Comparative analysis of X-ray image classification of pneumonia based on deep learning algorithm

Danqing Ma[1], Yuanfang Yang[2,7], Qiyuan Tian[3], Bo Dang[4] ,Zongqing Qi[5] , Ao Xiang[6]

[1]Computer Information Tech, Northern Arizona University, Flagstaff, AZ, U.S.
[2]Computer Science, Southern Methodist University, Dallas, TX, U.S.
[3]Information System Technology, George Washington University, Washington, DC, U.S.
[4]Computer Science, San Francisco Bay University, Fremont, CA, U.S.
[5]Computer Science, Stevens Institute of technology, Hoboken, Fremont, CA, U.S.
[6]Digital Media Technology, University of Electronic Science and Technology of China, Sichuan, CN

[7]madanqing376@gmail.com.

Abstract. In this paper, we compare and analyse the effectiveness of deep learning models such as AlexNet, Vgg, GoogleNet and MobileNet in the task of pneumonia image classification. During the training process, GoogleNet showed the fastest convergence speed, reaching convergence at the 3rd epoch; subsequently, the other three models gradually stabilised. In the end, the loss of AlexNet is 0.426, the loss of Vgg is 0.566, the loss of GoogleNet is 0.936, and the loss of MobileNet is 0.626. By selecting the weights of the round with the best training effect of each model, the results of classification accuracy are obtained: 88.9% for AlexNet, 92.6% for Vgg, and 92.6% for GoogleNet. 92.6%, GoogleNet is 85.2%, and MobileNet is 96.3%. MobileNet demonstrated the best prediction performance on the test set. These results provide a useful reference for the application of deep learning models in medical imaging..

Keywords: X-ray image、MobileNet、deep learning algorithm.

1. Introduction

Deep learning models [1], as one of the branches of machine learning [2], are widely used in the field of computer vision technology, and the field of image recognition, represented by medical image analysis, is also growing [3]. Artificial Intelligence has far-reaching impact on labour force [4], economic dynamics [5], and ethical issues [6], in which the study of classification of X-ray images of neocoronary pneumonia, X-ray images of healthy individuals and X-ray images of community-acquired pneumonia has attracted much attention. Neocoronavirus pneumonia is an infectious disease caused by a novel coronavirus, and its imaging manifestations in the lungs are characterised by certain features on X-ray images. By classifying and analysing X-ray images of neocoronavirus pneumonia, it can help doctors diagnose whether patients are infected with neocoronavirus more quickly and accurately, so as to take timely and effective treatment measures. As a comparison group, X-ray images of healthy people

play an important role in the classification of X-ray images of patients with neocoronavirus pneumonia. Through in-depth analysis and comparison of CT images of healthy people groups, the characteristic imaging manifestations of neocoronary pneumonia can be better distinguished and the accuracy of classification and diagnosis can be improved. Community-acquired pneumonia is an infectious disease of the lungs acquired in a community setting. Compared with neocoronal pneumonia, community-acquired pneumonia has different features on X-ray images. Therefore, differentiating community-acquired pneumonia from other types of lung diseases is also an important diagnostic problem in the field of medical imaging [7].

Deep learning algorithms play a vital role in these image classification problems [8]. Deep learning algorithms are able to automatically learn features from a large amount of well-labelled training data and build efficient models for image classification [9]. For example, convolutional neural network (CNN) is a deep learning algorithm commonly used for medical image classification tasks, which has strong adaptability and accuracy in processing medical images [10].

In addition, deep learning algorithms can be combined with techniques such as migration learning and data augmentation to further improve the generalisation ability and robustness of classification models [11]. By continuously optimising the algorithm structure and adjusting the hyperparameters, the deep learning model can be made to achieve more accurate and reliable image classification between different categories such as new coronary pneumonia, healthy people and community-acquired pneumonia [12].

It is of great significance to use deep learning algorithms to classify different types of X-ray images such as new crown pneumonia [13], healthy people and community-acquired pneumonia in the field of medical imaging [14]. In this paper, AlexNet, Vgg, GoogleNet and MobileNet models are introduced for comparative analysis to compare the effectiveness and accuracy of various deep learning models in pneumonia image classification. With the continuous progress of technology and the accumulation of data, we believe that deep learning algorithms will provide more effective auxiliary information for doctors in the future, and help make greater breakthroughs in clinical diagnosis [15].

2. Data set sources and data analysis

The dataset used in this paper is from the Kaggle open-source dataset (https://www.kaggle.com/datasets/pranavraikokte/covid19-image-dataset?rvi=1), which currently has 213 citations, and it contains about 137 X-ray images of COVID-19 images, in addition to 317 X-ray images containing community-acquired pneumonia and 118 normal chest X-ray images [16].The X-ray images of COVID-19 are shown in Fig. 1, the X-ray images of community-acquired pneumonia are shown in Fig. 2, and the normal chest X-ray images are shown in Fig. 3.

Figure 1. COVID-19.

Figure 2. Community-acquired pneumonia.

Figure 3. Normal chest.

3. Method

3.1. AlexNet

AlexNet is a deep convolutional neural network model proposed by Alex Krizhevsky, Ilya Sutskever, and Geoffrey Hinton in 2012 and is widely regarded as one of the major milestones in the field of deep learning. It achieved great success in the ImageNet image recognition challenge at that time and pushed deep learning technology to new heights [17]. The model structure of AlexNet is shown in Fig. 4.
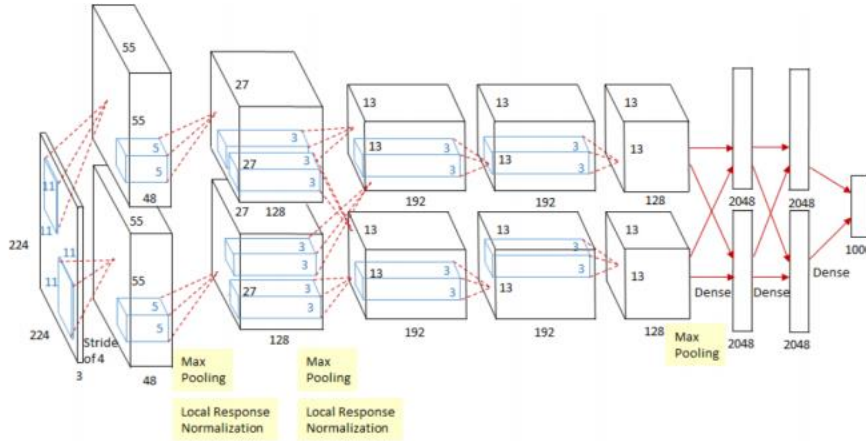

Figure 4. The model structure of AlexNet.

The structure of AlexNet consists of five convolutional layers and three fully connected layers. The first layer is a convolutional kernel of size 11x11, followed by a ReLU activation function and a maximum pooling layer. Subsequent convolutional layers follow a similar structure, extracting more abstract and complex features by continuously decreasing the size of the convolutional kernel and increasing the number of feature mappings. The fully connected layer is then responsible for classifying the features extracted by the convolutional layers. AlexNet introduces the Dropout technique, which prevents overfitting by randomly dropping some of the neurons during the training process, thus improving the generalisation of the model. In addition, AlexNet uses Data Enhancement technique to perform operations such as panning, flipping, and scaling on images in the training set in order to increase [18] diversity and thus improve the model performance.

3.2. Vgg

VGG is a deep convolutional neural network model proposed by the Computer Vision Group of the University of Oxford in 2014, whose main feature is the use of smaller convolutional kernels and deeper network structure [19]. The VGG model consists of 16 (VGG16) or 19 (VGG19) convolutional layers, which all use 3x3 sized convolutional kernels and convolutional operations with a step size of 1. Meanwhile every two convolutional layers are connected to a pooling layer for downsampling. The model structure of Vgg is shown in Fig. 5.
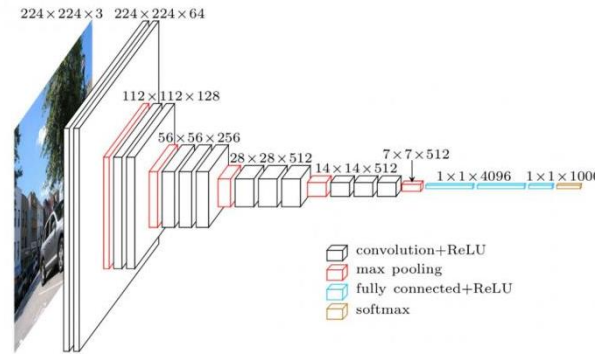
Figure 5. The model structure of Vgg.

The VGG model extracts features at different abstraction levels in an image by superimposing multiple convolutional layers, enabling the network to better understand the image content. Due to the use of smaller convolutional kernels and deeper network structure, the VGG model is able to reduce the number of parameters while maintaining a certain level of accuracy, which is conducive to training deeper neural networks [20].

During training, the VGG model usually uses optimisation methods such as cross-entropy loss function and stochastic gradient descent to continuously adjust the network parameters to improve the classification accuracy [21].

3.3. GoogleNetAlgorithm Optimisation
GoogLeNet is a deep convolutional neural network model proposed by the Google team in 2014, whose main feature is the use of the Inception module to extract features at different scales. Compared with the traditional approach of simply stacking convolutional layers, the Inception module is able to use different sized convolutional kernels at the same time to capture information at different scales in the image, thus improving the network's ability to understand the content of the image [22]. The model structure diagram of GoogLeNet is shown in Fig. 6.
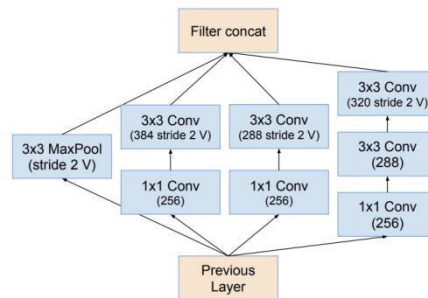

Figure 6. The model structure diagram of GoogLeNet.

GoogLeNet contains multiple Inception modules, each Inception module internally consists of multiple parallel convolutional layers including 1x1, 3x3, 5x5 sized convolutional kernels as well as a maximal pooling layer, which enriches the representation of features learned by the network by means of feature extraction at different scales and with different sensory fields.

3.4. MobileNet
MobileNet is a lightweight deep convolutional neural network model proposed by Google team in 2017, aiming to achieve efficient image recognition in resource-constrained environments such as mobile devices. MobileNet mainly uses deep separable convolution to reduce the number of parameters and computational complexity. The model principle of MobileNet is shown in Fig. 7.
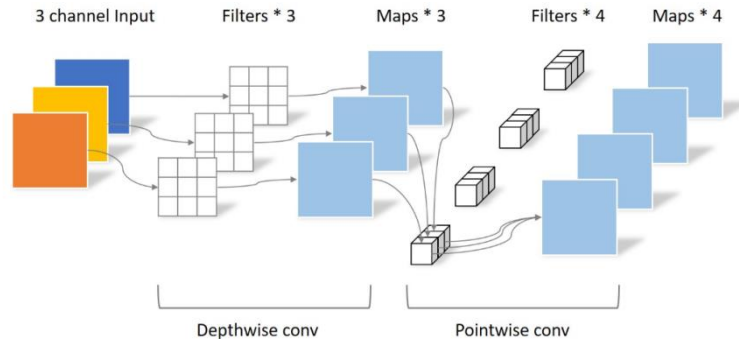
Figure 7. The model principle of MobileNet.

Deep separable convolution consists of two steps: deep convolution and point-by-point convolution. The deep convolution stage applies a separate convolution kernel to each input channel, while the point-by-point convolution stage uses a 1x1 convolution kernel to fuse the features of different channels together. This separation allows the network to drastically reduce the number of parameters while maintaining good performance. MobileNet also introduces a width multiplier and a resolution multiplier to further compress the model [23]. The width multiplier controls the number of channels in each layer of the network to reduce the complexity of the model, while the resolution multiplier adjusts the resolution of the input image to further reduce the amount of computation.

4. Result

In this paper, experiments are conducted using a standalone device with NVIDIA GeForce RTX 3090 graphics card for GPU, 32 GB of RAM, Intel Core i9 processor, and PyTorch, a deep learning framework, with Python version 3.9.

For the experimental parameter settings in this paper, Epoch is 50 rounds, Loss function is Cross Entropy Loss Function, Batch size is 32, Optimiser is Adam Optimiser, Initial Learning Rate is 0.001, Learning Rate Decay Strategy is used, Regularisation is L2 Regularisation Term, Regularisation Coefficient is 0.0001. The image size is uniformly adjusted to 224x224 pixels, and the data enhancement includes random horizontal flip, random rotation [24].

With the above experimental parameter settings, we will compare and analyse the effect of different models on the task of pneumonia X-ray image classification under the same conditions in order to evaluate their performance on this task [25]. The loss variation curves of each deep learning algorithm are output, as shown in Fig. 8.
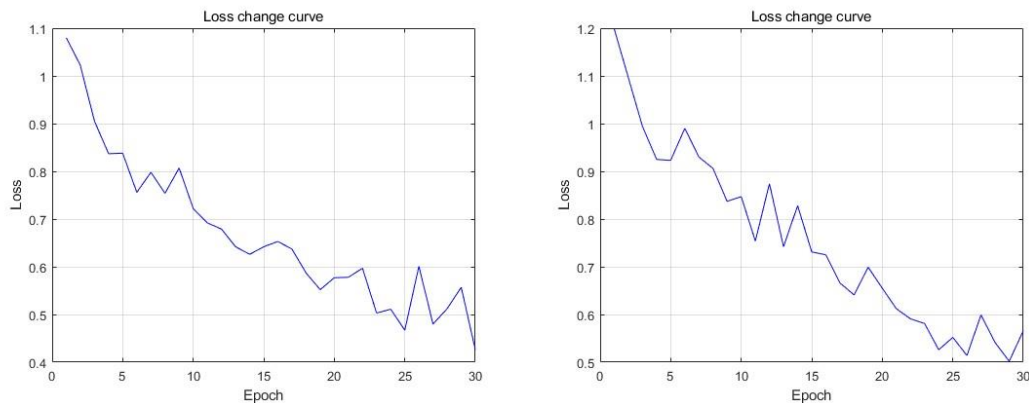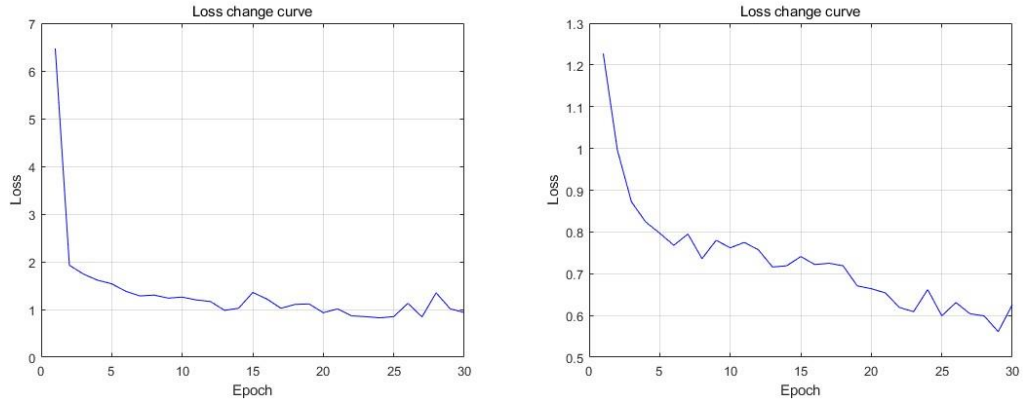
Figure 8. The loss variation curves of each deep learning algorithm.

As can be seen from Fig. 8, in terms of the convergence rate of the loss, GoogleNet reaches convergence first (3rd epoch), and the other three models gradually converge, and finally the loss of AlexNet is 0.426, the loss of Vgg is 0.566, the loss of GoogleNet is 0.936, and the loss of MobileNet is 0.626. The change curve of the validation set accuracy is shown in Fig. 9.
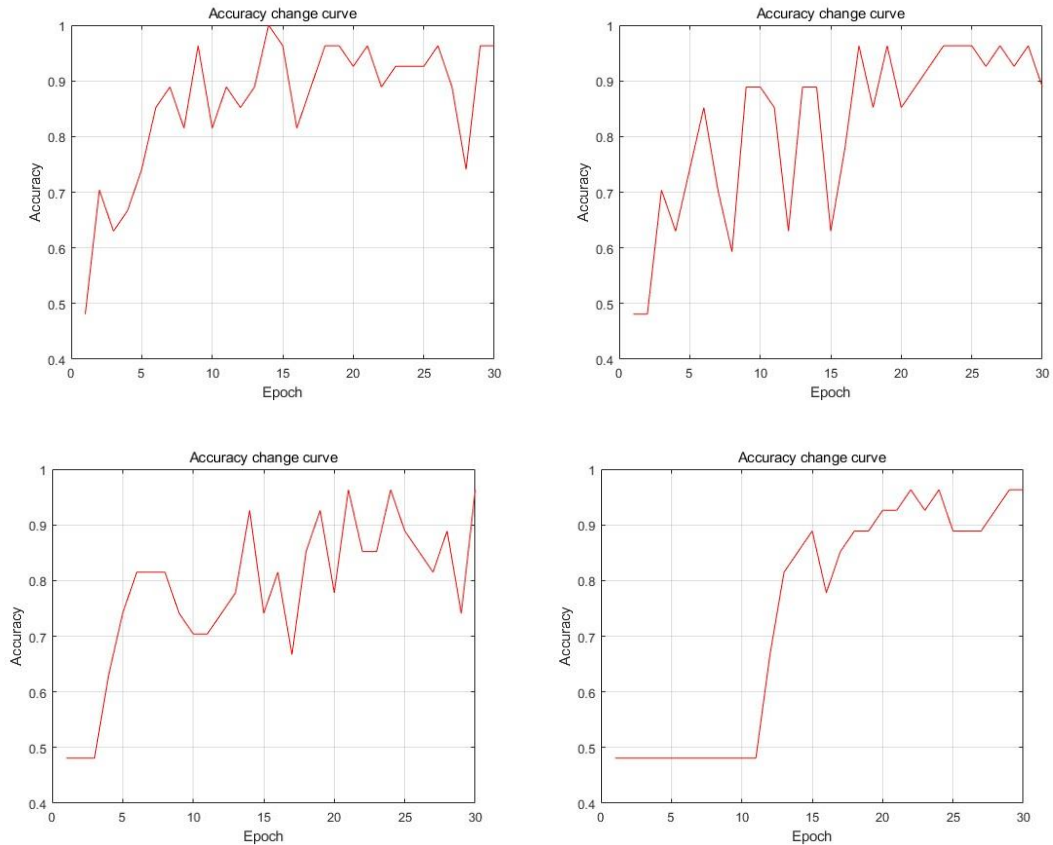

Figure 9. The change curve of the validation set accuracy.

The weights of the best trained round of the four models were selected for testing and the classification accuracy results were obtained: the Accuracy of AlexNet was 0.889, the Accuracy of Vgg was 92.6%, the Accuracy of GoogleNet was 85.2%, and the Accuracy of MobileNet was 96.3%. MobileNet showed the best prediction on the test set.

5. Conclusion

In this study, we introduced four deep learning models, AlexNet, Vgg, GoogleNet and MobileNet, to classify pneumonia images and compare their effectiveness and accuracy [26]. In terms of the convergence speed of the loss function, GoogleNet performs well and reaches the convergence state in only the 3rd epoch, while the other three models gradually stabilise [27]. The final loss values show 0.426 for AlexNet, 0.566 for Vgg, 0.936 for GoogleNet, and 0.626 for MobileNet. Further, after selecting the best performing round of training weights among the four models for testing, the classification accuracy results were obtained; AlexNet achieved 88.9% accuracy, Vgg 92.6%, GoogleNet 85.2%, and MobileNet up to 96.3% accuracy. This indicates that MobileNet shows the most superior prediction on the test set with more powerful classification ability [28]. Therefore, the comprehensive experimental results show that MobileNet performs well in the pneumonia image classification task [29], not only in terms of faster loss function convergence, but also far ahead of other models in terms of accuracy. Therefore, in practical applications, MobileNet can be considered as the preferred model for the pneumonia image classification task to obtain better classification performance and prediction results [30].

References

[1] Q. Zheng, C. Yu, J. Cao, Y. Xu, Q. Xing, and Y. Jin, "Advanced Payment Security System:XGBoost, CatBoost and SMOTE Integrated." 2024.

[2] Y. Weng and J. Wu, "Big data and machine learning in defence," *International Journal of Computer Science and Information Technology*, vol. 16, no. 2, 2024, doi: 10.5121/ijcsit.2024.16203.

[3] X. Li, W. Ye, M. Zou, F. Guo, and Z. Huang, "Analysis and research of retail customer consumption behavior based on support vector machine," in *2023 IEEE 3rd International Conference on Data Science and Computer Application (ICDSCA)*, 2023, pp. 713–718. doi: 10.1109/ICDSCA59871.2023.10392779.

[4] C. Zhou, Y. Zhao, S. Liu, Y. Zhao, X. Li, and C. Cheng, "Research on Driver Facial Fatigue Detection Based on Yolov8 Model." Jun. 2024. doi: 10.13140/RG.2.2.35207.20647.

[5] K. Mo, W. Liu, X. Xu, C. Yu, Y. Zou, and F. Xia, "Fine-Tuning Gemma-7B for Enhanced Sentiment Analysis of Financial News Headlines," *arXiv preprint arXiv:2406.13626*, 2024.

[6] C. Yu, Y. Xu, J. Cao, Y. Zhang, Y. Jin, and M. Zhu, "Credit Card Fraud Detection Using Advanced Transformer Model." 2024.

[7] C. Zhou, Y. Zhao, J. Cao, Y. Shen, X. Cui, and C. Cheng, "Optimizing Search Advertising Strategies: Integrating Reinforcement Learning with Generalized Second-Price Auctions for Enhanced Ad Ranking and Bidding," *arXiv reprint arXiv:2405.13381*. 2024.

[8] T. Liu, S. Li, Y. Dong, Y. Mo, and S. He, "Spam Detection and Classification Based on DistilBERT Deep Learning Algorithm," *Applied Science & Engineering Journal for Advanced Research*, vol. 3, no. 3, May 2024, doi: 10.5281/zenodo.11180575.

[9] J. Yuan, L. Wu, Y. Gong, Z. Yu, Z. Liu, and S. He, "Research on Intelligent Aided Diagnosis System of Medical Image Based on Computer Deep Learning," *arXiv preprint arXiv:2404.18419*. 2024.

[10] H. Liu, Y. Shen, W. Zhou, Y. Zou, C. Zhou, and S. He, "Adaptive speed planning for Unmanned Vehicle Based on Deep Reinforcement Learning," *arXiv preprint arXiv:2404.17379*. 2024.

[11] C. Jin, T. Che, H. Peng, Y. Li, and M. Pavone, "Learning from teaching regularization: Generalizable correlations should be easy to imitate," *arXiv preprint arXiv:2402.02769*, 2024.

[12] Z. Jiang *et al.*, "Weakly supervised spatial deep learning for earth image segmentation based on imperfect polyline labels," *ACM Transactions on Intelligent Systems and Technology (TIST)*, vol. 13, no. 2, pp. 1–20, 2022.

[13] J. Cao, Yanhui, Jiang, C. Yu, F. Qin, and Z. Jiang, "Rough Set improved Therapy-Based Metaverse Assisting System." 2024.

[14] X. Huang, Z. Zhang, F. Guo, X. Wang, K. Chi, and K. Wu, "Research on Older Adults' Interaction with E-Health Interface Based on Explainable Artificial Intelligence," *arXiv preprint arXiv:2402.07915*. 2024.

[15] X. Shen, Q. Zhang, H. Zheng, and W. Qi, "Harnessing XGBoost for Robust Biomarker Selection of Obsessive-Compulsive Disorder (OCD) from Adolescent Brain Cognitive Development (ABCD) data," *ResearchGate*, May 2024, doi: 10.13140/RG.2.2.12017.70242.

[16] J. Xu *et al.*, "Optimization of Worker Scheduling at Logistics Depots Using Genetic Algorithms and Simulated Annealing," *arXiv preprint arXiv:2405.11729*. 2024.

[17] J. Wang, X. Li, Y. Jin, Y. Zhong, K. Zhang, and C. Zhou, "Research on Image Recognition Technology Based on Multimodal Deep Learning," *arXiv preprint arXiv:2405.03091*. 2024.

[18] Y. Weng and J. Wu, "Fortifying the global data fortress: a multidimensional examination of cyber security indexes and data protection measures across 193 nations," *International Journal of Frontiers in Engineering Technology*, vol. 6, no. 2, 2024, doi: 10.25236/ijfet.2024.060206.

[19] C. Zhou *et al.*, "Predict Click-Through Rates with Deep Interest Network Model in E-commerce Advertising," *arXiv preprint arXiv:2406.10239*. 2024.

[20] C. Jin *et al.*, "Visual prompting upgrades neural network sparsification: A data-model perspective," *arXiv preprint arXiv:2312.01397*, 2023.

[21] F. Guo, "A Study of Smart Grid Program Optimization Based on K-Mean Algorithm," in *2023 3rd International Conference on Electrical Engineering and Mechatronics Technology (ICEEMT)*, 2023, pp. 711–714. doi: 10.1109/ICEEMT59522.2023.10263168.

[22] H. Guan and M. Liu, "Domain Adaptation for Medical Image Analysis: A Survey," *IEEE Trans Biomed Eng*, vol. 69, no. 3, pp. 1173–1185, Mar. 2022, doi: 10.1109/tbme.2021.3117407.

[23] W. He *et al.*, "Quantifying and reducing registration uncertainty of spatial vector labels on earth imagery," in *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 2022, pp. 554–564.

[24] Y. Chen, W. Huang, X. Liu, Q. Chen, and Z. Xiong, "Learning multiscale consistency for self-supervised electron microscopy instance segmentation," in *IEEE International Conference on Acoustics, Speech, and Signal Processing*, 2023.

[25] Z. Zhang, P. Li, A. Y. Al Hammadi, F. Guo, E. Damiani, and C. Y. Yeun, "Reputation-Based Federated Learning Defense to Mitigate Threats in EEG Signal Classification," *arXiv preprint arXiv:2401.01896*. 2023.

[26] Y. Cheng, Q. Yang, L. Wang, A. Xiang, and J. Zhang, "Research on Credit Risk Early Warning Model of Commercial Banks Based on Neural Network Algorithm," *arXiv preprint arXiv:2405.10762*. 2024.

[27] W. He, A. M. Sainju, Z. Jiang, D. Yan, and Y. Zhou, "Earth Imagery Segmentation on Terrain Surface with Limited Training Labels: A Semi-supervised Approach based on Physics-Guided Graph Co-Training," *ACM Transactions on Intelligent Systems and Technology (TIST)*, vol. 13, no. 2, pp. 1–22, 2022.

[28] T. Deng, H. Xie, J. Wang, and W. Chen, "Long-Term Visual Simultaneous Localization and Mapping: Using a Bayesian Persistence Filter-Based Global Map Prediction," *IEEE Robot Autom Mag*, vol. 30, no. 1, pp. 36–49, 2023.

[29] Y. Chen, C. Liu, W. Huang, S. Cheng, R. Arcucci, and Z. Xiong, "Generative Text-Guided 3D Vision-Language Pretraining for Unified Medical Image Segmentation," *arXiv preprint arXiv:2306.04811*, 2023.

[30] W. Xu, J. Chen, Z. Ding, and J. Wang, "Text Sentiment Analysis and Classification Based on Bidirectional Gated Recurrent Units (GRUs) Model," *arXiv preprint arXiv:2404.17123*. 2024.