is high, resulting in insufficient data available for training.
- (2) There is a serious data imbalance between different categories of diabetic retinopathy.
- (3) The black box form of deep learning models leads to poor interpretability of models.
- (4) The model has the problem of insufficient generalization ability and overfitting.
- (5) Retinopathy is a small sample target, and there is a problem of similarity between the categories of lesions, so it is difficult to accurately distinguish the model.
- (6) The deployment of the model to the real clinical environment involves compliance, security, privacy protection and other issues, and it is difficult to be practically applied to the clinical environment.

REFERENCES

- [1] Wang, Y., Shen, Z., Hu, K., Yang, J., & Li, C. (2025). AI End-to-End Autonomous Driving.
- [2] Chen, J., Zhang, X., Wu, Y., Ghosh, S., Natarajan, P., Chang, S. F., & Allebach, J. (2022). One-stage object referring with gaze estimation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 5021-5030).
- [3] Xie, Y., Li, Z., Yin, Y., Wei, Z., Xu, G., & Luo, Y. (2024). Advancing Legal Citation Text Classification A Conv1D-Based Approach for Multi-Class Classification. *Journal of Theory and Practice of Engineering Science*, 4(02), 15–22. [https://doi.org/10.53469/jtpes.2024.04\(02\).03](https://doi.org/10.53469/jtpes.2024.04(02).03)
- [4] Xu, G., Xie, Y., Luo, Y., Yin, Y., Li, Z., & Wei, Z. (2024). Advancing Automated Surveillance: Real-Time Detection of Crown-of-Thorns Starfish via YOLOv5 Deep Learning. *Journal of Theory and Practice of Engineering Science*, 4(06), 1–10. [https://doi.org/10.53469/jtpes.2024.04\(06\).01](https://doi.org/10.53469/jtpes.2024.04(06).01)
- [5] Yin, Y., Xu, G., Xie, Y., Luo, Y., Wei, Z., & Li, Z. (2024). Utilizing Deep Learning for Crystal System Classification in Lithium - Ion Batteries. *Journal of Theory and Practice of Engineering Science*, 4(03), 199–206. [https://doi.org/10.53469/jtpes.2024.04\(03\).19](https://doi.org/10.53469/jtpes.2024.04(03).19)
- [6] Luo, Y., Wei, Z., Xu, G., Li, Z., Xie, Y., & Yin, Y. (2024). Enhancing E-commerce Chatbots with Falcon-7B and 16-bit Full Quantization. *Journal of Theory and Practice of Engineering Science*, 4(02), 52–57. [https://doi.org/10.53469/jtpes.2024.04\(02\).08](https://doi.org/10.53469/jtpes.2024.04(02).08)

- [7] Yan, H., Wang, Z., Xu, Z., Wang, Z., Wu, Z., & Lyu, R. (2024, July). Research on image super-resolution reconstruction mechanism based on convolutional neural network. In Proceedings of the 2024 4th International Conference on Artificial Intelligence, Automation and High Performance Computing (pp. 142-146).
- [8] Pang, H., Zhou, L., Dong, Y., Chen, P., Gu, D., Lyu, T., & Zhang, H. (2024). Electronic Health Records-Based Data-Driven Diabetes Knowledge Unveiling and Risk Prognosis. arXiv preprint arXiv:2412.03961.
- [9] Long, Y., Gu, D., Li, X., Lu, P., & Cao, J. (2024, September). Enhancing Educational Content Matching Using Transformer Models and InfoNCE Loss. In 2024 IEEE 7th International Conference on Information Systems and Computer Aided Education (ICISCAE) (pp. 11-15). IEEE.
- [10] Tang, Y., & Zhao, S. (2025). Research on Relationship Between Aging Population Distribution and Real Estate Market Dynamics based on Neural Networks.
- [11] Zhao, S., Lu, Y., Gong, C., & Xu, Q. (2025). Research on Labour Market Efficiency Evaluation Under Impact of Media News Based on Machine Learning and DMP Model.
- [12] Chen, K., Zhao, S., Jiang, G., He, Y., & Li, H. (2025). The Green Innovation Effect of the Digital Economy. International Review of Economics & Finance, 103970.
- [13] Yao, T. (2024, August). Research on the Local Head Loss Coefficient in Short-Tube Hydraulic Testing. In 2024 3rd International Conference on Applied Mechanics and Engineering Structures (AMES 2024) (pp. 89-97). Atlantis Press.
- [14] Xiangyu, G., Yao, T., Gao, F., Chen, Y., Jian, X., & Ma, H. (2024). A new granule extrusion-based for 3D printing of POE: studying the effect of printing parameters on mechanical properties with “response surface methodology”. Iranian Polymer Journal, 1-12.
- [15] Wu, W. (2025). Fault Detection and Prediction in Models: Optimizing Resource Usage in Cloud Infrastructure.
- [16] Ge, J. (2024). Technologies in Peace and Conflict: Unraveling the Politics of Deployment. Review International Journal of Research Publication and Reviews (IJRPR), 5(5), 5966-5971.