Exploring the Application of Deep Learning in Lung Cancer Prediction

Jie Cheng

Tianhua College, Shanghai Normal University, shanghai, 201800, China
100246@yzpc.edu.cn

Abstract. Deep learning is a branch of artificial intelligence. Deep learning can make learning predictions by simulating the working principles between neurons in the human brain. Lung cancer is a major problem in the medical field, and the prediction of lung cancer based on deep learning is of great significance. Therefore, this article will discuss and study the current status of lung cancer and how deep learning predicts it. This article concludes that deep learning is more effective than other methods for complex diseases such as lung cancer. Among them, convolutional neural networks can have higher accuracy. The emergence of deep learning makes complex medical content simple to quantify. Make it easier for doctors to judge treatment.The prediction and research of cancer through deep learning can improve the treatment effect and benefit mankind. This article explores the help and contribution of deep learning in the field of lung cancer through the characteristics of deep learning.

Keywords: Artificial intelligence, Lung cancer prediction, Deep learning, Convolutional neural network.

1 Introduction

Lung cancer has become a major cancer disease that threatens human life and health. The incidence of lung cancer has ranked first among all cancers. At the same time, the mortality rate of lung cancer is high, which seriously endangers human health. At present, there are 1.3 million new lung cancer patients and about 1.2 million deaths from lung cancer every year in the world, and the proportion of lung cancer in young and middle-aged people is increasing. China is the country with the largest number of lung cancer patients, with about 400,000 people diagnosed with lung cancer every year, and the incidence rate is as high as 61.4/100,000 [1]. In the medical field, based on years of investigation and research, lung cancer has diverse pathogenic factors, diverse characteristics of disease stages, diverse cases of cancer patients, diverse forms of cancer cell mutations, and diverse treatment methods for different patients. Today, with the changes in information technology, the birth of artificial intelligence technology has enabled all walks of life to make more accurate and intelligent judgments and analyses, and the same is true in the field of medical cancer research.

Deep learning is a specialized field in machine learning. Compared with machine learning, deep learning is more "intelligent". Machine learning can autonomously learn

Y. Wang (ed.), *Proceedings of the 2024 2nd International Conference on Image, Algorithms and Artificial Intelligence (ICIAAI 2024)*, Advances in Computer Science Research 115,
https://doi.org/10.2991/978-94-6463-540-9_13

the features of things that need to be learned through multi-layer deep neural networks and then combine them with low-level features to become more abstract high-level features that describe attribute categories. Deep learning's autonomous learning of the features of things can greatly reduce the cost of manual analysis. It can quickly learn to match in such complex and small matters, identify the features of different cells, and autonomously match mutually suitable models, and enhance knowledge through learning. Deep learning can develop and infer more information about cancer diseases. Reduce the large amount of consumption and loss of manpower and material resources, improve prediction accuracy, and make the prediction of lung cancer more accurate and detailed.

2 Deep Learning

In human history's development, three crucial historical nodes have pushed human life onto a more efficient path. When the information technology of the third industrial revolution made a disruptive breakthrough, with the continuous development of the information age, mankind once again faced new challenges and breakthroughs. The driving force behind this scientific and technological revolution is artificial intelligence. 2006 is regarded as the first year of deep learning, when HINTON proposed a solution to the gradient vanishing problem in deep network training. First, the structure of the training data is learned by autonomous learning, and then supervised training is performed on this structure. Compared with manually designed features or shallow feature representations, multi-layer convolutional structures are one of the most advanced and attractive structures in the current computer vision field. Different from traditional Classification And Detection (CAD) systems, deep learning is a new field of machine learning. It has achieved many breakthroughs in image recognition, speech recognition and other fields, and has become a recognized revolutionary technology in the field of artificial intelligence [2]. Deep learning is the process of learning from large-scale data based on deep neural networks[3,4].

2.1 Current Status and Characteristics of Deep Learning

There are many algorithms in deep learning, and different algorithms are suitable for different scenarios. Among them, neural networks, recurrent neural networks, and convolutional neural networks are algorithms often used in machine learning.

As the algorithm basis of deep learning, the principle of neural networks is inspired by the working principle of human brain nerves, imitating the connection between neurons to learn, calculate, and predict through signal transmission between layers of neurons. Based on this feature, it can be seen that neural networks are very suitable for calculating and predicting complex nonlinear models.

Recurrent neural network is referred to as " Recurrent Neural Network (RNN)". This algorithm is suitable for neural networks with similar sequence data. It has a strong memory function and can keenly obtain the dependency between sequence data. This algorithm can be used to process natural language, time series prediction, etc.

A convolutional neural network is referred to as "Convolutional Neural Networks (CNN)". This algorithm is a neural network structure used to process images and videos. It extracts features through convolutional layers and pooling layers for analysis and prediction.

2.2 Advantages of deep learning in disease prediction

In the treatment of cancer, the analysis of medical images is an essential part. Medical personnel can understand, predict and give appropriate treatment plans by analyzing the patient's medical images. However, this manual judgment method often increases the workload of medical personnel, and if there is a mistake in the judgment and prediction of diseases such as cancer, it may be a cause of error. In recent years, with the development of information technology, medical image analysis has also entered the era of information analysis. Information technology can assist medical personnel in conducting rigorous analyses and judgments and calculating helpful patient information from a large amount of data. Therefore, the status of artificial intelligence-assisted medical analysis and prediction is becoming increasingly important. CAD based on machine learning and image processing technology has gradually become a research hotspot in the medical field. However, CAD requires manual analysis to extract the model and then perform fitting and comparison. In the medical field, there are often a large number of images that need to be analyzed, calculated and predicted. Traditional CAD has never been able to surpass the recognition accuracy of humans for similar diseases, and the calculation speed is very slow, which makes it difficult to provide assistance for busy clinical work. This method is not well applicable [5].

In 2006, Hinton's team published a report on deep learning. They concluded that, unlike previous machine learning methods, deep learning is a hierarchical model inspired by the simulation of human brain neurons, which can extract, analyze, predict, and learn massive amounts of data, and find the connection between data. More importantly, deep learning can extract from pixel-level raw data step by step. Its novel features provide important help in the field of medical analysis. Compared with CAD, deep learning helps improve the ability to make auxiliary diagnoses[6].

Deep Learning (DL) is a branch of machine learning that emphasizes learning from a series of continuous representation layers. DL can train neural networks on massive amounts of medical image data abstract features to quantify medical data, analyze and study diseases, and then predict them. As a key feature in DL in machine learning, CNN can analyze data features from convolutional structure data and automatically learn to extract predictions. Machine learning can process massive amounts of medical data and improve the accuracy and precision of medical predictions; automatic feature extraction can accurately obtain features and relationships between complex medical data; high flexibility can be applied to the prediction and judgment of different diseases to improve efficiency. The high accuracy of machine learning makes medical predictions and judgments more accurate.

3 Application of deep learning in lung cancer prediction

3.1 Characteristics and influencing factors of lung cancer

Among the diseases nowadays, cancer is still a major disease area that is still being conquered in the medical field. Lung cancer, as the most common cancer disease with a high incidence and mortality rate, has always been widely concerned. In recent years, the incidence and mortality of lung cancer in China have risen sharply, becoming a major disease problem that needs to be solved urgently. According to relevant research, it is estimated that 2.1 million new cases of lung cancer will occur in 2018 (accounting for 11.6% of all new tumors) and 1.8 million deaths (18.4%). Lung cancer is common in elderly patients over 70 years old, while the incidence of lung cancer in young patients under 40 years old is relatively low (1% to 10%). Studies from Europe and Japan show that the incidence of lung cancer in young and middle-aged people has an upward trend. Early studies have shown that the malignancy rate and mortality rate of lung cancer patients under 40 years old are higher than those of elderly patients. For patients with advanced lung cancer, the prognosis is also poor, which may be since young people have better compensatory ability of physical functions and less obvious clinical symptoms, so the prognosis is also poor. Arnold et al. found that the median survival time of patients with lung cancer stages I to IV under the age of 46 was 60.0, 59.1, 18.4, and 8.8 months, respectively [7]. This shows that the treatment and intervention of lung cancer is already urgent.

There are many pathological characteristics of lung cancer, such as gender, age, genetic history of the disease, smoking history, work environment, living habits, cancer tumor type, tumor location, medication history, etc. Because of such complex factors, different patients have different characteristics. If machine learning is used, it can well extract and analyze the disease characteristics of different patients, conduct targeted analysis and prediction, and specify appropriate treatment plans.

3.2 Characteristics and influencing factors of lung cancer

For the treatment of lung diseases, taking Computed Tomography (CT) images is an indispensable means of treating and screening lung diseases. CT images often have the characteristics of high image information integration and many lung nodules. Relying solely on clinicians for analysis and judgment will greatly impact the accuracy and treatment efficiency of the disease. If artificial intelligence technology is introduced in this direction, efficiency can be greatly improved. The corresponding neural network model can be analyzed and predicted by using CNN to analyze CT images of lung cancer patients. The nature of lung cancer tumors and the stage of the tumor can be obtained through the CNN algorithm, and the patient's condition can be accurately judged to provide a basis for subsequent treatment.

In addition, artificial intelligence can use various machine learning and deep learning algorithms to learn and quantify many disease characteristics autonomously; the process of quantifying disease characteristics is radiomics. With the widespread and in-depth application of deep learning in clinical medicine, the prediction of lung cancer

treatment is based not only on the patient's disease characteristics and the tumor status presented by CT images but also on individual mutation maps. Through the CNN algorithm, a large number of scientific eyes can be obtained with high-throughput quantitative image features, which can be widely used in the field of cancer gene mutation. Epidermal Growth Factor Recepto (EGFR) and Kirsten Rat Sarcoma viral oncogene homolog (KRAS) mutation status is crucial for the targeted treatment of Non-Small Cell Lung Cancer (NSCLC) patients because they are key genes for understanding lung cancer pathogenicity before targeted treatment. In particular, Epidermal Growth Factor Receptor Tyrosine Kinase Inhibitors (EGFR-TKIS) inhibitors have improved survival for patients with sensitive mutations. Some studies have reported the research progress of using radiomics to identify EGFR and KRAS mutations[8].

3.3 Deep learning for lung cancer disease prediction

Although traditional machine learning algorithms such as k－Nearest Neighbor (KNN) and random forest can effectively analyze lung cancer mutations, they still require many clinicians to interpret and find models suitable for the algorithm, often requiring a lot of labor costs. Although breakthrough progress has been made, it has not fundamentally solved the problem of lung cancer disease prediction. Deep learning algorithms can effectively improve this problem.

The CNN convolutional neural network in the deep learning algorithm can have more powerful feature extraction and classification analysis in lung cancer disease prediction because of its powerful neural network characteristics. The relevant data record that the accuracy of the CNN model in identifying EGFR mutation status is 71.5%, while the recognition accuracy of the fusion model of CNN and clinical characteristics (gender, smoking history) can reach 77.2%. This shows that deep learning is more suitable for analyzing the cancer field than other machine learning algorithms for lung cancer outcome prediction [5].

The prediction of cancer gene mutations by the convolutional neural network CNN algorithm is as high as 75%. This shows that the powerful neural network through deep learning can be more suitable for lung cancer patients and can reduce the workload of clinicians; at the same time, it is more applicable than other machine learning algorithms.

4 Recommendations for Deep Learning in Cancer Prediction

The era of artificial intelligence has begun, and deep learning plays a very prominent role in the medical field. It solves many complex, tedious and error-prone tasks, allows the work to be accurately quantified, and objectively makes a better and more scientific plan for patients. However, deep learning still has many problems and challenges[9].

Although deep learning has more powerful learning and analytical computing functions than machine learning, the factors that cause cancer diseases are diverse, and the laws of gene mutations are currently uncontrollable. These factors will lead to different

data sets, which greatly increases the difficulty of model training. Too much model training will invest more labor costs, and the efficiency will be greatly reduced.

In deep learning, such as the variability of convolutional neural network models, there will be non-repeatability and non-reappearance in the model training process, so it is very challenging to establish a richer and more stable data set for analysis and prediction.

Finally, the model's predictions are unexplainable. The prediction results obtained from the training and analysis of a model are based on algorithms, but this unexplainability is very difficult to explain and illustrate in the medical field. Therefore, improving accuracy to reduce negative data is a challenge that artificial intelligence will face in the future[10].

5 Conclusion

By exploring the prediction and help of deep learning in the field of lung cancer, this article concludes that deep learning, as a branch of artificial intelligence, is more suitable for prediction in the field of cancer than machine learning. Among them, the prediction accuracy of convolutional neural networks is better. Deep learning is currently a more suitable method for analysis and prediction for complex cancers such as lung cancer, which have diverse causes and cancer cells are prone to mutation. The characteristics and functions of deep learning can well adapt to the mutation of cancer cells to quantify abstract biological characteristics and make predictions. It is convenient for clinicians to have a more scientific and accurate grasp and predict the patient's condition. Patients can also be given a suitable treatment plan through the predicted results to improve treatment efficiency and accuracy. In the future, people still need to develop and explore the help of deep learning in the field of lung cancer to solve current problems, so that deep learning can better benefit mankind and realize the value of artificial intelligence.

References

1. Zhong J.: Understanding lung cancer. Cancer Rehabilitation (1): 16-18, (2020).
2. Zhang C.J.:Artificial intelligence and deep learning. Science and Technology and Innovation, (13): 25-27(2019).
3. Dong H.Z.: AI automated testing. Beijing: Machinery Industry Press, (2020).
4. Fan Y., ed. Smart cities and information security in the era of big data. Beijing: Electronic Industry Press,(2018).
5. Song Y.: Progress in the clinical application of artificial intelligence deep learning in gastric cancer. China Medical GuideVol. 21(28): 63-66(2023).
6. Chen S.H, Liu W.X, Qin J, Chen L.L., BingG., Zhou Y.X., Wang T.F., Huang B.S.: Research progress of computer-aided diagnosis of cancer based on deep learning and medical images. Journal of Biomedical Engineering, 2017.
7. Mo C.S., Hong Q., Kong Z.H. Clinical pathological characteristics of young and middle-aged patients with lung cancer. Chinese Practical MedicineVol. 15(24):36-39(2020).

8. Jia L.Y., Ren X.T., Zhao J.J., et al. Research and progress of artificial intelligence in lung cancer imaging genomics. Journal of Taiyuan University of TechnologyVol. 53(4):571-587(2022).
9. Li X.L. Research progress of artificial intelligence in medical image processing. Chinese Journal of Medical Computer ImagingVol. 29(4):454-457(2023).
10. Hu P., Yan T.F., Zhou H.Z, Zhu X.G. Research progress of artificial intelligence in early detection of intracranial hemorrhage and hematoma segmentation in non-enhanced CT images. Chinese Journal of Cerebrovascular Diseases (Electronic Edition) 17(4): 410-416,(2023)