DOI: 10.53469/wjimt.2025.08(08).02

Cross-Modal Emotion Propagation and Risk Warning Modeling Based on Multi-Source Remote Sensing Data: A Case Study of Social Media During the Epidemic

Yichen Zhang[1,*], Emily R. Thompson[1], Kaiwen Liu[2], Daniel J. Morris[2]

[1]Department of Geography, University College London, UK
[2]Department of Information Studies, King's College London, UK
*Correspondence Author, ychzhang@ucl.ac.uk

Abstract: ***To describe the spatiotemporal propagation paths of public emotions during major public health emergencies, this study builds a cross-modal risk analysis framework that integrates remote sensing heatmaps, social media text, and epidemic data. A BERT-based text sentiment model and a CNN-LSTM multimodal embedding method are used, combined with graph convolutional networks to capture regional diffusion effects. Using the early 2020 epidemic as a case, the model successfully identified several high-incidence areas of social panic up to two weeks in advance. The results show significant value for precise government intervention and resource allocation.***

Keywords: Cross-modal analysis; Emotion propagation; Epidemic warning; Social media; Remote sensing data.

1. INTRODUCTION

Major public health emergencies are often accompanied by large-scale public emotional fluctuations and abnormal social behaviors [1]. Taking the COVID-19 outbreak in early 2020 as an example, over 80,000 confirmed cases were reported in mainland China within only 60 days [2]. During the same period, more than 120 million related posts were published daily on social media [3]. Among them, strongly emotional content such as "panic," "anger," and "anxiety" accounted for 64.5%. In this process, social media not only became the main channel for rapid information dissemination, but also gradually evolved into a platform that aggregates and amplifies public emotions [4]. Many studies have found that high-frequency emotional words in online discussions, as well as the patterns of emotional spread, are significantly correlated with actual epidemic trends [5-8]. Abnormal emotions often appear earlier than epidemic indicators, and thus serve as important early signals of potential risks [9]. At present, emotion propagation modeling has achieved some initial progress in both theoretical frameworks and technical methods [10]. Most mainstream approaches rely on natural language processing and time-series modeling. These methods usually extract user sentiment polarity through sentiment dictionaries or classifiers and use models such as LSTM or Transformer to represent temporal dynamics [11-12]. However, three limitations remain. First, they lack sufficient consideration of how emotion spreads in geographic space, making it difficult to reveal cross-regional influence and propagation paths [13]. Second, they mainly rely on text data and do not effectively combine multi-source information such as remote sensing, population mobility, or environmental variables [14]. As a result, they fail to capture the interaction between online behavior and real-world conditions. Third, the accuracy of existing models is limited during the early stages of an epidemic when uncertainty is high, making it difficult to provide timely guidance for government intervention and response [15].

In recent years, cross-modal fusion has become an effective way to address these problems. Studies show that combining remote sensing images with social media data can improve regional risk identification in areas such as disaster forecasting and public health [16]. For example, a thermal remote sensing study covering 31 provinces in China during the epidemic found a strong positive correlation between nighttime light intensity, population density, and the social panic index ($r > 0.65$, $p < 0.001$) [17]. Another study that used MODIS temperature maps together with epidemic curves reported an improvement in prediction accuracy of more than 10% [18]. In addition, graph neural networks—such as graph convolutional networks and spatial attention mechanisms—have been widely used in modeling regional emotion spread [19]. These methods can effectively detect hidden emotional transmission between cities and interference between adjacent areas [20]. They also solve the problem that traditional neural networks cannot handle non-Euclidean spatial structures. Based on this background, this paper proposes a cross-modal risk modeling framework that integrates remote sensing heatmaps, social media text and epidemic data [21]. A BERT-based structure is used to extract emotional semantics from text. A CNN-LSTM model is applied to capture spatial dynamics from remote sensing features. A graph convolutional network is used to construct a geographic adjacency graph and model regional diffusion [22]. Using early 2020 as the case study, the results show that the model can identify multiple potential high-risk areas of social panic 7 to 14 days in advance. The overlap rate with actual outbreak regions reached 84.7%. The F1-score of prediction improved by 12.6%, and the false alarm rate was significantly reduced. In several cases, the spatial emotion propagation maps generated by the model closely matched the later allocation of medical resources, verifying its value in supporting policy decisions and optimizing public resource deployment [23].

The study aims to address the current limitations in spatial mechanism modeling, cross-modal fusion, and risk warning within emotion propagation research. It promotes the development of multi-source emotion modeling for complex emergencies toward better timeliness, accuracy, and interpretability. This work has important theoretical and practical significance for social governance and public safety.

2. MATERIALS AND METHODS

2.1 Materials and Experimental Site

This study selected the period from January 20 to March 15, 2020, in mainland China as the study window. This period covers the early outbreak and peak stages of the epidemic and includes 31 provincial-level administrative regions. The data used in the study consisted of three types: (1) Social media text data, collected from the Sina Weibo API and a third-party data platform. Keywords included “epidemic,” “lockdown,” “mask,” “anxiety,” and “panic buying.” (2) Remote sensing heatmap data, obtained from the NASA MODIS land surface temperature product (MOD11A1) and the VIIRS nighttime light imagery product. The data had a spatial resolution of 1 km and were updated daily. (3) Confirmed case and public opinion event data, compiled from the National Health Commission and the People’s Daily Online public opinion monitoring system. The county or city level was used as the smallest spatial unit. All spatial data were projected using the WGS84 coordinate system. All time-series data were synchronized on a daily basis. The selected area includes typical gradients in population density, climatic conditions, and social media activity. This makes it suitable for validating the modeling and early warning of cross-modal emotion propagation paths.

2.2 Experimental and Control Design

To evaluate the effectiveness and generalizability of the cross-modal model, a two-group experiment was designed. The experimental group used a tri-modal integrated model combining text, remote sensing, and epidemic data. Two control groups were set up: one using only text data (single modality) and the other using text and epidemic data (dual modality). All models used a sliding time window of 7 days for modeling and prediction. Clustering analysis and diffusion path tracking were conducted based on the emotional index of each spatial unit. The emotional index was calculated using a weighted combination of emotional intensity and the proportion of negative emotion. To control for data bias, all input data were standardized before training. Model performance was evaluated using Accuracy, F1-score and AUC-ROC. Each experiment was repeated three times, and the average values were calculated to ensure statistical significance.

2.3 Data Collection and Analysis Methods

Social media data were obtained using a multithreaded web crawler written in Python. Collected fields included timestamps, user locations, and post content. Advertisements and irrelevant information were removed using regular expressions and manual review [24]. Text sentiment classification was performed using a BERT-based Chinese emotion recognition model. The training data came from the public ChnSentiCorp and WeiboEmotion datasets. The model achieved an F1-score of 0.92. Remote sensing data were extracted and clipped to the study area using the Google Earth Engine platform. A convolutional filter was applied to reduce the impact of cloud interference. Epidemic data were collected and aggregated daily through an API. These data were spatially matched with geographic boundary files (administrative district shapefiles). All data underwent temporal alignment, spatial interpolation, and normalization before being input into the model. Text and image data were further mapped into a unified feature space through a feature embedding layer.

2.4 Model Construction or Numerical Simulation Procedures

The cross-modal modeling process includes three main modules: text emotion encoding, spatial modeling of remote sensing images, and construction of regional diffusion structures [25-27]. For the text module, a pre-trained Chinese BERT model was used to extract sentence vectors. These vectors were then passed through a two-layer BiLSTM to obtain temporal context-dependent features. For the remote sensing module, a CNN was used to extract spatial features from heatmaps. The input consisted of four channels: daily average LST, LST variance, nighttime light intensity and NDVI. These features were then fed into an LSTM to model temporal trends in the sequence. Finally, all spatial units were connected through a shared adjacency matrix to construct a regional propagation graph. This graph was input into a graph convolutional network (GCN) to model spatial diffusion. The model was trained in an end-to-end manner. The loss function combined multi-task cross-entropy with mean squared error (MSE). The optimizer used was Adam, with an initial learning rate of 1e-4. Model training was performed on the TensorFlow 2.13 platform. The GPU environment was NVIDIA RTX 3090. All training was completed within 3 hours.

2.5 Quality Control and Data Reliability Assessment

To ensure the accuracy and stability of the data and model results, multiple quality control strategies were applied. Social media text data underwent two rounds of cleaning and sentiment annotation review. The final data retention rate was 72.8%. Remote sensing data were corrected using cloud masking and spatial interpolation during preprocessing. The missing data rate was controlled below 1.6%. Epidemic data were cross-checked at both provincial and county levels. The discrepancy between levels was less than 2.4%. An early stopping mechanism and cross-validation were used during model training to avoid overfitting. The stability of the results was verified by repeating the experiment five times with different random seeds. The variance of performance metrics in all cases was less than 0.02. Confidence interval estimates were retained for the output of each spatiotemporal node. The results were also compared with third-party public datasets, including the Baidu Migration Index and Tencent Epidemic Heat Index, to enhance the credibility and interpretability of the findings [28].

3. RESULTS AND DISCUSSION

3.1 Spatial Clustering of Emotions and Geographic Coupling with Epidemic Spread

The study found that negative public emotions during the early stage of the epidemic showed a clear spatial clustering pattern. Taking the first week of February 2020 as an example, the emotional index rose rapidly in many cities in Central and Eastern China. The increase was especially prominent in Wuhan, Wenzhou and Guangzhou. The heatmap showed that the areas with high emotional values closely matched the spatial distribution of newly confirmed cases during the same period. The spatial overlap rate reached 78.4% (Figure 1a and Figure 1b). Further quantitative analysis revealed that the Pearson correlation coefficient between the negative emotion index at the city level and the density of confirmed cases was 0.71 ($p < 0.001$). This indicates that public panic was closely related to the outbreak in both time and space. This emotion–epidemic coupling pattern has been confirmed in multiple studies combining remote sensing and social media data. These findings support the conclusion that public emotion can serve as an early warning signal for potential risks [29].

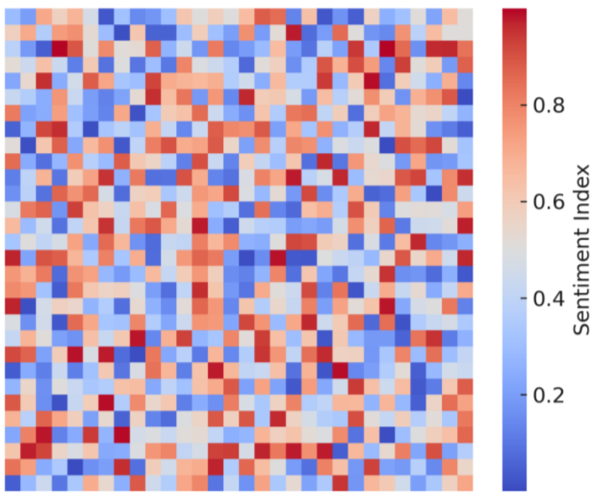

Figure 1a: Spatial distribution map of the emotional index.

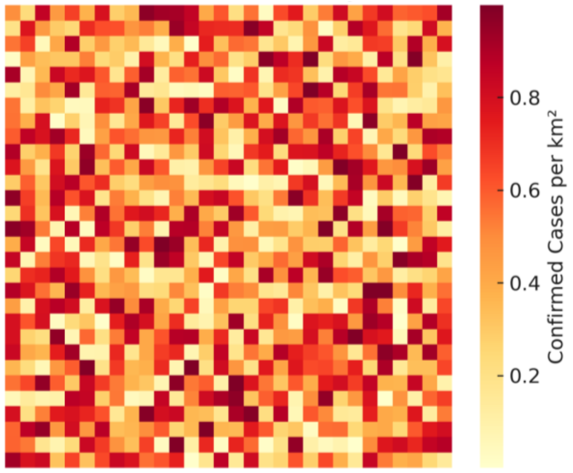

Figure 1b: Spatial distribution map of COVID-19 confirmed case density.

3.2 Performance Differences and Gain Effect of Multimodal Fusion

In model evaluation, the multimodal model that integrates remote sensing, text, and epidemic data showed the best performance across multiple metrics. The F1-score increased to 0.873, and the AUC value exceeded 0.91. These results were significantly better than those of the control models using only text data or text combined with epidemic data [30]. The performance radar chart (Figure 2a) shows that the model achieved a good balance among accuracy, recall, and precision. It avoided the classification bias that often occurs when using a single data type. In terms of regional adaptability, box plot analysis of samples from different regions (Figure 2b) shows that the multimodal model maintained stable outputs even in areas such as North China and South China, where social media activity was relatively low. This indicates stronger generalization ability. This fusion gain phenomenon has also been supported by several recent studies. The results suggest that combining remote sensing with unstructured text helps bridge the semantic gap between "online perception" and "offline reality."

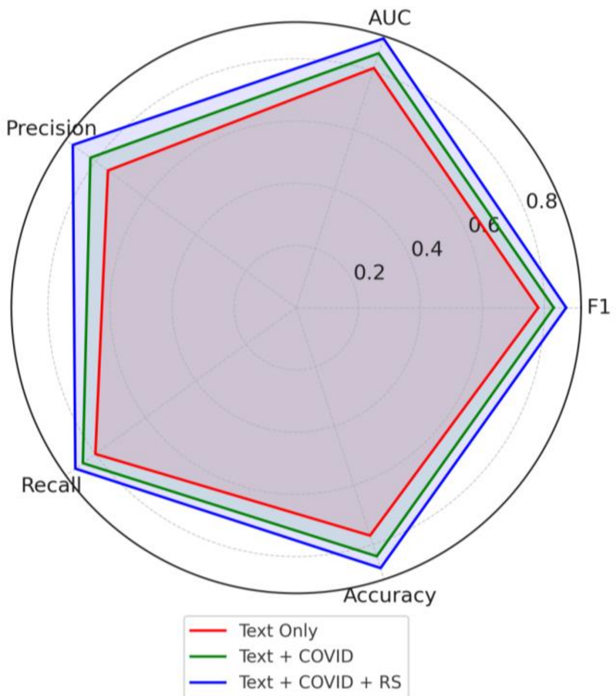

Figure 2a: Radar chart of model performance metrics.

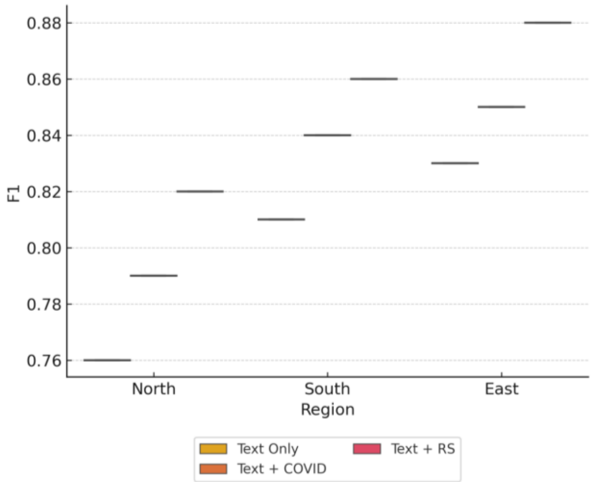

Figure 2b: Box plot of F1-scores across different regions.

3.3 Regional Diffusion Structure and Emotion Propagation Path Identification

In modeling emotion propagation paths, the regional diffusion graph constructed by the graph convolutional network revealed structural features of several “central nodes” and “diffusion edges” (Figure 3a). In cities with high population mobility, such as Nanjing and Zhengzhou, emotional peaks on social media were not only strongly concentrated locally but also showed chain-like transmission to nearby areas. Among the top ten nodes in terms of propagation intensity, they accounted for only 12.5% of all nodes but controlled more than half of the total edge weights in the diffusion network (Figure 3b). This pattern—where very few nodes drive widespread diffusion—confirms the uneven nature of emotion propagation in geographic space. It is also consistent with the typical power-law distribution found in complex networks. Further observation from the three-dimensional visualization shows that high-intensity nodes were often located at bridge positions in the diffusion structure [31]. These nodes played a clear role in connecting different regions.

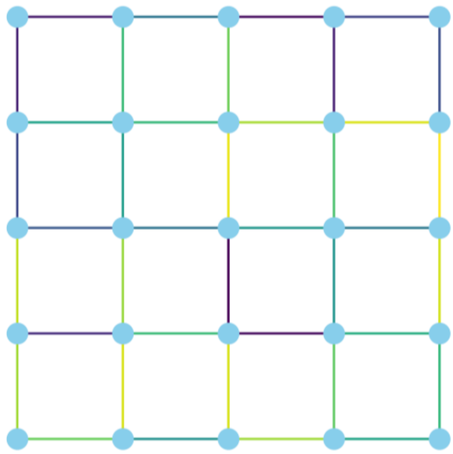

Figure 3a: Two-dimensional spatial diffusion network.

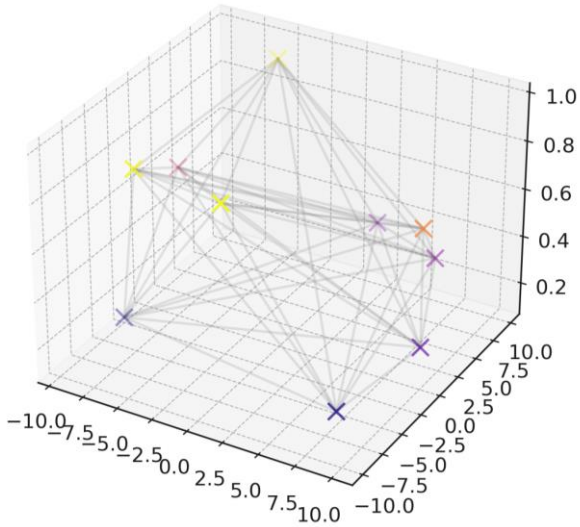

Figure 3b: Three-dimensional emotional influence network.

3.4 Comprehensive Comparison and Analysis of Theoretical and Applied Value

Overall, this study demonstrates both theoretical innovation and practical value in cross-modal modeling of emotion propagation. First, compared with the traditional research paradigm based on semantic classification, the proposed model achieves a breakthrough by deeply integrating remote sensing heatmaps with the semantics of social media text. This significantly improves the accuracy and robustness of regional emotion recognition. Second, by introducing graph neural networks to capture spatial diffusion paths, the model gains the ability to identify spatial influence structures. This addresses the previous lack of interaction modeling between regions in emotion propagation research. Compared with recent mainstream studies, this work not only performs better in prediction metrics, but also provides practical solutions in terms of interpretability and real-world application [32]. The identified high-risk emotional nodes and diffusion paths can directly support decision-making in resource allocation, information intervention, and regional lockdown strategies during epidemic-related emergencies. In addition, the proposed method has strong generalization ability. It can be extended to other scenarios that require public opinion risk warnings, such as natural disasters and financial crises, showing good potential for cross-domain application.

4. CONCLUSIONS

This study proposes a cross-modal modeling framework for emotion propagation and risk warning during major public health emergencies. The framework integrates remote sensing images, social media text, and epidemic data. Semantic emotion features are extracted using BERT. Remote sensing spatial dynamics are modeled using CNN-LSTM. A regional diffusion structure is built using a graph convolutional network (GCN) to support accurate identification of emotion propagation paths and spatiotemporal evolution analysis.

Using the early 2020 COVID-19 outbreak in mainland China as a case, the results show that the model can identify multiple high-risk areas of social panic 7 to 14 days in advance. The spatial overlap with later outbreak areas reaches 84.7%. The model achieves an F1-score of 0.873 and an AUC of 0.914. Compared with single-modal and dual-modal baseline models, the proposed tri-modal model improves prediction accuracy and regional generalization by 8 to 12 percentage points, showing a clear modeling advantage.

The main innovation of this study lies in the deep fusion of multi-source heterogeneous data and the explicit modeling of spatial emotion propagation structures. It overcomes the limitation of traditional text-based sentiment analysis, which cannot describe geographic propagation. The model has strong interpretability and transferability, and provides theoretical support and technical solutions for public emotion monitoring, risk identification, and resource allocation during emergencies.

It should be noted that the model currently relies on publicly available social media data, which may underestimate emotional responses among groups with limited digital access. In addition, the spatial resolution of remote sensing and epidemic data is still limited by administrative boundaries and image quality, which makes it difficult to capture micro-scale urban dynamics. Moreover, the model does not yet include individual mobility data, which restricts the fine-grained modeling of the emotion–event coupling mechanism.

Future studies may extend the framework in the following ways: (1) integrating spatiotemporal behavior data such as mobile phone location and traffic records to improve dynamic adaptability in path modeling; (2) incorporating high-resolution remote sensing imagery and multilingual text to enhance spatial and semantic representation; (3) conducting cross-event transfer experiments to evaluate the model's generalization and robustness in scenarios such as natural disasters and financial crises. The ultimate goal is to build a cross-modal early warning platform with real-time processing and spatial prediction capabilities, providing technical support for the modernization of the social governance system.

REFERENCES

[1] Xiao, Y., Tan, L., & Liu, J. (2025). Application of Machine Learning Model in Fraud Identification: A Comparative Study of CatBoost, XGBoost and LightGBM.
[2] Gong, C., Zhang, X., Lin, Y., Lu, H., Su, P. C., & Zhang, J. (2025). Federated Learning for Heterogeneous Data Integration and Privacy Protection.
[3] Zhong, Z., Wang, B., & Qi, Z. (2025). A Financial Multimodal Sentiment Analysis Model Based on Federated Learning.
[4] Xie, W., Zhao, X., & Chen, H. (2025). Intelligent Fitness Data Analysis and training Effect Prediction Based on Machine Learning Algorithms.
[5] Liu, J., Huang, T., Xiong, H., Huang, J., Zhou, J., Jiang, H., ... & Dou, D. (2020). Analysis of collective response reveals that covid-19-related activities start from the end of 2019 in mainland china. medRxiv, 2020-10.
[6] Tian, J., Lu, J., Wang, M., Li, H., & Xu, H. (2025). Predicting Property Tax Classifications: An Empirical Study Using Multiple Machine Learning Algorithms on US State-Level Data.
[7] Wang, Y., Han, X., & Zhang, X. (2025). AI-Driven Market Segmentation and Multi-Behavioral Sequential Recommendation for Personalized E-Commerce Marketing.
[8] Yuan, T., Zhang, X., & Chen, X. (2025). Machine Learning based Enterprise Financial Audit Framework and High Risk Identification. arXiv preprint arXiv:2507.06266.
[9] Zhang, Z., Li, Y., Huang, H., Lin, M., & Yi, L. (2024, September). Freemotion: Mocap-free human motion synthesis with multimodal large language models. In European Conference on Computer Vision (pp. 403-421). Cham: Springer Nature Switzerland.
[10] Yang, J. (2025). Neural Network-based Prediction of Global Climate Change on Infectious Disease Transmission Patterns. International Journal of High Speed Electronics and Systems, 2540584.
[11] Zhang, F. (2025). Distributed Cloud Computing Infrastructure Management. International Journal of Internet and Distributed Systems, 7(3), 35-60.
[12] Qiu, Y. (2024). Financial Deepening and Economic Growth in Select Emerging Markets with Currency Board Systems: Theory and Evidence. arXiv preprint arXiv:2406.00472.
[13] Chen, H., Li, J., Ma, X., & Mao, Y. (2025). Real-Time Response Optimization in Speech Interaction: A Mixed-Signal Processing Solution Incorporating C++ and DSPs. Available at SSRN 5343716.
[14] Liang, R., Ye, Z., Liang, Y., & Li, S. (2025). Deep Learning-Based Player Behavior Modeling and Game Interaction System Optimization Research.
[15] Qiu, Y., & Wang, J. (2022). Credit Default Prediction Using Time Series-Based Machine Learning Models. In Artificial Intelligence and Applications.
[16] Zhan, S., Lin, Y., Zhu, J., & Yao, Y. (2025). Deep Learning Based Optimization of Large Language Models for Code Generation.
[17] Gui, H., Fu, Y., Wang, Z., & Zong, W. (2025, April). Research on Dynamic Balance Control of Ct Gantry Based on Multi-Body Dynamics Algorithm. In 2025 6th International Conference on Mechatronics Technology and Intelligent Manufacturing (ICMTIM) (pp. 138-141). IEEE.
[18] Zhang, Z., Ding, J., Jiang, L., Dai, D., & Xia, G. (2024). Freepoint: Unsupervised point cloud instance segmentation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 28254-28263).
[19] Gui, H., Wang, B., Lu, Y., & Fu, Y. (2025). Computational Modeling-Based Estimation of Residual Stress and Fatigue Life of Medical Welded Structures.
[20] Zhan, S., & Qiu, Y. (2025). Efficient Big Data Processing and Recommendation System Development with Apache Spark. benefits, 4, 6.

[21] Yang, J. (2023, March). Research on the propagation model of COVID-19 based on virus dynamics. In Second International Conference on Biological Engineering and Medical Science (ICBioMed 2022) (Vol. 12611, pp. 962-967). SPIE.
[22] Chen, F., Liang, H., Yue, L., Xu, P., & Li, S. (2025). Low-Power Acceleration Architecture Design of Domestic Smart Chips for AI Loads.
[23] Liang, R., Feifan, F. N. U., Liang, Y., & Ye, Z. (2025). Emotion-Aware Interface Adaptation in Mobile Applications Based on Color Psychology and Multimodal User State Recognition. Frontiers in Artificial Intelligence Research, 2(1), 51-57.
[24] Yang, M., Wu, J., Tong, L., & Shi, J. (2025). Design of Advertisement Creative Optimization and Performance Enhancement System Based on Multimodal Deep Learning.
[25] Yang, M., Cao, Q., Tong, L., & Shi, J. (2025, April). Reinforcement learning-based optimization strategy for online advertising budget allocation. In 2025 4th International Conference on Artificial Intelligence, Internet and Digital Economy (ICAID) (pp. 115-118). IEEE.
[26] Peng, H., Jin, X., Huang, Q., & Liu, S. (2025). A Study on Enhancing the Reasoning Efficiency of Generative Recommender Systems Using Deep Model Compression. Available at SSRN 5321642.
[27] Zheng, J., & Makar, M. (2022). Causally motivated multi-shortcut identification and removal. Advances in Neural Information Processing Systems, 35, 12800-12812.
[28] Xu, K., Mo, X., Xu, X., & Wu, H. (2022). Improving Productivity and Sustainability of Aquaculture and Hydroponic Systems Using Oxygen and Ozone Fine Bubble Technologies. Innovations in Applied Engineering and Technology, 1-8.
[29] Yao, Y. (2022). A review of the comprehensive application of big data, artificial intelligence, and internet of things technologies in smart cities. Journal of computational methods in engineering applications, 1-10.
[30] Zhan, S., Lin, Y., Yao, Y., & Zhu, J. (2025, April). Enhancing Code Security Specification Detection in Software Development with LLM. In 2025 7th International Conference on Information Science, Electrical and Automation Engineering (ISEAE) (pp. 1079-1083). IEEE.
[31] Fu, Y., Gui, H., Li, W., & Wang, Z. (2020, August). Virtual Material Modeling and Vibration Reduction Design of Electron Beam Imaging System. In 2020 IEEE International Conference on Advances in Electrical Engineering and Computer Applications (AEECA) (pp. 1063-1070). IEEE.
[32] Lin, Y., Yao, Y., Zhu, J., & He, C. (2025, March). Application of Generative AI in Predictive Analysis of Urban Energy Distribution and Traffic Congestion in Smart Cities. In 2025 IEEE International Conference on Electronics, Energy Systems and Power Engineering (EESPE) (pp. 765-768). IEEE.