Comparison of Supervised Learning and Unsupervised Learning in the Diagnosis of Medical Imaging Diseases

Xinhai Yang

Shanghai Weiyu High School, Shanghai, China yxh140820@163.com

Abstract: Medical image registration represents a pivotal element in the field of disease image analysis, acting as the essential precursor for a multitude of sophisticated analytical tasks. In recent years, traditional methodologies have encountered significant challenges in meeting the evolving demands of clinical practice. In contrast, deep learning-based strategies have emerged as powerful alternatives, showcasing their ability to facilitate more rapid and accurate registration processes, thereby exerting a profound impact on clinical applications. Within the specialized domain of medical imaging, the intricate level of expertise required for domain knowledge imposes rigorous standards on annotators, which in turn leads to increased annotation costs. As a result, the efficacy of supervised learning approaches compared to unsupervised learning methodologies can exhibit substantial variability in real-world applications. This paper systematically utilizes a diverse array of medical imaging datasets to rigorously assess the performance outcomes of both supervised and unsupervised learning techniques, specifically in relation to their practical applications in the medical imaging landscape.

Keywords: Machine learning, medical image processing, supervised learning, unsupervised learning

1. Introduction

In recent years, with the development of artificial intelligence technology represented by deep neural networks, deep learning has been widely used in various fields. From satellite remote sensing image detection to crop pest and disease identification, deep learning technology has spread to all corners of national defense and people's livelihood. In the field of medical imaging, doctors or researchers often need to know some details of a specific internal tissue and organ in order to make the right treatment decision when performing quantitative analysis, real-time monitoring, and treatment planning. Therefore, medical image processing has become an indispensable part of disease diagnosis and treatment, and it is becoming increasingly important. However, China's current doctor-patient ratio is only 1:950, ranking 96th in the world. Among them, doctors in tertiary hospitals have an average of 6.7 visits per day, and this data has risen to 9.1 in grassroots hospitals [1], compared with the average daily diagnosis and treatment of about 5 doctors in Europe and the United States, which shows that China's doctor resources are in a scarce state. How to apply deep learning technology to the field of medical imaging, improve the efficiency of diagnosis and treatment, and relieve the pressure of doctors' diagnosis and treatment has attracted more and more attention.

Traditional methods are inadequate for clinical needs; deep learning techniques offer faster, more precise registration, significantly influencing clinical practice. In medical imaging, the specialized knowledge required imposes high costs for annotation. Consequently, the effectiveness of supervised versus unsupervised learning methods can differ markedly in application. This paper will analyze results from both approaches across multiple datasets, highlighting their respective advantages and limitations, necessitating tailored methods for varying scenarios.

2. Theoretical Basis for Supervised and Unsupervised Learning

2.1. Overview of Machine Learning

Neural networks have achieved significant success across various research domains due to their robust feature learning capabilities, including speech and image recognition, image segmentation, and natural language processing. The deep learning implementation process involves feature extraction from data to develop a model capable of generalizing to unknown data. Deep learning is categorized into supervised learning, which addresses specific tasks through training data pairs, and unsupervised learning, which leverages unlabeled data to learn prior information for diverse tasks. While supervised learning offers stable performance for specific tasks, it lacks robustness. In contrast, unsupervised learning capitalizes on the data distribution of unlabeled images, eliminating the need for cumbersome data labeling, which can compromise model accuracy.

2.2. Supervised Learning Theory

Well-known algorithms for supervised learning include support vector machines (SVMs), decision trees, random forests, and convolutional neural networks (CNNs) and recurrent neural networks (RNNs) in the form of deep learning. There are two main algorithms commonly used in medical images:

A support vector machine (SVM) is a classifier that identifies distinct classes by determining the optimal hyperplane in feature space, maximizing the margin between categories defined by the nearest support vectors. SVMs excel with small to medium datasets, particularly when category boundaries are clear, and effectively address linearly separable and non-separable problems using kernel methods. They are widely applied in medicine for categorical assessments, such as diagnosing osteoporosis.

Convolutional Neural Networks (CNNs) are prominent deep learning algorithms known for their self-learning and intelligent processing capabilities. Comprising an input layer, convolutional layer, activation function layer, pooling layer, and fully connected layer, CNNs process original image data to extract features, reduce computational load, and adapt to image dislocation. They are crucial in medical imaging interpretation and serve as a leading recognition method in computer vision, facilitating tasks such as image recognition, classification, segmentation, labeling, and generation through extensive data training.

2.3. Unsupervised Learning Theory

The biggest feature of the unsupervised learning method [2-4] is that the trained model can be used as general prior information in image reconstruction under different imaging conditions. At the same time, the separation of prior terms and data fidelity terms in the reconstruction of the algorithm eliminates the data sensitivity caused by data acquisition bias. From the perspective of learning data distribution, unsupervised learning methods vary. In general, the following unsupervised algorithms are available in medical images: autoencoder (AE) learning [5, 6] (e.g., denoising autoencoding (DAE)

[5] and variational autoencoding (VAE) [6], generative adversarial networks (GANs) [7], SimCRL, etc.

SimCLR (Simple Contrastive Learning of Representations) is a self-supervised learning method used to learn useful feature representations from unlabeled data. The core idea of SimCLR is to learn by contrastively comparing the relative relationships between positive samples (similar image pairs) and negative samples (dissimilar image pairs) [8].

The learning of an autoencoder (AE) simply retains the information of the original input data and does not ensure a useful representation of the features. This is because the autoencoder may simply copy the original input, or simply pick features that slightly alter the reconstruction error, but do not contain particularly useful information. In order to avoid the above situation and to be able to learn better feature representations, it is necessary to give certain constraints to the data representation. Denoising autoencoders can solve this problem by reconstructing input data that contains noise [9].

The function of the denoising autoencoder is to learn the raw data of the superimposed noise, and the features it learns are almost the same as those learned from the data of the unsuperimposed noise, but the denoising autoencoder learns the features from the input of the superimposed noise is more robust, and can avoid the above problems encountered by the autoencoder and simply learn the same eigenvalues.

A generative adversarial network (GAN) comprises a generative model and a discriminative model. The generative model captures the sample data distribution, while the discriminative model acts as a binary classifier to distinguish real data from generated samples. The optimization process is a "binary minima game," where one model's parameters are fixed while the other's are updated alternately, enabling the generative model to estimate the data distribution. GANs have significantly advanced unsupervised learning and image generation, expanding from initial applications in image generation to various computer vision domains, including image segmentation, video prediction, and style transfer.[10]

3. The Application of Supervised Learning and Unsupervised Learning in the Diagnosis of Medical Imaging Diseases

3.1. Case Overview

Diabetic retinopathy is one of the ocular complications due to diabetes that involves diseases and abnormal changes in the retina. Diabetic retinopathy can be divided into two stages: (1) Early non-proliferative diabetic retinopathy (NPDR). (2) Proliferative diabetic retinopathy (PDR). If diabetic retinopathy worsens, it can lead to retinal ischemia, which stimulates the growth of new blood vessels. Early diagnosis and treatment can slow the progression of diabetic retinopathy and reduce the risk of blindness [11].

3.2. Supervised Learning Cases

In the classification criteria for diabetic retinopathy, the division of fundus image quadrants requires the network to have the knowledge of the location of the image elements, and if the network model wants to have a better disease classification ability, the network model needs to have the ability to process location information. In order to solve the problem, Hinton et al. [12] proposed a new type of convolutional neural network, capsule network, to solve the above problems. The structure of the capsule network is precisely to enable the neural network to correctly recognize the relationship between the parts in the image, and it advocates to preserving the relative relationship between the parts of the image and carrying out further processing. This method achieves excellent results in identifying the same object at different angles. Therefore, on this basis, this chapter proposes a capsule network-based diagnostic model for diabetic retinopathy DRCaps, and achieves a binary accuracy of 0.9658 and a Penta classification accuracy of 0.8183 on the Kaggle APTOS 2019 dataset. The methodology proposed in this chapter is analyzed in detail below. The overall capsule network model consists of a feature extraction module, a capsule module, and a decoder module.

The Feature Extraction Module is the part of DRCaps that completes the feature extraction function, which is composed of a frequency domain attention module and a convolutional layer module for extracting features, which makes the model pay more attention to the lesion part by applying the attention mechanism to the feature map, and the residual module with a jump connection ensures the depth of the network, and the features extracted in the front part of the network can be transmitted to the rear network.

The Capsule Layer consists of the Primary Capsule Layer and the Digital Capsule Layer, which encodes the feature map for the digital capsule layer, converts it into a vector, and inputs the digital capsule layer for classification tasks.

The decoder part is composed of a deconvolution layer, which forms an encoder-decoder structure with the front feature extraction network, and the decoder part tries to reconstruct the feature map and make the image as close to the original image as possible.

This paper compares its method with classical network structures and existing literature. The comparison is divided into two parts: the first part enhances the persuasiveness of network data through experiments on various parameters across multiple architectures, ultimately selecting the best results for comparison. The second part evaluates experimental results from prior studies. Table 3.9 presents classification task results from three models on Kaggle APTOS 2019: VGG16 at 0.7271, Inception-V3 at 0.7217, and ResNet50 at 0.7135, with this paper achieving 0.8183. Additionally, the dichotomous classification results show VGG16 at 0.929, Inception-V3 at 0.935, and ResNet50 at 0.845, while this paper's method outperforms with 0.9658.

3.3. Unsupervised Learning Cases

Machine learning for diabetic retinopathy leverages computational and statistical techniques to analyze medical images, aiding in the diagnosis and treatment of diabetes-related ocular complications. However, annotating fundus images is a labor-intensive and costly endeavor, necessitating expert input and often resulting in subjective inconsistencies. Self-supervised learning, an unsupervised approach, addresses this by learning from unlabeled data, thereby alleviating the annotation burden on clinicians and reducing associated costs. Given the scarcity of labeled diabetic retinopathy data, self-supervised learning can utilize unlabeled datasets to enhance training sample availability. This method mitigates the subjective biases inherent in manual labeling, as it typically operates independently of human-generated labels. In this context, the SImCRL model exemplifies the application of self-supervised learning.

Data augmentation

Data augmentation augments the training dataset by applying various random transformations to the input image to generate multiple different perspectives, which helps the model learn more robust and useful feature representations. The model randomly selects a sub-region of the fundus image, crops a fixed-size image from it, randomly adjusts the fundus image to different sizes, and fills in the blank parts. This helps the model learn features at different sizes. This is followed by a random horizontal flip to flip the image horizontally with a certain probability. Finally, the color dithering process is performed with a certain probability to randomly adjust the brightness, contrast, saturation and hue of the image. This helps the model learn features under different lighting conditions.

Embedding coding

The embedding encoding part is the process of converting a data-enhanced image into a feature vector. This is usually done through a neural network. The goal of embedding encoding is to map the image into a low-dimensional feature space for subsequent contrast loss calculations.

Comparison loss

Contrast loss is the core component used to train the model so that fundus image features that have undergone data augmentation and embedding coding are better clustered together in the feature space. Contrast loss is achieved by maximizing the similarity between positive samples (similar image pairs) and minimizing the similarity between negative samples (dissimilar image pairs).

Experimental process

The Kaggle APTOS dataset is a retinopathy detection dataset on the Kaggle platform, which is called "Diabetic Retinopathy Detection". This dataset contains a series of fundus images for the detection of diabetic retinopathy. The dataset includes 3662 training set images and 1928 test set images, each with a resolution of 2588×1958 pixels, divided into five levels of 0~4.

The accuracy of the results compared to three different supervised learning methods on Kaggle APTOS 2019. The accuracy of the proposed method on the five classification tasks of the Kaggle APTOS dataset is 0.582 and 0.831 on the two classification tasks, indicating that the method has achieved good results on multiple datasets.

4. Comparison of Supervised and Unsupervised Learning

In the task of medical image disease diagnosis, supervised learning and unsupervised learning have their own advantages and limitations, and there are significant differences in their application scenarios and performance.

4.1. Data Dependency and Annotation Costs

Supervised learning necessitates substantial high-quality labeled datasets, with model performance closely tied to data size and quality. For instance, the DRCaps model attained notable accuracy in diabetic retinopathy classification (5-category 0.8183, binary classification 0.9658) due to accurately annotated fundus images from the Kaggle APTOS dataset. However, medical image annotation demands professional involvement, making it time-intensive and costly. Conversely, unsupervised learning employs self-supervised or generative techniques (e.g., SimCRL, GAN) to derive features from unlabeled data, significantly mitigating reliance on manual annotation. After training a SimCRL-based model on unlabeled data, the accuracy for the five-class task is 0.582 and 0.831 for the two-class task, which, while lower than supervised learning outcomes, effectively addresses labeling resource shortages. Furthermore, data augmentation methods (e.g., random cropping, color jitter) enhance the robustness of unsupervised models against data noise and imaging condition variability.

4.2. Model Performance and Task Suitability

Supervised learning is particularly effective for targeted classification and segmentation tasks, such as diabetic retinopathy grading. Models like CNNs and capsule networks can accurately capture lesion features through end-to-end training, though they risk overfitting due to labeling bias or limited data distribution. Traditional CNNs may lose spatial information from pooling, whereas capsule networks preserve spatial relationships via dynamic routing, enhancing classification accuracy. In contrast, unsupervised learning excels in generating feature representations, making it suitable for complex data distributions and multitasking.

4.3. Generalization Ability and Robustness

Supervised learning models are sensitive to labeled data distribution, leading to reduced generalization when training data differs from real-world scenarios (e.g., images from various

devices). For instance, ResNet50's accuracy on the Kaggle dataset (0.7135) is notably lower than that of capsule networks, highlighting its vulnerability to data variability. In contrast, unsupervised learning captures broader feature distributions from unlabeled data, demonstrating greater adaptability to variations in imaging conditions (e.g., lighting and noise).

5. Conclusion

Supervised learning is particularly well-suited for annotation tasks where ample resources and welldefined objectives are available, such as in specialized disease screening within tertiary care facilities. Its high accuracy and interpretability align with the stringent reliability standards required for clinical diagnostics. Conversely, unsupervised learning proves beneficial in scenarios characterized by resource scarcity or the necessity for rapid adaptation to novel tasks, exemplified by initial screenings or epidemiological investigations in primary healthcare settings. Furthermore, generative models, including Generative Adversarial Networks (GANs), can facilitate the synthesis of medical images, enhance data augmentation, and support physician training, thereby broadening their applicability. In conclusion, the selection between these methodologies must consider data availability, task specifications, and budgetary limitations in real-world applications. Future investigations may delve into semi-supervised or hybrid learning paradigms, merging the precision of supervised approaches with the adaptability of unsupervised techniques, to foster the comprehensive advancement of medical imaging diagnostic technologies.

References

- [1] Fang, B. T. (2023). The release of the 2022 Statistical Bulletin on the Development of Health and Health Undertakings in China. Journal of Traditional Chinese Medicine Management, 31(19), 116.
- [2] Bengio, Y., & LeCun, Y. (2007). Scaling learning algorithms towards AI. In Large-scale kernel machines (pp. 1-41).
- [3] Erhan, D., Courville, A., Bengio, Y., et al. (2010). Why does unsupervised pre-training help deep learning? In Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics (pp. 201-208). JMLR Workshop and Conference Proceedings.
- [4] Bengio, Y., Courville, A. C., & Vincent, P. (2012). Unsupervised feature learning and deep learning: A review and new perspectives. CoRR, abs/1206.5538, 1.
- [5] Vincent, P., Larochelle, H., Bengio, Y., et al. (2008). Extracting and composing robust features with denoising autoencoders. In Proceedings of the 25th International Conference on Machine Learning (pp. 1096-1103).
- [6] Rifai, S., Vincent, P., Muller, X., et al. (2011). Contractive auto-encoders: Explicit invariance during feature extraction. ICML.
- [7] Radford, A., Metz, L., & Chintala, S. (2015). Unsupervised representation learning with deep convolutional generative adversarial networks. arXiv preprint arXiv:1511.06434.
- [8] Van Oord, A., Kalchbrenner, N., & Kavukcuoglu, K. (2016). Pixel recurrent neural networks. In International Conference on Machine Learning (pp. 1747-1756). PMLR.
- [9] Kingma, D. P., & Dhariwal, P. (2018). Glow: Generative flow with invertible 1x1 convolutions. arXiv preprint arXiv:1807.03039.
- [10] Li, B. Y., & Li, Y. (2019). Application of supervised learning ultrasound backscattering method in bone quality evaluation.
- [11] Liang, F., Hu, D. Y., & Shen, Z. J. (2014). 2014 American Diabetes Association guidelines: Standards of medical care in diabetes. Chinese Journal of Clinical Physicians: Electronic Edition, (6).
- [12] Sabour, S., Frosst, N., & Hinton, G. E. (2017). Dynamic routing between capsules. Advances in Neural Information Processing Systems, 30.