Computers, Materials & Continua

Tech Science Press

Doi:10.32604/cmc.2025.066663

EDITORIAL

Guest Editorial Special Issue on the Next-Generation Deep Learning Approaches to Emerging Real-World Applications

Yu Zhou[1], **Eneko Osaba**[2] **and Xiao Zhang**[3,*]

[1]College of Computer Science and Software Engineering, Shenzhen University, Shenzhen, 518060, China
[2]TECNALIA, Basque Research and Technology Alliance (BRTA), Derio, 48160, Spain
[3]Department of Computer Science, South-Central Minzu University, Wuhan, 430074, China

*Corresponding Author: Xiao Zhang. Email: xiao.zhang@my.cityu.edu.hk

Received: 14 April 2025; Accepted: 30 April 2025; Published: 09 June 2025

1 Introduction

Deep learning (DL), as one of the most transformative technologies in artificial intelligence (AI), is undergoing a pivotal transition from laboratory research to industrial deployment. Advancing at an unprecedented pace, DL is transcending theoretical and application boundaries to penetrate emerging real-world scenarios such as industrial automation, urban management, and health monitoring, thereby driving a new wave of intelligent transformation. In August 2023, Goldman Sachs estimated that global AI investment will reach US$200 billion by 2025 [1]. However, the increasing complexity and dynamic nature of application scenarios expose critical challenges in traditional deep learning, including data heterogeneity, insufficient model generalization, computational resource constraints, and privacy-security trade-offs. The next generation of deep learning methodologies needs to achieve breakthroughs in multimodal fusion, lightweight design, interpretability enhancement, and cross-disciplinary collaborative optimization, in order to develop more efficient, robust, and practically valuable intelligent systems. Fueled by algorithmic innovations, enhanced computational capabilities, and explosive data growth, deep learning has rapidly evolved from laboratory research to industrial implementation, demonstrating transformative potential across critical domains. This special issue aims to explore cutting-edge advancements in applying deep learning to address complex real-world challenges and to showcase next-generation methodologies in emerging scenarios. Submissions for this special issue in *Computers, Materials & Continua* are open from 13 November 2023 to 31 October 2024 and contain 13 outstanding papers in the above research fields.

2 Articles Included in the Special Issue

Game-theoretic distributed optimization bridges privacy preservation and resource allocation, addressing efficiency-fairness trade-offs in decentralized systems through incentive mechanisms and combinatorial solutions. Lu et al. [2] proposed ENTIRE, a contract-based dynamic incentive mechanism for federated learning (FL), which addresses the challenges of model bias and privacy protection arising from dynamic client participation. Targeting challenges including client resource heterogeneity, non-IID data distributions, budget limitations, and concealed participation preferences, ENTIRE leverages contract design to achieve personalized allocation of participation levels and payments, thereby balancing model performance with incentive fairness. The study derives the convergence upper bound for federated learning under non-convex

loss functions, revealing a direct correlation between client participation probability and global model accuracy, and optimizes the contract schemes via the Lagrangian multiplier method. The proposed method undergoes a comprehensive experimental evaluation on the MNIST, Fashion-MNIST (Non-IID), and CIFAR-10 (IID) datasets. The results demonstrate a significant 12.9% improvement in model performance. This research represents the first integration of contract theory with FL, resolving preference concealment and incentive compatibility issues in dynamic participation scenarios. It establishes an extensible collaborative framework for privacy-sensitive domains such as healthcare and finance. Zhang et al. [3] proposed a RL-based approach for solving the knapsack problem, enhancing combinatorial optimization efficiency through state representation optimization and Noisy layers injection. Addressing the limitations of dynamic programming (DP) with exponential computational complexity and greedy algorithms prone to suboptimal solutions, their study innovatively normalized item weights and volumes as ratios relative to the knapsack's capacity while mapping item values to percentage contributions of total value, thereby eliminating scale bias and improving model generalization. By integrating Noisy Layers into the Dueling DQN framework to introduce weight stochasticity, the method enhanced exploration capabilities and avoided local optima. Experimental results demonstrate that the proposed method achieves a 5% improvement in accuracy over state-of-the-art RL-based approaches (e.g., Transformer models) while exhibiting a 9000× speedupcompared to DP, offering a scalable solution for logistics route optimization and real-time resource allocation in industrial applications.

Edge intelligence enables real-time perception in resource-constrained environments, where lightweight models and efficient attention mechanisms are critical for applications like UAV inspection and infrastructure monitoring. Said et al. [4] proposed CFEMNet, a UAV-based vehicle detection model for complex urban environments, addressing challenges including small vehicle targets, severe occlusion, and low computational efficiency in aerial imagery. To overcome the limitations of traditional detection models in balancing high-resolution details and semantic information, CFEMNet introduces a Context-aware Feature Extraction Module (CFEM) based on the HRNet architecture. This module employs parallel convolutional paths and Multi-head Self-Attention (MSA) to capture local semantic features and global spatial features separately, with dynamic weight-based adaptive fusion of these features, significantly enhancing detection capabilities for occluded and small targets. To reduce computational overhead, the model adopts local window MSA, achieving a 37% reduction in memory consumption while sacrificing only 0.3% accuracy. Evaluated on the VisDrone-DET2018 dataset, CFEMNet achieves 44.7% mAP, surpassing the baseline HRNet by 4.8%, with an inference speed of 7.1 FPS (NVIDIA GTX 960). Furthermore, its anchor-free detection head directly predicts vehicle center points and scales, eliminating hyperparameter tuning for preset anchor boxes. Zhou et al. [5] addressed challenges in underground pipeline defect detection—including target morphology diversity, complex backgrounds, and stringent real-time requirements—by proposing SDH-FCOS, an enhanced fully convolutional one-stage detector (FCOS). Built upon the FCOS framework, the model incorporates a Spatial Pyramid Pooling-Fast (SPPF) module to fuse multi-scale local details and global contextual features through hierarchical pooling, enhancing capture capabilities for defects like cracks and sediment deposits. Simultaneously, the Feature Pyramid Network (FPN) structure was optimized by adding top-down information flow paths to combine shallow and deep features and accelerate the circulation of shallow information. For scale variations in pipeline defects, SDH-FCOS employs a dual detection heads strategy, grouping feature layers into two sets processed by independent heads to reduce task conflicts and enhance efficiency. Experiments on the Sewer-ML dataset demonstrate that SDH-FCOS achieves 85.96% mAP, outperforming Faster R-CNN and YOLOv4 by 7.49% and 3.68%, respectively, while attaining real-time performance at 30.3 FPS (NVIDIA RTX 3060Ti) with a 12% reduction in missed detection rate. This study provides a high-precision, low-latency solution for automated pipeline

inspection. Ellouze et al. [6] proposed a human activity recognition framework integrating Convolutional Neural Network-Long Short-Term Memory (CNN-LSTM) with Dempster-Shafer (DS) evidence theory, addressing noise interference and classification uncertainty in smartphone sensor data under distributed environments. To overcome traditional models' limitations in joint spatial-temporal feature modeling, this research designed a hybrid architecture: CNN extracts local spatial patterns from accelerometer data, LSTM captures temporal dependencies of activities, and DS evidence theory fuses predictions from multiple models (CNN-LSTM, bidirectional LSTM, etc.), effectively reducing misclassification rates. Validated on the MMASH dataset, the DS fusion achieved 98% classification accuracy, providing real-time monitoring tools for chronic disease prevention.

Cognitive computing engines decode user behaviors and psychological states through multimodal analysis, enabling interpretable decision-making in human-AI interaction systems. Xing et al. [7] proposed a Multi-Head Encoder Shared Model (MESM) to address the challenge of collaborative modeling between intent and emotion in task-oriented dialogue systems. Targeting the limitations of traditional Dialogue State Tracking (DST) that neglects emotional information and suffers from scarce annotated data, MESM constructs a unified framework through multi-dataset joint training. The model converts intent slot-values and emotion labels into natural language summaries and designs a Feature Fusioner to dynamically integrate features from dual encoders. Specifically, using emotion encoder outputs as Query and intent encoder outputs as Key/Value, the model extracts cross-task correlations via multi-head attention mechanisms while preserving original features through residual connections. Experimental results demonstrate that MESM achieves 53.26% Joint Goal Accuracy (JGA) on MultiWoZ 2.1, outperforming DS2 and DualLoRA by 0.8%, with weighted emotion recognition accuracy (W-avg) reaching 63.48%–8.46% higher than Text-CNN. MESM establishes a novel emotion-enhanced dialogue management paradigm applicable to medical customer service and intelligent assistants. Zhang et al. [8] developed a User-aspect-sentence Graph Convolutional Neural Network (U-ASGCN) that integrates user features with syntactic dependencies for aspect-level sentiment analysis, overcoming traditional methods' neglect of user subjectivity and text-structure relationships. Addressing existing models' limitations in capturing user historical comment patterns and fine-grained semantics, the study designed a User Comment Feature Extraction (UCFE) module that dynamically fuses user historical features through BERT encoding and multi-head attention to generate preference-representative vectors. Simultaneously, an enhanced Graph Convolutional Network constructs dual-channel adjacency matrices (sentence-level and aspect-term level) based on syntactic dependency trees, modeling contextual associations and inter-aspect relationships through position-encoding weighting to enhance semantic capture. Validated on Yelp-2014 and a self-built User-Aspect dataset, U-ASGCN achieves 83.4% accuracy and 71.29% F1-score. This approach pioneers the joint modeling of user historical behavior and syntactic structures while establishing the first aspect-level dataset containing user information, providing more interpretable solutions for e-commerce review analysis and personalized recommendation systems. Zhu et al. [9] proposed a Dual-Layer User Representation model (DLUR) that enhances recommendation system interpretability and dynamic adaptability through the integration of semantic and sequential features. For temporal dynamics, its sequence layer employs Transformer-based encoding of user behavioral sequences to capture users' interests and hobbies. The model introduces an interactive attention mechanism to dynamically compute word-level user-item review correlations, generating explainable recommendation rationales. Experimental results on Amazon and Yelp datasets demonstrate DLUR's superiority, achieving 12%–15% improvements in Hit Ratio (HR) and Normalized Discounted Cumulative Gain (NDCG) over DeepCoNN and GRU4Rec, with particularly pronounced advantages in sparse data scenarios such as Clothing category recommendations. DLUR provides transparent recommendation solutions for e-commerce

and streaming platforms, with future potential for expansion into video content recommendations through multimodal feature integration.

Knowledge-enhanced decision systems leverage structured domain knowledge to improve semantic understanding and enable precise decision-making in complex scenarios. Wang et al. [10] developed a Diffusion Sampling and Label-Driven Co-attention Neural Network (DSLD) to address the limitations of conventional Transformers in long-range dependency modeling and label-induced semantic bias within complex semantic contexts. Targeting the insufficient semantic sensitivity of traditional Transformers in long-text processing, DSLD employs a multi-channel diffusion sampling encoder that generates text subsets with varying information densities through binomial distribution. These multi-scale representations are fused via parallel convolutional operations on attention matrices to strengthen contextual semantic encoding. Concurrently, its label-driven encoder injects label semantics into text representations by constructing label-text joint correlation matrices through embedded label vectors, thereby rectifying semantic biases from open-domain training. Evaluations on seven datasets including AG News reveal DSLD's 0.16%–0.47% accuracy improvements over Transformer and LEAM baselines. This end-to-end model demonstrates significant potential for long-text applications such as medical document classification and legal contract parsing without requiring complex pretraining. Chen et al. [11] proposed an emergency intent recognition model based on knowledge graph (KG) and data augmentation, effectively addressing the challenges of annotated data scarcity and semantic comprehension deviations in the urban rail transit (URT) domain. To leverage unstructured emergency knowledge (e.g., contingency plans), they constructed a Neo4j-based KG containing 500+ cases through entity extraction and relation mapping, generating 40,232 labeled samples via 25-class question templates. The model adopted an XLNet-BiLSTM-CNN architecture: XLNet captured long-range dependencies via permutation language modeling, BiLSTM extracted temporal features, CNN focused on keywords, and NLPCDA tools enhanced data diversity through synonym substitution and syntactic variation. Experiments achieved 96.98% accuracy and 96.62% F1-score in emergency intent classification, outperforming BERT-BiLSTM by 0.36%. This framework has been deployed in a metro emergency system, enabling sub-3-s precise responses for fault consultation.

Spatiotemporal signal analysis enables accurate prediction of urban dynamics through advanced decomposition and modeling techniques, supporting smart city planning and management. Zhou et al. [12] proposed a multi-view enhanced Transformer model (MVformer) to address the challenges of noise interference in cellular base station data and long-term time-series prediction of population density. To resolve the limitations of conventional temporal models in capturing high-frequency variations and periodic features, MVformer decomposes time series into varying-length subsequences through multi-view division. It incorporates both time-domain Transformer architecture and frequency-domain enhancement techniques, effectively integrating multi-scale temporal features through a cross-view attention mechanism. The study establishes systematic data preprocessing procedures to eliminate any irregular or redundant data. Experimental results on Shenzhen's base station dataset (about 7.8 million users) demonstrate that MVformer achieves 34.7% MSE reduction for univariate prediction and 31.3% error reduction for multivariate forecasting compared to FEDformer, with precise detection of population fluctuations caused by holidays and emergencies. This model provides high-precision tools for smart city traffic management and public safety alerts. Li et al. [13] developed a hybrid model combining CEEMDAN signal decomposition with ConvLSTM for taxi demand prediction, addressing nonlinear forecasting challenges influenced by weather, holidays, and other factors in urban transportation. Targeting conventional models' inadequacy in non-stationary time series modeling, the study employs CEEMDAN to decompose original demand sequences into high-frequency, medium-frequency, and low-frequency modal components. Sample entropy-based clustering reconstruction reduces error accumulation and enhances decomposition efficiency. The reconstructed

components and external factors are jointly input into a ConvLSTM network, where convolutional layers capture spatial hotspot regions and LSTM layers model temporal dependencies. Experiments on Beijing taxi data show CEEMDAN-ConvLSTM achieves 12.7% SMAPE, outperforming LSTM and ARIMA by 21.03% and 34.5%, respectively, with significant advantages over comparative models like EEMD-ConvLSTM. This provides a robust solution for dynamic urban traffic scheduling, demonstrating substantial improvements in handling complex spatiotemporal correlations. Awan and Mehmood [14] proposed the Accident Severity Level Prediction Deep Learning (ASLP-DL) lightweight deep learning framework, addressing challenges of data heterogeneity and insufficient model generalization in traffic accident severity prediction. Targeting traditional methods' difficulties in capturing nonlinear relationships among multifactorial interactions (road surface conditions, temporal factors, weather patterns), ASLP-DL integrates DNN, D-CNN, and D-RNN models through grid search-optimized hyperparameter tuning, combined with data augmentation to mitigate class imbalance. Cross-regional experimental results demonstrated the D-RNN model's 89.03% accuracy and 0.751 AUC score, outperforming SVM and TabNet by 43.63% and 20.03%, respectively. The framework achieved prediction errors below 5% in high-risk scenarios like nighttime wet road conditions, providing transportation authorities with dynamic risk assessment tools.

3 Conclusion

Next-generation deep learning methods for emerging real-world applications are overcoming traditional technical bottlenecks and driving the large-scale deployment of intelligent systems across industrial inspection, urban management, and multimodal interaction scenarios. Lightweight architectures (e.g., CFEMNet and SDH-FCOS) significantly enhance edge computing efficiency, enabling high-precision real-time analysis for UAV-based inspections and underground pipeline defect detection. Multimodal fusion frameworks (e.g., MESM and U-ASGCN) optimize interpretability in dialogue systems and recommendation services through emotion-aware enhancement and user-specific feature modeling. Privacy-preserving frameworks (e.g., ENTIRE for federated learning) and noise-augmented reinforcement learning address challenges in dynamic decision-making and data heterogeneity, providing robust solutions for medical collaboration and logistics optimization. As technology continues to evolve, we stand at a pivotal juncture of the intelligent revolution. Future research should focus on exploring meta-learning-based dynamic architecture adaptation techniques to enable edge devices to self-adjust to data drift; integrating causal reasoning with multimodal fusion to build more interpretable decision systems; developing joint modeling approaches combining spatiotemporal knowledge with neural networks to enhance long-term prediction accuracy; designing hardware-aware neural architecture search algorithms to optimize the energy efficiency of edge computing; and investigating blockchain-empowered federated learning frameworks to ensure secure collaboration with sensitive data. These innovative directions will propel intelligent systems toward greater efficiency, robustness, and trustworthiness, providing sustained momentum for industrial intelligent transformation—particularly in addressing critical challenges such as handling complex dynamic environments, safeguarding data privacy, and improving system transparency.

Funding Statement: This work was supported in part by Guangdong Basic and Applied Basic Research Foundation under Grant 2024A1515012485, in part by Shenzhen Fundamental Research Program under Grant JCYJ20220810112354002, in part by Shenzhen Science and Technology Program under Grant KJZD20230923114111021, in part by the Fund for Academic Innovation Teams and Research Platform of South-Central Minzu University under Grant XTZ24003 and Grant PTZ24001, in part by the Knowledge Innovation Program of Wuhan-Basic Research through Project 2023010201010151, in part by the Research Start-up Funds of South-Central Minzu University under Grant YZZ18006, and in part by the Spring Sunshine Program of Ministry of Education of the People's Republic of China under Grant HZKY20220331.

Conflicts of Interest: The authors declare no conflicts of interest to report regarding the present study.

References

1. He M. Investment in R&D of AI: evidence from the global south. In: Tech transformation and AI readiness: pioneering paths for the global south. Cham: Springer Nature Switzerland; 2025. p. 61–85.
2. Lu J, Huang T, Xie Y, Cao S, Li B. A federated learning incentive mechanism for dynamic client participation: unbiased deep learning models. Comput Mater Contin. 2025;83(1):619–34. doi:10.32604/cmc.2025.060094.
3. Zhang Z, Yin HY, Zuo LD, Lai P. Reinforcement learning for solving the knapsack problem. Comput Mater Contin. 2025;84(1):919–36. doi:10.32604/cmc.2025.062980.
4. Said Y, Alassaf Y, Saidani T, Ghodhbani R, Rhaiem OB, Alalawi AA. Context-aware feature extraction network for high-precision UAV-based vehicle detection in urban environments. Comput Mater Contin. 2024;81(3):4349–70. doi:10.32604/cmc.2024.058903.
5. Zhou B, Li B, Lan W, Tian C, Yao W. SDH-FCOS: an efficient neural network for defect detection in urban underground pipelines. Comput Mater Contin. 2024;78(1):633–52. doi:10.32604/cmc.2023.046667.
6. Ellouze A, Kadri N, Alaerjan A, Ksantini M. Combined CNN-LSTM deep learning algorithms for recognizing human physical activities in large and distributed manners: a recommendation system. Comput Mater Contin. 2024;79(1):351–72. doi:10.32604/cmc.2024.048061.
7. Xing X, Chen J, Zhang X, Zhou S, Zhang R. Multi-head encoder shared model integrating intent and emotion for dialogue summarization. Comput Mater Contin. 2025;82(2):2275–92. doi:10.32604/cmc.2024.056877.
8. Zhang M, Chai J, Cao J, Ji J, Yi T. Aspect-level sentiment analysis based on deep learning. Comput Mater Contin. 2024;78(3):3743–62. doi:10.32604/cmc.2024.048486.
9. Zhu F, Xie J, Alshahrani M. Learning dual-layer user representation for enhanced item recommendation. Comput Mater Contin. 2024;80(1):949–71. doi:10.32604/cmc.2024.051046.
10. Wang C, Shang W, Yi T, Zhu H. Enhancing deep learning semantics: the diffusion sampling and label-driven co-attention approach. Comput Mater Contin. 2024;79(2):1939–56. doi:10.32604/cmc.2024.048135.
11. Chen Y, Wu X, Fan J, Zhu G. A data-enhanced deep learning approach for emergency domain question intention recognition in urban rail transit. Comput Mater Contin. 2025;84(1):1597–613. doi:10.32604/cmc.2025.062779.
12. Zhou Y, Lin B, Hu S, Yu D. An enhanced multiview transformer for population density estimation using cellular mobility data in smart city. Comput Mater Contin. 2024;79(1):161–82. doi:10.32604/cmc.2024.047836.
13. Li M, Gu Y, Geng Q, Yu H. A combination prediction model for short term travel demand of urban taxi. Comput Mater Contin. 2024;79(3):3877–96. doi:10.32604/cmc.2024.047765.
14. Awan S, Mehmood Z. ASLP-DL—a novel approach employing lightweight deep learning framework for optimizing accident severity level prediction. Comput Mater Contin. 2024;78(2):2535–55. doi:10.32604/cmc.2024.047337.