Proceedings of the 2023 International Conference on Machine Learning and Automation
DOI: 10.54254/2755-2721/33/20230268

Application of machine learning in medicine

QingWei Meng

Information college, Liaoning University, No.1-12 Chongshan West Road, Huanggu District, Shenyang, Liaoning, China, 110031

1848507895@qq.com

Abstract. In the past decades, the quantity and complexity of medical data have been increasing, such as medical images, genomics data, physiological signals and so on. These data contain a lot of valuable information, but traditional analysis methods and manual feature extraction have been unable to effectively deal with these large-scale and high-dimensional data. The rapid development of machine learning technology has brought new opportunities for medical research and clinical practice. Machine learning algorithms can learn and discover patterns, laws and prediction models from large-scale data, thus helping doctors and researchers to make more accurate and personalized diagnosis and treatment decisions. The application of machine learning in medicine has become a research field of great concern in recent years. This paper studies the application of machine learning in medicine, such as medical image diagnosis, genomics and drug discovery, and analyzes the relevant technical methods and ideas of machine learning in medicine, and analyzes the main algorithms and usage methods used in medical image diagnosis, such as convolution neural network and other deep learning algorithms. Through detailed analysis and research, it is found that machine learning provides a new method and tool, which can effectively process large-scale and complex medical data in medicine, bring more possibilities for medical diagnosis, treatment and research, and provide support for individualized medical care.

Keywords: artificial intelligence, machine learning, medical field, medical imaging, genomics, drug research and development, prediction model.

1. Introduction

The development of machine learning has gone through several important stages. The first stage is Symbolic artificial intelligence (AI), which focuses on the research of AI based on logical reasoning and uses rules and knowledge-based methods to solve problems. The second stage is Connectionism, in which the neural network model is introduced to simulate the information processing process of the human brain. Neural networks learn and reason by simulating the connections between neurons in the human brain. The third stage is Statistical Learning, which emphasizes learning probability models from data and making inferences and decisions by using statistical methods. Classical statistical learning methods include linear regression, logical regression, decision tree and so on. The fourth and final stage is Deep Learning. With the increase in computer processing capacity and data volume, deep learning shows an excellent ability to handle large-scale and complex data. Deep learning uses multi-layer neural network structure for feature extraction and representation learning, and has achieved great success in image recognition, speech recognition, natural language processing and

DOI: 10.54254/2755-2721/33/20230268

other fields. The deep neural network is a model based on neuron connection and hierarchical structure in machine learning. It consists of multiple hidden layers, and each hidden layer has multiple neuron nodes. These hidden layers can carry out nonlinear transformation and feature extraction, so that the network can learn a higher level of abstract representation from the original data. A. Ng, Y. LeCun, Y. Bengio and others discussed and analyzed the application of deep neural networks in machine learning and its revolutionary progress in recognition, classification and model generation, and introduced the importance and application of deep neural networks in the field of machine learning [1, 2]. This paper analyzes the application of machine learning in medicine, collects many applications and ideas about technology, such as medical imaging, genomics, etc., and analyzes and studies the algorithms designed in medical image diagnosis, genomics and drug discovery, analyzes their principles and adaptability, and studies the use methods and principles of some related algorithms. Generally speaking, the background of machine learning is that in the ever-evolving field of computer science and AI, machines can learn and improve automatically by using data and statistical methods, so as to realize automation and intelligence of various tasks. The research of machine learning helps us to better understand the data, make accurate predictions and decisions, and promote the development and application of artificial intelligence technology. It has a broad application prospect in various fields and is of great significance to promoting social progress and solving practical problems.

2. Application of machine learning in the medical field

2.1. Medical imaging diagnosis

In the field of medical image diagnosis, machine learning algorithms and technologies have made remarkable progress. Medical image diagnosis uses some machine learning algorithms, such as Convolutional Neural Networks (CNN), which is a deep learning algorithm, especially suitable for image processing tasks. In medical image diagnosis, CNN is widely used in image classification, target detection and segmentation. It extracts features from images through a multi-layer convolution layer and pooling layer, and then uses a fully connected layer for classification or location. Support Vector Machines (SVM) algorithm, which is a supervised learning algorithm, classifies by finding an optimal hyperplane. In medical image diagnosis, SVM can be used for classification tasks, such as judging benign and malignant tumors and detecting abnormal regions. Researchers such as Esteva used more than 129,000 clinical image data sets of different types of skin lesions to train CNN. CNN has achieved performance comparable to that of professional dermatologists in distinguishing malignant melanoma from benign lesions. Esteva's study highlights the potential of deep learning technology in improving diagnostic accuracy and helping to detect skin cancer early [3]. The Random Forest algorithm is an integrated learning algorithm, which is composed of multiple decision trees and classified by a voting mechanism. In medical image diagnosis, the random forest can be used for image segmentation, lesion detection and classification. In addition to CNN, other deep learning algorithms are also widely used in medical image diagnosis. For example, Recurrent Neural Networks (RNN) can be used to process sequence data, such as time series physiological signals. Generative Adversarial Networks (GAN) can be used to synthesize medical image data. Transfer Learning algorithm refers to a model that has been trained in one field, and it can be fine-tuned or transferred to another related field. In medical imaging diagnosis, transfer learning can use the trained model to speed up the training process of new data and improve performance. With the development of deep learning technology, the black-box characteristics of neural networks have become a challenge. Therefore, researchers are trying to develop methods to explain the decision-making process of neural networks, so as to improve their interpretability and credibility.

The application of these machine learning algorithms and technologies in medical image diagnosis is helpful to assist doctors in tasks such as lesion detection, classification, segmentation and location. Rajpurkar and others proposed a method based on deep learning [4]. They used a large-scale chest X-ray dataset to train the algorithm, aiming at achieving pneumonia detection at the radiology level. The research results show that the deep learning algorithm has reached a level comparable to that of

DOI: 10.54254/2755-2721/33/20230268

professional radiologists in pneumonia detection. This study shows the potential of deep learning technology in the field of medical image recognition and helps to improve the accurate diagnosis of pneumonia. They can improve the accuracy and speed of diagnosis, and provide support for precision medicine. However, it should be noted that these algorithms and techniques are only used as auxiliary tools, and the professional judgment and experience of doctors are still indispensable factors in the diagnosis process.

2.2. Genomics

The application of machine learning in genomics includes many fields, including gene recognition, gene regulation, genome variation, and so on.

Gene prediction and annotation, through supervised learning algorithms, such as SVM and Random Forest, models can be trained to predict and annotate genes. These models can predict the position and function of genes according to the characteristics of DNA sequence, such as open reading frame (ORF), promoter and splicing site. In genomics at high latitudes, the accuracy of prediction can be achieved by a small number of samples [5]. Gene expression analysis, using unsupervised learning methods such as cluster analysis, principal component analysis (PCA) and factor analysis, can classify and reduce the dimension of gene expression data according to similarity, thus revealing gene expression patterns and functions. In addition, deep learning methods such as CNN and RNN are gradually applied to the analysis and prediction of gene expression data. DNA sequence analysis, using sequence alignment and sequence pattern recognition techniques, combined with the hidden Markov model (HMM) and conditional random field (CRF) and other algorithms, can analyze functional elements in DNA sequences, such as promoters, splicing sites, transcription factor binding sites and so on. These methods can help reveal the structure and function of gene regulatory networks and DNA sequences. Genome variation analysis, through machine learning algorithms, such as support vector regression (SVR) and Random Forest Regression, can predict the relationship between genotype and phenotype according to genome variation data, so as to identify disease-related genetic variation and provide support for precision medicine. Protein structure prediction, using neural networks and deep learning methods, such as CNN and RNN, can predict the three-dimensional structure of protein and help to understand the function and interaction of protein. In the study of population genetics, machine learning methods, such as cluster analysis and correlation analysis, can be used to model and analyze the genomic differences between populations and reveal human evolution, population structure and genetic diversity.

The application of these machine learning algorithms and technologies in genomics is helpful to speed up data processing and analysis, discover potential biological knowledge, and provide new tools and methods for the research and application of genomics. Gulshan et al. developed a method based on a deep learning algorithm to detect diabetic retinopathy in retinal fundus photos. Researchers use large-scale data sets to train and verify the algorithm. The results show that the performance of the deep learning algorithm in diagnosing diabetic retinopathy is comparable to that of professional ophthalmologists. This study emphasizes the potential of deep learning technology in the detection of diabetic retinopathy [6]. However, it should be noted that the machine learning algorithm is only used as an auxiliary tool, and it still needs to be combined with experimental verification and expert knowledge for comprehensive analysis and interpretation in genomics research.

2.3. Drug discovery

The application of machine learning in drug discovery can help accelerate the screening and optimization of candidate compounds and improve the efficiency of drug research and development. Through literature investigation, many related algorithm technologies have been found.

Drug screening, by using machine learning algorithms, such as SVM [7], Random Forest and neural network, large-scale compound libraries can be virtually screened to predict their binding affinity, selectivity and activity with specific targets. This method can help to screen candidate compounds with potential drug activity. Drug property prediction, using machine learning algorithms,

DOI: 10.54254/2755-2721/33/20230268

such as regression analysis and deep learning models, can predict various properties of drugs, such as solubility, pharmacokinetics and toxicity, according to molecular structure and physical and chemical properties. This helps to take these properties into account when designing compounds and reduces the need for experiments. The prediction of drug action mechanism can predict the action mechanism of new drugs by analyzing the relationship data between known drugs and targets, combined with machine learning algorithms, such as cluster analysis and network analysis. This is of great significance for understanding the mode of action of drugs, finding new targets and designing multi-target drugs. Drug molecular optimization, using machine learning algorithms and deep learning models, can optimize the design of drug molecules. By generating and evaluating a large number of compound structures, such as GAN and Variational Automatic Encoder (VAE), potential candidate compounds can be provided and their structures can be optimized. Drug side effects prediction uses machine learning algorithms and data mining technology, such as association rule mining and classification algorithm, which can predict and identify potential drug side effects from a large-scale drug database. This is helpful for early detection and avoidance of potential safety problems in drug research and development. Through machine learning algorithms and optimization methods, different drug combinations can be simulated and optimized to achieve better curative effects and reduce side effects. This is of great significance for designing personalized treatment schemes and developing multidrug resistance.

The application of these machine learning algorithms and technologies can help accelerate the process of drug discovery, reduce costs, and promote the development of personalized drug research and development and precision medicine. Bychkov and others used deep learning technology to analyze the tissue samples of colorectal cancer and predicted the prognosis of patients [8]. Researchers trained the deep learning algorithm using a large number of tissue slice image data sets. The results show that the algorithm can accurately predict the prognosis of patients with colorectal cancer. This study shows the potential of deep learning in cancer prognosis evaluation and helps to guide clinical treatment decisions. However, it should be noted that the machine learning algorithm is only used as an auxiliary tool, and it still needs to be combined with experimental verification and expert knowledge for comprehensive analysis and decision-making in drug discovery.

In a word, the future development prospect of machine learning in the medical field is very broad. The future development of machine learning in the medical field will focus on individualized medical care, medical image analysis, interdisciplinary cooperation, intensive learning and independent decision-making, data privacy and security. These advances will bring more accurate, efficient and personalized medical services and improve the level of medical research and clinical practice.

3. Realization of machine learning related technologies in the medical field

The history of machine learning in the medical field can be traced back to the 1950s. In the early stage (1950s-1960s), machine learning was mainly used in pattern recognition and pattern classification tasks. Medical researchers use statistical methods and linear regression models to analyze medical data such as electrocardiograms and radiological images. In the period of further development (1970s-1980s), with the improvement of computing power and data acquisition technology, machine learning has been more widely used in the medical field. Artificial Neural Networks have been applied to medical image and signal processing. Then it entered the stage of vigorous development (1990s-2000s). With the progress of machine learning algorithms and medical data accumulation, the application of machine learning in the medical field became more diversified and complicated. Algorithms such as Support Vector Machines and Decision Trees are widely used in medical image analysis, cancer prediction, drug design and so on. Finally, in the period of the rise of deep learning (from the 2010s to the present), that is, in recent years, with the rise of deep learning technology, the development of machine learning in the medical field has entered a new stage. Convolutional Neural Networks have made great breakthroughs in medical image analysis, disease diagnosis and individualized treatment.

DOI: 10.54254/2755-2721/33/20230268

In addition to these milestones, there are many specific applications that promote the development of machine learning in the medical field. For example, using machine learning to analyze genomic data to find genetic variations related to diseases; Predicting patients' reactions to drugs through machine learning to realize individualized treatment; Assisting doctors in disease diagnosis and medical image analysis. These applications not only improve the efficiency of medical research and clinical practice, but also bring new innovations and opportunities to the medical field.

The core idea of machine learning in the medical field is to learn patterns and laws from a large number of medical data to help solve medical problems and provide better medical services. First of all, machine learning relies on a large number of medical data, including clinical data, biological sample data, medical images and so on. By analyzing and learning these data, machine learning can extract features, build models and make predictions. Data monitoring of machine learning was mentioned in the study of evaluating breast cancer. This study aims to evaluate the performance of a system based on artificial intelligence in breast cancer screening. The research team trained a deep learning algorithm by using a large-scale breast X-ray image data set and compared it with professional doctors around the world [9]. The results show that the performance of the artificial intelligence system in breast cancer detection is comparable to that of professional doctors, and even exceeds the average level in some cases. This study reveals the potential of machine learning in breast cancer screening and proves its importance and application value in the medical field. Secondly, through pattern recognition and classification, machine learning can identify patterns and laws in medical data and classify them into different categories. For example, identify tumor types, predict disease risks, and so on.

Then, machine learning can predict future results or events by learning existing medical data. For example, predict the possible disease development of patients and drug reactions, etc. This kind of prediction and decision support can help doctors make more effective treatment plans and decisions. In addition, machine learning can provide personalized medical services for each patient according to his individual characteristics and medical history. By learning a large number of patient data, machine learning can generate individualized treatment plans and suggestions. Machine learning also has the advantages of automation and high efficiency, which can automatically extract information and knowledge from a large number of medical data, reduce the burden on doctors and improve work efficiency. For example, in medical image analysis, machine learning can automatically identify tumors or abnormal areas and assist doctors in diagnosis.

In short, the development of machine learning in the medical field has experienced decades of evolution and innovation. With the increase of data, the improvement of algorithm models and the improvement of computing power, the application of machine learning in the medical field has become more and more extensive and mature. The development of machine learning in the medical field has also gone through the process of data collection, feature selection, model training and verification, practical application and continuous improvement, which has brought many potential benefits and opportunities for medical research and clinical practice. The core idea of machine learning in the medical field is to learn patterns and laws from a large number of medical data in a data-driven way, so as to achieve the goals of prediction, classification, decision support and individualized medical care. Researchers such as Poplin used large-scale data sets to train deep learning algorithms to predict patients' cardiovascular health [10]. The results show that the algorithm can accurately predict the cardiovascular risk factors of patients, such as hypertension and high cholesterol, from simple non-invasive retinal scanning. Poplin’s study highlights the application potential of deep learning in the medical field and provides a new method for the prevention and diagnosis of cardiovascular diseases. These ideas make machine learning an indispensable tool and method in medical research and clinical practice.

4. Conclusion

In this paper, the application of machine learning in the medical field and its potential advantages in diagnosis, treatment and prediction are analyzed. Through the comprehensive investigation of related

DOI: 10.54254/2755-2721/33/20230268

research and cases, it is can know that machine learning has great potential in the medical field and can help doctors improve accuracy and efficiency. For example, in medical image diagnosis, machine learning algorithms can automatically detect abnormal areas, and assist in screening and classifying disease types. Through machine learning, big data analysis technology can be used to mine patterns and knowledge in medical data, thus improving disease risk assessment and prediction. This is very important for finding potential disease risk factors in advance and making personalized treatment plans. Machine learning can also help improve the drug research and development process and clinical trial design. By analyzing large-scale bioinformatics data, machine learning can accelerate the discovery of new drugs, improve drug safety and individualized treatment effect.

Although machine learning has broad application prospects in the medical field, it still faces some challenges and limitations. For example, data privacy and security, algorithm interpretability, model verification and adaptability need to be further studied and solved. The application of machine learning in the medical field provides new opportunities for improving medical practice and patient care. But supervision and regulation in technology, morality and law should be strengthened, and so that the application of machine learning meets ethical and legal requirements and truly benefits human health.

References

[1] A. Ng, (2004)."Feature selection, L1 vs. L2 regularization, and rotational invariance," in Proceedings of the 21st International Conference on Machine Learning (ICML), 04 Proceedings of the twenty-first international conference on Machine learning, Stanford.

[2] Y. LeCun, Y. Bengio, and G. Hinton. (2015). "Deep learning," Nature, 521 (7553), 436-444.

[3] Andre Esteva, Brett Kuprel, Roberto A. Novoa et al. (2017).Dermatologist-level classification of skin cancer with deep neural networks. Nature,542(7639),115-118.

[4] Pranav Rajpurkar, Jeremy Irvin, Kaylie Zhu, et al. (2017). CheXNet: Radiologist-level pneumonia detection on chest x-rays with deep learning.arXiv preprint arXiv: 1711.05225.

[5] Trevor Hastie, Robert Tibshirani, Jerome Friedman (2009). The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Springer. 33.

[6] Varun Gulshan, Lily Peng, Marc Coram, Martin C. Stumpe et al. (2016). Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. JAMA,316(22),2402-2410.

[7] Bishop, C.M. (2006). Pattern Recognition and Machine Learning. Springer. 351.

[8] Dmitrii Bychkov, Nina Linder, Riku Turkki et al. (2018). Deep learning based tissue analysis predicts outcome in colorectal cancer. Scientific Reports,8(1),1-11.

[9] Scott M. McKinney, Marc Sieniek, Varun Godbole, et al. (2020). International evaluation of an AI system for breast cancer screening. Nature, 577(7788), 89-94.

[10] Ryan Poplin, Avinash Varadarajan, Katy Blumer et al. (2018). Prediction of cardiovascular risk factors from retinal fundus photographs via deep learning. Nature Biomedical Engineering,2(3),158-164.