Journal of Artificial Intelligence Practice (2024)
Clausius Scientific Press, Canada

DOI: 10.23977/jaip.2024.070214
ISSN 2371-8412 Vol. 7 Num. 2

Application Research of Machine Learning Algorithms in Medical Diagnosis

Yunfan Deng*

Guilin University of Electronic Technology, Guilin, Guangxi, 541004, China
**Corresponding author: 1753161295@qq.com*

Keywords: Machine Learning; Medical Diagnosis; Multimodal Data Fusion; Automated Diagnosis; Deep Learning

Abstract: This paper discusses the application of machine learning algorithms in medical diagnosis, aiming to explore their impact on the accuracy and efficiency of disease diagnosis. Through case analysis and literature review, the paper elaborates on the application of machine learning algorithms in medical diagnosis, including multimodal data fusion, automated diagnosis, and deep learning algorithms. The research indicates that machine learning algorithms can improve the accuracy and efficiency of disease diagnosis through learning and analysis of large amounts of data. While machine learning algorithms have great potential in medical diagnosis, it is essential to ensure data quality, privacy, and algorithm transparency. Interdisciplinary collaboration is recommended to promote sustainable development in healthcare.

1. Introduction

Medical diagnosis is a crucial process for determining the causes of diseases and selecting treatment plans, the accuracy of which is vital for patients' health and treatment outcomes. However, the complexity of the medical field and the diversity of diseases pose significant challenges to the diagnostic process. Traditional medical diagnostic methods often rely on the experience and intuition of doctors, but this approach carries risks of subjectivity and misdiagnosis.

With the rapid development of machine learning technology, an increasing body of research indicates the immense potential and significance of machine learning algorithms in medical diagnosis. Machine learning algorithms can learn and analyze large amounts of medical data to discover hidden patterns and regularities in diseases, thereby providing rapid and accurate diagnostic results[1,2]. Compared to traditional methods, machine learning algorithms offer better objectivity and predictive ability[3-5], assisting doctors in making more accurate diagnoses.

In this paper, firstly, several commonly used machine learning algorithms for medical diagnosis are introduced to lay the foundation for subsequent application research. Subsequently, specific cases are cited to elaborate on the application of machine learning algorithms in medical diagnosis, explaining how they enhance diagnostic accuracy and efficiency. Lastly, the future prospects and potential challenges of machine learning algorithms in medical diagnosis are summarized, along with relevant recommendations. Through this research, the aim is to provide valuable reference and inspiration for researchers in the medical field and related areas, to attract more attention to the

application of machine learning algorithms in medical diagnosis, and to promote further development and utilization in this field.

2. Overview of Machine Learning Algorithms

2.1 Support Vector Machine (SVM)

Support Vector Machine (SVM) is a common supervised learning algorithm suitable for binary classification problems. In medical diagnosis, SVM can automatically learn and identify disease features by training on a large amount of disease data, enabling disease classification and prediction. For example, in lung cancer diagnosis, SVM can analyze patients' CT scan data to automatically detect and classify early-stage lung cancer.

2.2 Random Forest (RF)

Random Forest (RF) is an ensemble learning algorithm based on decision trees, which improves the accuracy and stability of the model by combining the predictions of multiple decision trees. In medical diagnosis, RF can classify and predict tumor types by analyzing patients' biomarker data (such as gene expression profiles). The advantage of RF is its ability to handle high-dimensional data and its robust performance in the presence of noise and missing data.

2.3 Decision Tree (DT)

Decision Tree (DT) is a supervised learning algorithm based on partitioning the dataset by feature values. It makes decisions in a tree-like structure, with each internal node representing a feature, each branch representing a feature value, and each leaf node representing the final classification result. Decision trees are easy to understand and interpret, suitable for classification and regression problems. In medical diagnosis, decision trees can analyze patients' medical history, signs, and laboratory test results to assist doctors in making diagnostic decisions. For example, by analyzing patients' clinical symptoms and examination results, decision trees can help doctors determine whether patients have a certain disease.

2.4 Deep Learning (DL)

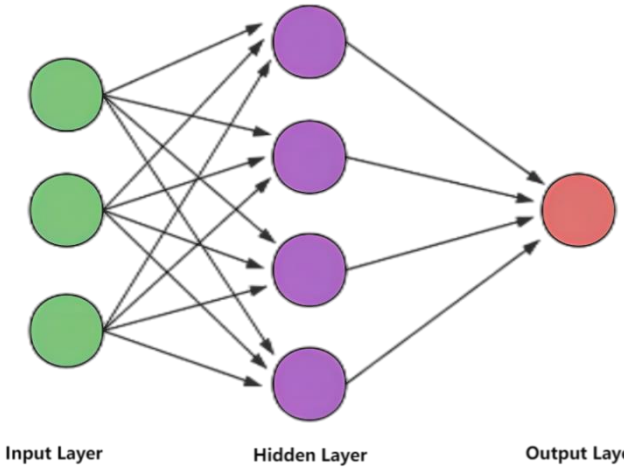

Figure 1: Basic Structure of Deep Learning Neural Network

Deep Learning (DL) constructs neural network models[6,7], as shown in Figure 1, to simulate the neural network structure of the human brain, with powerful feature learning and representation capabilities. In medical imaging diagnosis, DL can automatically identify and diagnose diseases by analyzing pixel features, texture features, etc., of medical images. For example, in breast cancer diagnosis, DL can analyze patients' X-rays and MRI data to automatically identify and label potential abnormal areas, helping doctors accurately locate diseases.

3. Application of Machine Learning Algorithms in Medical Diagnosis

3.1 Rapid Screening and Preliminary Diagnosis

In the process of medical diagnosis, doctors typically need to conduct a series of examinations and assessments to determine the etiology of patients, which is often time-consuming and tedious. Machine learning algorithms can rapidly process large amounts of data to perform preliminary disease classification and screening, thereby improving the working efficiency of medical professionals. For example, machine learning algorithms can quickly identify suspected cases, assisting doctors in conducting more in-depth diagnosis and treatment. Additionally, these algorithms can predict disease progression and outcomes based on patients' clinical features and historical data, providing doctors with more accurate treatment plans.

To illustrate the application of machine learning algorithms in rapid screening and preliminary diagnosis more intuitively, we take lung cancer screening as an example. Lung cancer screening is an important public health issue, as early detection and treatment of lung cancer can significantly improve the chance of survival. However, conventional lung cancer screening methods (such as chest X-rays) often have lower sensitivity and specificity. In recent years, with the widespread use of computed tomography (CT) technology, a large amount of CT image data has been generated. Utilizing machine learning algorithms to analyze these image data can achieve efficient screening for lung cancer.

The following is a typical application process of machine learning algorithms[8] in lung cancer screening:

(1) Data Collection: Collect a large number of chest CT image data, including data from both normal individuals and lung cancer patients.

(2) Data Preprocessing: Preprocess the raw image data, such as image enhancement and noise reduction, to improve the performance of the algorithm.

(3) Feature Extraction: Extract key features conducive to lung cancer from the preprocessed image data, such as shape, texture, edges, etc.

(4) Model Training: Train machine learning models, such as SVM, RF, etc., using the extracted feature data.

(5) Model Evaluation: Evaluate the performance of the model through methods such as cross-validation, accuracy, recall, etc.

(6) Clinical Application: Apply the trained model to clinical lung cancer screening to assist doctors in rapid diagnosis.

3.2 Automated Diagnosis

Automated diagnosis refers to the process of utilizing computer algorithms to analyze medical images, physiological data, etc. In the medical field, the application of automated diagnosis greatly improves the accuracy and efficiency of diagnosis. Among them, deep learning algorithms, as powerful tools, perform exceptionally well in automated diagnosis. Taking the Convolutional Neural Network (CNN) as an example, it can extract features from a large amount of medical image

data through learning and make accurate diagnoses[9,10].

Figure 2 illustrates the application results of CNN in breast cancer detection. By training on a large amount of breast X-ray data, CNN can accurately identify potential breast lesions, thereby improving the accuracy of diagnosis.

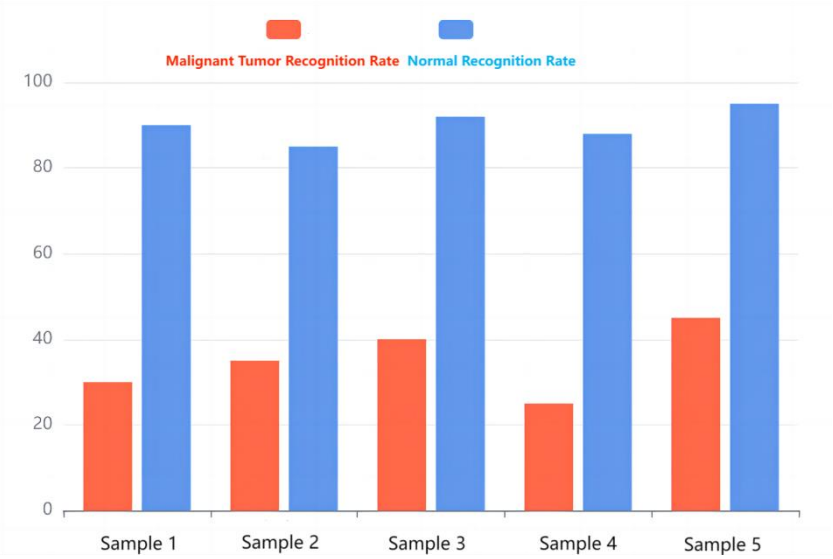

Figure 2: Application Results of CNN in Breast Cancer Detection

From Figure 2, it can be observed that in samples 1 to 5, the recognition of malignant tumors by CNN is significantly lower than that of normal tissues. This indicates that CNN can accurately distinguish malignant tumors from normal tissues in breast cancer detection, facilitating timely intervention and treatment by doctors and improving patients' survival rates.

3.3 Multimodal Data Fusion

In the field of medicine, machine learning algorithms play a crucial role in addressing the issue of multimodal data fusion. Firstly, they can handle various types of data, including imaging data, laboratory test data, and genomic data. These data have different representations and formats, but machine learning algorithms can effectively fuse them together by learning the inherent patterns and relationships in the data. Secondly, machine learning algorithms can automatically extract features from the data and use these features for tasks such as classification and prediction. In multimodal data fusion, machine learning algorithms can build multiple decision trees and average their predictions to reduce the risk of overfitting. Additionally, ensemble learning methods such as RF, gradient boosting trees, etc., can fully utilize the predictions of multiple models to further improve classification accuracy and generalization ability.

A typical application case of multimodal data fusion is tumor diagnosis. In tumor diagnosis, imaging data can display the location, size, and morphology of tumors, while genomic data can reveal biomarkers of tumors. By fusing these two types of data, the nature of the tumor can be more accurately determined, providing personalized treatment plans for patients. Moreover, machine learning algorithms can also be applied to other medical fields such as cardiovascular diseases, neurological diseases, etc., to provide strong support for the diagnosis, treatment, and prevention of diseases through multimodal data fusion.

4. Conclusion

This paper discusses the application and advantages of machine learning algorithms in medical

diagnosis. Through automated diagnosis and multimodal data fusion, machine learning algorithms have significantly improved the accuracy and efficiency of disease diagnosis. In the future, with further development of data collection and storage technologies, machine learning algorithms will be more widely applied in medical diagnosis. Personalized healthcare, precision diagnosis, and treatment will become more popular, thereby enhancing the quality and efficiency of medical services.

However, machine learning algorithms in medical diagnosis also face some challenges. Concerns regarding data quality, privacy protection, and algorithm interpretability require further attention and resolution. Additionally, legal regulations and ethical issues in medical diagnosis require stricter standards and regulations. Therefore, interdisciplinary cooperation is recommended, including efforts from medical experts, data scientists, policymakers, and the general public, to ensure the sustainable development of machine learning algorithms in medical diagnosis, bringing more benefits to the cause of human health.

References

[1] Song X.Q., Zhang X.L., Ma X.M., et al. Research progress on the application of machine learning in clinical decision support systems for infectious diseases [J]. Nursing Research, 2024, 38(03).

[2] Zehra Karapinar Senturk. Early diagnosis of Parkinson's disease using machine learning algorithms[J]. Medical Hypotheses. Volume 138, Issue. 2020. PP 109603.

[3] Ren X.X. Research on medical image segmentation algorithm based on multi-feature fusion [D]. Shandong: Qilu University of Technology, 2023.

[4] Gao G., Ma S., Wang X.Y. Basic principles and application progress of computer-aided diagnosis in medical imaging diagnosis [J]. Radiology Practice, 2016, 31(12).

[5] V A Binson, Sania Thomas, M Subramoniam, et al. A Review of Machine Learning Algorithms for Biomedical Applications [J]. Annals of Biomedical Engineering. Volume, Issue. 2024.

[6] Li N. Application research of machine learning model in survival prediction of endometrial cancer [D]. Shandong: Shandong University of Finance and Economics, 2024.

[7] Yu K.M., Bao X.D., Li B., et al. Research on machine learning algorithms for Raman spectroscopy image in biomedical detection [J]. Chinese Medical Equipment, 2021, 36(08).

[8] Du Z.Z. Application of machine learning algorithm based on electronic medical records in the prediction of cardiovascular diseases [D]. Hubei: Wuhan Institute of Posts and Telecommunications, 2020.

[9] Zhang F., Hou G.D., Wu N.X., et al. Performance study of machine learning algorithms for breast cancer detection[J]. Journal of Jiamusi University (Natural Science Edition), 2020, 38(01).

[10] Chen S. Research on medical auxiliary diagnosis algorithm based on machine learning [D]. Fujian: Fuzhou University, 2020.