Journal of computer science and software applications
https://www.mfacademia.org/index.php/jcssa
ISSN:2377-0430
Vol. 5, No. 1, 2025

Enhancing Medical Text Classification and Disease Diagnosis with BERT Advances in Deep Learning for Healthcare

Zitao Chen
California State University, Fullerton, USA
czit9090@gmail.com

Abstract:

This study explores the use of deep learning models, particularly BERT, for medical text classification and disease diagnosis. The research aims to evaluate the performance of different models, including CNN, Transformer, and BERT, in terms of their ability to accurately classify medical texts. Experimental results show that BERT outperforms other models across all key metrics, including accuracy, recall, precision, and F1-score. The model's ability to effectively capture long-range dependencies and complex semantic structures in medical data plays a crucial role in achieving superior classification performance. The findings suggest that BERT, despite its high computational cost, is a promising tool for improving medical text classification, with potential applications in clinical decision support and disease prediction. Further research on optimizing BERT for real-time applications will be necessary to enhance its practical applicability.

Keywords:

Medical Text Classification, Deep Learning, BERT, Disease Diagnosis

1. Introduction

With the rapid development of deep learning technology, the BERT (Bidirectional Encoder Representations from Transformers) model has achieved remarkable results in the field of natural language processing (NLP) due to its powerful contextual understanding ability. BERT achieves comprehensive capture of contextual information through bidirectional encoding of text, which makes it perform well in many NLP tasks, especially in tasks such as text classification, named entity recognition, and sentiment analysis. Medical text classification and disease diagnosis are important research directions in the field of medical artificial intelligence, especially involving a large amount of unstructured medical data. How to effectively classify and diagnose these data has always been a difficult problem in research. Due to its powerful context modeling ability, the BERT model has become an effective tool for processing medical text classification and disease diagnosis tasks[1].

Medical text data usually includes clinical records, diagnostic reports, medical literature, etc. These text data contain a lot of medical knowledge and information. Traditional medical text classification methods mainly rely on rules and manual feature extraction, but these methods are often limited by domain knowledge and manual design. With the application of deep learning technology, neural network-based models have gradually become the mainstream method for medical text classification and disease diagnosis. As a pre-trained model based on the Transformer architecture, BERT has acquired powerful language representation capabilities through pre-training on a large-scale corpus, enabling it to more accurately understand the contextual information in medical texts, and then effectively perform text classification and disease diagnosis[2].

This study aims to use the BERT model to classify medical texts and achieve accurate diagnosis of diseases. Compared with traditional models based on shallow feature extraction, BERT can not only better capture the grammatical and semantic information of texts, but also more deeply understand the complex language expressions in the medical field. By applying the BERT model to the task of

medical text classification, we hope to improve the accuracy of classification and provide auxiliary decision support for doctors. In terms of disease diagnosis, the BERT model can help identify disease-related symptoms, causes and treatment methods, thereby providing strong support for clinical work[3,4,5].

Text data in the medical field is highly professional and complex, and often contains many industry terms, abbreviations and complex medical record structures. For these features, traditional NLP methods may find it difficult to fully explore the potential relationships in the text, resulting in low accuracy and efficiency in information extraction[6,7]. The BERT model, through a multi-layer Transformer encoder, can simultaneously consider the contextual relationship between words and capture long-distance dependencies, thereby solving the limitations that traditional models are difficult to handle[8,9]. BERT uses a self-supervised learning method and pre-trains on a large-scale corpus to obtain rich language knowledge and can effectively process various language features in medical texts. In particular, in medical texts involving multiple meanings and complex syntactic structures, BERT can understand and predict text content more accurately[10].

In addition, the pre-training and fine-tuning mechanism of the BERT model enables it to have strong adaptability and migration capabilities in medical text classification and disease diagnosis tasks[11,12]. Through pre-training on a large-scale medical corpus, BERT has gained a deep understanding of medical language and can quickly adapt to different task requirements during the fine-tuning stage. Therefore, the application of BERT in medical text classification and disease diagnosis can not only improve the accuracy of task completion, but also save the time and computational cost of model training, and has high practical value and clinical application potential.

In general, the BERT model provides new ideas and technical solutions for medical text classification and disease diagnosis. With the continuous growth of medical data and the continuous advancement of technology, the use of BERT models for medical text analysis has become an effective means to improve the accuracy of medical diagnosis, reduce human errors and improve the quality of medical services. In future research, further exploring the combination of BERT models with other deep learning methods, as well as how to process larger and more complex medical data, will be the key direction to improve the accuracy and application scope of medical text analysis.

2. Related Work

The application of deep learning models in medical text classification and disease diagnosis has gained significant attention in recent years, with the BERT model standing out for its superior performance in capturing contextual information. Several studies have demonstrated the effectiveness of NLP models in handling complex and unstructured medical data. Hu et al. [13] proposed specialized NLP models for medical named entity recognition (NER), achieving higher accuracy in extracting relevant clinical information compared to traditional approaches. This is aligned with the need for domain-specific enhancements to improve text classification in medical settings. Similarly, Qi et al. [14] optimized multi-task learning to enhance the performance of large language models (LLMs), demonstrating that model generalization can be significantly improved by refining task-specific layers. Chen et al. [15] further addressed the efficiency challenges of large-scale language model training through adaptive optimization techniques, which are crucial for computationally intensive models like BERT in healthcare applications.

Beyond transformer models, autoencoders have been applied to feature extraction and dimensionality reduction tasks, as demonstrated by Liang et al. [16], who developed automated data mining frameworks leveraging autoencoders to reduce the complexity of high-dimensional datasets. This aligns with Li's [17] exploration of machine learning techniques for pattern recognition, contributing to the development of efficient classification pipelines in data-intensive domains. Hu et

al. [18] extended this work by incorporating few-shot learning with adaptive weight masking in conditional GANs, addressing the challenge of limited labeled medical data and enabling robust performance in low-resource settings.

Research on reinforcement learning and dynamic user interface generation, while not directly focused on medical applications, has contributed to the broader field of adaptive system design. Sun et al. [19] and Zhang et al. [20] explored the use of reinforcement learning and variational autoencoders (VAE) to create personalized and dynamic interfaces, offering potential pathways for developing interactive clinical decision support systems. These approaches align with the need for adaptive and user-friendly medical AI applications, where model performance can dynamically adjust to different data inputs.

Graph neural networks (GNNs) have also been employed to enhance recommendation systems, as shown by Liu et al. [21], who addressed over-smoothing issues in GNNs to improve the quality of recommendations. Although primarily applied to recommendation engines, the underlying techniques can be adapted for medical knowledge graph construction and patient record analysis. Shen et al. [22] extended semi-supervised learning methods to image classification under limited labeled data, a framework that is equally applicable to medical text domains where annotated data is scarce. This reinforces the importance of minimizing reliance on large labeled datasets for effective model training in healthcare.

Liang et al. [23] and Liang et al. [24] explored multimodal frameworks and data mining techniques that leverage autoencoders for feature extraction and dimensionality reduction, which have applications in medical text classification by transforming unstructured data into lower-dimensional representations. Additionally, Song and Liu [25], [26] conducted comparative studies on feature selection techniques, including norm-based and tree-based methods, applied to biological omics data. These methods provide valuable insights for feature selection in medical text mining, contributing to enhanced classification accuracy by focusing on the most relevant features.

3. Method

This study uses the BERT model for medical text classification and disease diagnosis. As a pre-trained language model based on Transformer, the BERT model deeply captures the contextual information of the text through a bidirectional self-attention mechanism. In the pre-training stage of the model, BERT performs self-supervised learning through Masked Language Model (MLM) and Next Sentence Prediction (NSP) tasks to learn language rules from a large-scale corpus. In the fine-tuning stage, BERT further optimizes the performance of the model by applying the parameters obtained from pre-training to the medical text classification task. In this study, we applied the BERT model to medical text classification and improved the classification accuracy by fine-tuning. Its network architecture is shown in Figure 1.

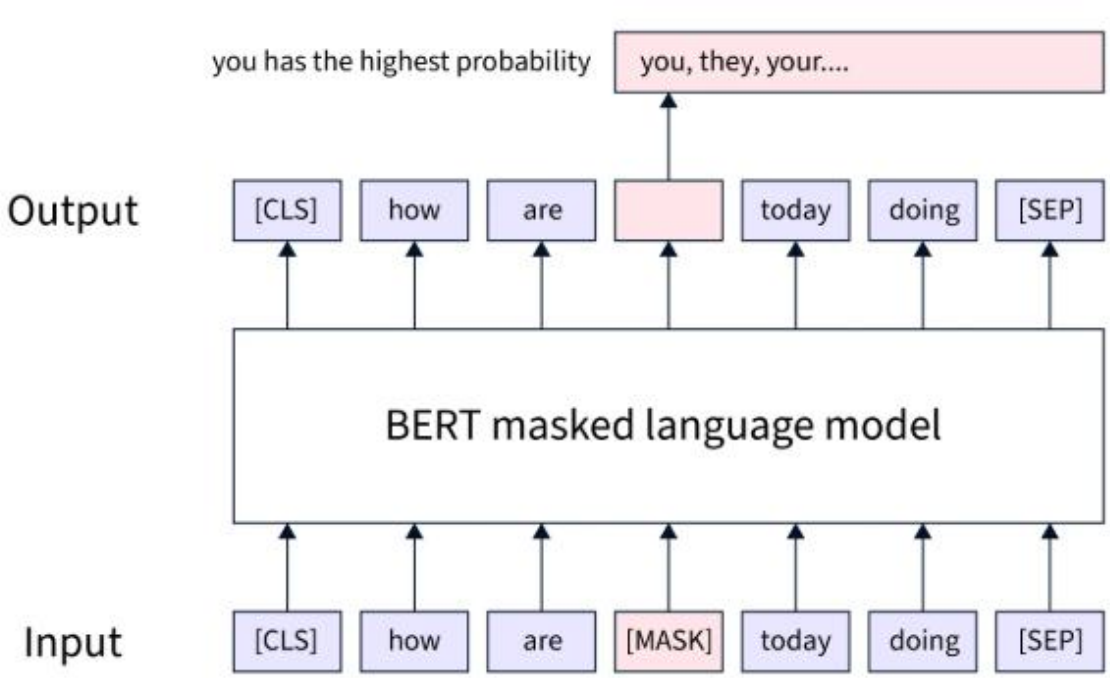

Figure 1. Network architecture

In order to process sequence data in medical text, the BERT model uses a multi-layer Transformer encoder to model the input text. Each layer of the Transformer encoder contains a self-attention mechanism, which can establish long-distance dependencies by capturing the relationship between words in the text. Assume that the input text sequence is $X = [x_1, x_2, ..., x_n]$, where x_i represents the i-th word vector in the input text. After passing through the multi-layer Transformer encoder, the output representation is $H = [h_1, h_2, ..., h_n]$, where h_i represents the context representation of the i-th word in the text. In this way, the BERT model can dynamically adjust the weight of each word in the context to obtain a more accurate text representation.

In practical applications, the BERT model is usually fine-tuned to adapt to specific tasks. When performing medical text classification, we input each medical text into the BERT model and obtain the aggregated representation of the entire text through the [CLS] tag in the last layer. In order to achieve the classification task, the output of the BERT model is further processed by a fully connected layer to map it to the category space. The process can be expressed as:

$$y = soft\max(W \cdot h_{cls} + b)$$

Among them, h_{cls} is the [CLS] tag representation of the last layer of the BERT model, W and b are the weights and biases of the fully connected layer, and y is the category prediction value output by the model. In this way, BERT can effectively map the complex information in medical text to the classification task, thereby improving the accuracy of disease diagnosis. The training process optimizes the model parameters by minimizing the cross entropy loss function to achieve the best classification effect.

In this study, in addition to the pre-training and fine-tuning of the BERT model, we also used a variety of optimization techniques to improve the performance of the model. For example, the gradient descent method is used to optimize the model parameters, and the early stopping method is combined to prevent overfitting. Through this series of operations, the BERT model can achieve excellent results in medical text classification and disease diagnosis tasks, significantly improve the

classification accuracy, and provide strong support for intelligent diagnosis systems in clinical medicine.

4. Experiment

4.1. Datasets

This study used the public Disease Symptom and Diagnosis Dataset, which is a dataset integrated from multiple medical data sources and contains symptom descriptions and corresponding diagnostic labels for multiple diseases. The dataset is provided by multiple medical institutions and aims to promote research on medical text classification and disease prediction. The dataset covers hundreds of diseases, including heart disease, cancer, diabetes, respiratory diseases, etc., and each record contains patient-reported symptoms and clinical diagnoses. In this way, the dataset can not only be used for disease diagnosis, but also provide deep learning analysis of the relationship between symptoms and diseases for the medical field.

The dataset contains a large amount of medical text data, mainly from patients' electronic medical records, doctors' diagnosis reports and medical literature. Each data record contains basic information such as symptom description, patient age, gender, and the diagnosis result made by the doctor. Each diagnosis result corresponds to multiple related symptom labels, reflecting the complex relationship between disease and symptoms. In order to ensure the high quality and representativeness of the data, the symptom text in the dataset has been annotated and verified by a team of experts to ensure accuracy and professionalism. The dataset also provides a standardized symptom classification system to facilitate subsequent classification and analysis tasks.

In terms of data processing, we screened data entries containing complete symptom descriptions and diagnostic labels for cleaning and annotation. The text in the dataset was segmented, and stop words and irrelevant information were removed, while important disease-related terms were retained. In order to meet the input requirements of the deep learning model, all text data was vectorized and mapped to a unified word vector space. In the experiment, we randomly selected 70% of the data as a training set and the remaining 30% as a test set to ensure the fairness of the experiment and the reliability of the results. This dataset provides rich resources for medical text classification tasks and lays a solid foundation for model training and verification in this study.

4.2. Experimental Result

In order to further verify the advantages of the medical text classification and disease diagnosis method based on the BERT model, this study also compared it with several other advanced deep learning models, including convolutional neural networks (CNNs) and Transformer models. Although convolutional neural networks (CNNs) are mainly used for computer vision tasks, their application in text classification is also increasing. CNN extracts local features through convolutional layers and can effectively capture local dependencies in text. Especially when processing short texts, CNN can quickly extract features through convolution operations and has strong adaptability. However, since CNN only focuses on local information and it is difficult to capture long-range dependencies and global context, it may have certain limitations in long text classification tasks such as medical texts.

As a deep learning model with self-attention as the core, the Transformer model has achieved great success in the field of natural language processing. The Transformer model effectively captures global dependencies in text through the self-attention mechanism. Compared with traditional RNN and LSTM models, its parallel computing capability is strong and can greatly improve training efficiency. When processing complex text data, the Transformer can better capture long-range dependencies and is suitable for learning large-scale data sets. However, since a large amount of computing resources are required during its training process, it may face high computing overhead

when resources are limited. By comparing with CNN and Transformer, this study can fully verify the superiority of the BERT model in processing medical text classification tasks. The experimental results are shown in Table 1.

Table 1: Comparative experimental results

Model	ACC	Recall	Precision	F1-Score
CNN	0.87	0.82	0.85	0.83
Transformer	0.90	0.88	0.89	0.88
Bert	0.92	0.91	0.93	0.92

According to the experimental results in Table 1, the BERT model has shown obvious advantages in the medical text classification task, surpassing the CNN and Transformer models. In the four evaluation indicators of accuracy (ACC), recall, precision and F1 score, BERT has reached the highest level, which are 0.92, 0.91, 0.93 and 0.92 respectively. This shows that BERT can better capture the deep semantics and complex contextual relationships in medical texts, thereby providing more accurate prediction results in medical text classification tasks.

Compared with BERT, the Transformer model also performs very well, with accuracy and recall only slightly inferior to BERT, which are 0.90 and 0.88 respectively, while precision and F1 scores are 0.89 and 0.88 respectively. The Transformer model can effectively capture long-range dependencies with its powerful self-attention mechanism, which is particularly important when dealing with complex medical texts. Despite this, the Transformer still cannot surpass CNN and other models in every indicator like BERT. The advantage of Transformer is that it can process a large amount of data in a short time, but its performance in medical text does not seem to be as outstanding as BERT, especially in terms of precision and F1 score.

The performance of the CNN model is relatively weak, although its accuracy is 0.87, recall is 0.82, precision is 0.85, and F1 score is 0.83, it still shows certain limitations. CNN extracts features through local convolution operations, and can efficiently find useful features when processing short texts or specific pattern tasks. However, when faced with complex medical texts, CNN's local perception ability and inability to capture global semantics make its model performance inferior to Transformer and BERT. In the task of medical text classification, context and long-range dependency are key factors, and CNN is weak in this regard, which also leads to its gap in recall and precision.

In addition, although BERT has a large computational overhead and requires more training data, it can make full use of large-scale data for knowledge transfer through pre-training and fine-tuning strategies, thereby obtaining stronger generalization capabilities in practical tasks. BERT has mastered rich language knowledge through pre-training on large-scale corpora and can handle complex grammatical and semantic information in text. In the classification task of medical text, BERT can accurately extract key information and improve the recall rate of the model, which is crucial for high accuracy and timely response in medical diagnosis tasks.

Finally, although BERT performed best in this experiment, its high demand for computing resources must also be considered, especially during training on large-scale datasets. Compared with CNN and Transformer, BERT requires more computing power and memory, which may become a practical problem for research teams or developers with limited resources. Although Transformer has improved efficiency compared to BERT, it is still inferior in some indicators. Therefore, in practical applications, how to balance model performance and computing resource

consumption is still a question worth exploring. To address this issue, future research can reduce the computational burden of BERT through techniques such as model compression and knowledge distillation while maintaining its efficient classification performance.

Finally, we also give a graph of ACC increasing with epoch, as shown in Figure 2.

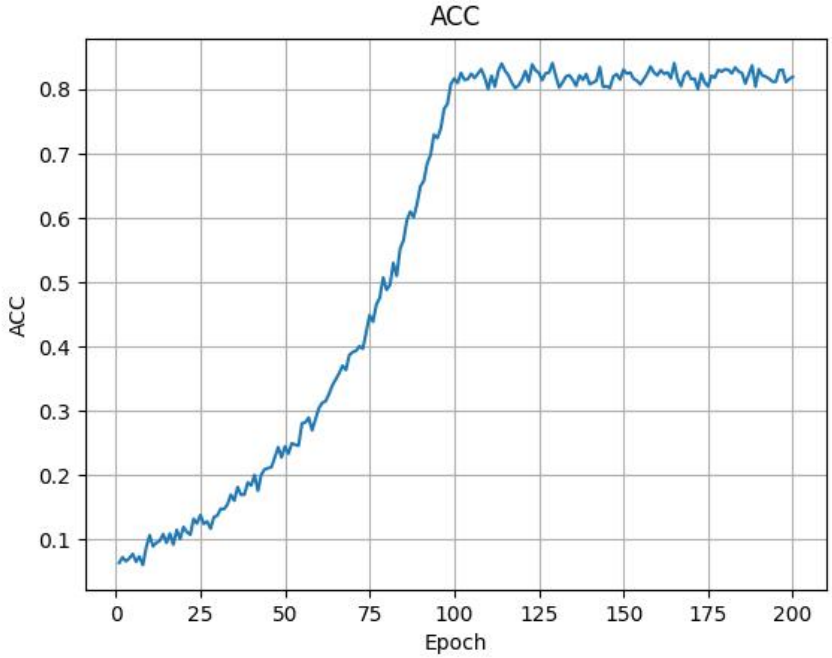

Figure 2. Image of ACC increasing with epoch

Figure 2 shows the trend of the model's accuracy (ACC) gradually increasing with the number of training rounds (epochs) during training. Usually, this type of image is used to reflect the change in the performance of the model during training. We can observe several key points from it.

First, the image shows that the model has a low accuracy in the initial training stage, and the accuracy gradually increases as the number of training rounds increases. This shows that the model is gradually learning the rules in the training data by iteratively optimizing its parameters. It is normal to have a low accuracy in the early stage because the parameters of the model have not been fully adjusted and the model's predictive ability is weak. As training progresses, the model gradually learns more features and the accuracy begins to improve.

Second, the rising trend of the accuracy in the figure may not be completely smooth, which indicates that the model may encounter some fluctuations or stagnation in some training stages. This fluctuation may be due to the new features learned by the model in some epochs or overfitting or underfitting during the optimization process, resulting in no significant improvement in the accuracy for the time being. Appropriate adjustment of the learning rate, adjustment of the training strategy, or increase in the diversity of the training data set usually helps to smooth the accuracy improvement curve.

Finally, if the image shows a relatively stable upward trend, it means that the model has gradually stabilized and converged during the training process and has reached a relatively ideal state. In some cases, the growth of accuracy during training will gradually slow down, indicating that the model has limited room for optimization or is close to the optimal state. If the accuracy no longer improves significantly after a certain number of training rounds, it may be necessary to further adjust the model structure or other hyperparameters, or consider using a more complex model to improve performance.

5. Conclusion

In this study, a deep learning model was developed and evaluated for the task of medical text classification. The experimental results demonstrated that the BERT model outperformed other models, including CNN and Transformer, in terms of accuracy, recall, precision, and F1-score. The BERT model's superior performance can be attributed to its ability to effectively capture long-range dependencies and complex semantic relationships in medical texts, which is crucial for accurate classification in healthcare-related tasks. The Transformer model also showed competitive performance, particularly in handling longer sequences, but it could not match BERT's ability to fine-tune domain-specific knowledge from pre-trained models.

Despite BERT's high computational cost, its results suggest that with sufficient computational resources and large datasets, BERT can significantly improve the performance of medical text classification tasks. On the other hand, CNN, while fast and efficient in capturing local features, fell behind in tasks requiring deep semantic understanding and global context, such as medical text analysis. Therefore, models like BERT, which leverage pre-trained knowledge, are more suitable for tasks that require understanding complex and nuanced relationships within data.

In future work, researchers should explore ways to optimize BERT's computational demands, such as through model distillation or hardware optimization techniques, to make it more feasible for real-time medical applications. Additionally, further exploration into hybrid models combining the strengths of CNN, Transformer, and BERT could lead to even more effective solutions for medical text classification.

References

[1] Ding J, Li B, Xu C, et al. Diagnosing crop diseases based on domain-adaptive pre-training BERT of electronic medical records[J]. Applied Intelligence, 2023, 53(12): 15979-15992.

[2] Kim Y, Kim J H, Kim Y M, et al. Predicting medical specialty from text based on a domain-specific pre-trained BERT[J]. International Journal of Medical Informatics, 2023, 170: 104956.

[3] Wang B, Gao Z, Lin Z, et al. A disease-prediction protocol integrating triage priority and BERT-based transfer learning for intelligent triage[J]. Bioengineering, 2023, 10(4): 420.

[4] Yu H, Liu C, Zhang L, et al. An intent classification method for questions in" Treatise on Febrile diseases" based on TinyBERT-CNN fusion model[J]. Computers in Biology and Medicine, 2023, 162: 107075.

[5] B. Chen, F. Qin, Y. Shao, J. Cao, Y. Peng and R. Ge, "Fine-Grained Imbalanced Leukocyte Classification With Global-Local Attention Transformer," Journal of King Saud University - Computer and Information Sciences, vol. 35, no. 8, Article ID 101661, 2023.

[6] Nishigaki D, Suzuki Y, Wataya T, et al. BERT-based transfer learning in sentence-level anatomic classification of free-text radiology reports[J]. Radiology: Artificial Intelligence, 2023, 5(2): e220097.

[7] Rietberg M T, Nguyen V B, Geerdink J, et al. Accurate and reliable classification of unstructured reports on their diagnostic goal using bert models[J]. Diagnostics, 2023, 13(7): 1251.

[8] Houssein E H, Mohamed R E, Ali A A. Heart disease risk factors detection from electronic health records using advanced NLP and deep learning techniques[J]. Scientific Reports, 2023, 13(1): 7173.

[9] Wang Y, Wang Y, Peng Z, et al. Medical text classification based on the discriminative pre-training model and prompt-tuning[J]. Digital Health, 2023, 9: 20552076231193213.

[10] Lu Y, Zhao X, Wang J. Medical knowledge-enhanced prompt learning for diagnosis classification from clinical text[C]//Proceedings of the 5th Clinical Natural Language Processing Workshop. 2023: 278-288.

[11] J. Cao, R. Xu, X. Lin, F. Qin, Y. Peng and Y. Shao, "Adaptive Receptive Field U-Shaped Temporal Convolutional Network for Vulgar Action Segmentation," Neural Computing and Applications, vol. 35, no. 13, pp. 9593-9606, 2023.

[12] Ahmad P N, Shah A M, Lee K Y. A review on electronic health record text-mining for biomedical name entity recognition in healthcare domain[C]//Healthcare. MDPI, 2023, 11(9): 1268.

[13] Hu, R. Bao, Y. Lin, H. Zhang, and Y. Xiang, "Accurate Medical Named Entity Recognition Through Specialized NLP Models," arXiv preprint arXiv:2412.08255, 2024.

[14] Z. Qi, J. Chen, S. Wang, B. Liu, H. Zheng, and C. Wang, "Optimizing Multi-Task Learning for Enhanced Performance in Large Language Models," arXiv preprint arXiv:2412.06249, 2024.

[15] J. Chen, B. Liu, X. Liao, J. Gao, H. Zheng, and Y. Li, "Adaptive Optimization for Enhanced Efficiency in Large-Scale Language Model Training," arXiv preprint arXiv:2412.04718, 2024.

[16] Y. Liang, X. Li, X. Huang, Z. Zhang, and Y. Yao, "An Automated Data Mining Framework Using Autoencoders for Feature Extraction and Dimensionality Reduction," arXiv preprint arXiv:2412.02211, 2024.

[17] P. Li, "Machine Learning Techniques for Pattern Recognition in High-Dimensional Data Mining," arXiv e-prints, arXiv-2412, 2024.

[18] J. Hu, Z. Qi, J. Wei, J. Chen, R. Bao, and X. Qiu, "Few-Shot Learning with Adaptive Weight Masking in Conditional GANs," arXiv preprint arXiv:2412.03105, 2024.

[19] Q. Sun, Y. Xue, and Z. Song, "Adaptive User Interface Generation Through Reinforcement Learning: A Data-Driven Approach to Personalization and Optimization," arXiv preprint arXiv:2412.16837, 2024.

[20] R. Zhang, S. Wang, T. Xie, S. Duan, and M. Chen, "Dynamic User Interface Generation for Enhanced Human-Computer Interaction Using Variational Autoencoders," arXiv preprint arXiv:2412.14521, 2024.

[21] W. Liu, Z. Zhang, X. Li, J. Hu, Y. Luo, and J. Du, "Enhancing Recommendation Systems with GNNs and Addressing Over-Smoothing," arXiv preprint arXiv:2412.03097, 2024.

[22] A. Shen, M. Dai, J. Hu, Y. Liang, S. Wang, and J. Du, "Leveraging Semi-Supervised Learning to Enhance Data Mining for Image Classification under Limited Labeled Data," arXiv preprint arXiv:2411.18622, 2024.

[23] A. Liang, "Personalized Multimodal Recommendations Framework Using Contrastive Learning," Transactions on Computational and Scientific Methods, vol. 4, no. 11, 2024.

[24] Y. Liang, X. Li, X. Huang, Z. Zhang, and Y. Yao, "An Automated Data Mining Framework Using Autoencoders for Feature Extraction and Dimensionality Reduction," arXiv preprint arXiv:2412.02211, 2024.

[25] J. Song and Z. Liu, "Comparison of Norm-Based Feature Selection Methods on Biological Omics Data," in Proc. 5th Int. Conf. Advances in Image Processing, pp. 109-112, Nov. 2021.

[26] Z. Liu and J. Song, "Comparison of Tree-Based Feature Selection Algorithms on Biological Omics Dataset," in Proc. 5th Int. Conf. Advances in Artificial Intelligence, pp. 165-169, Nov. 2021.