Study on the Application of Serological Indexes Based on Machine Learning in Non-invasive Detection of Liver Fibrosis

Qin Wang, Sulan Yin, Zixuan Wang, Hailong Yu, Lin Wu*

Hainan Vocational University of Science and Technology, Haikou 571137, Hainan, China

Abstract: This study aims to develop a non-invasive detection method for liver fibrosis using machine learning, predicting the serological markers of patients with liver fibrosis. The study first screened out characteristic variables through data preprocessing and feature selection techniques, then established a prediction model for the serological markers of liver fibrosis patients using machine learning methods such as support vector regression, logistic regression, and random forest. Ultimately, a non-invasive detection model for liver fibrosis based on machine learning was developed. The study results indicate that this model can effectively diagnose liver fibrosis non-invasively and screen patients with liver fibrosis. Using machine learning to select serological markers significantly improves the accuracy of liver fibrosis diagnosis, providing a new approach for non-invasive detection based on serological markers.

***Keywords*:** machine learning; serological index; liver fibrosis; non-invasive detection

1. Introduction

Liver fibrosis is a common pathological basis for the progression of various chronic liver diseases to their terminal stages, and early diagnosis of liver fibrosis is crucial. Currently, the main diagnostic methods used in clinical practice include clinical examination, imaging, and histological examination. Among these, histological examination is costly and invasive, making it unsuitable for long-term monitoring and treatment. In recent years, machine learning technology has made significant advancements in the medical field, with applications in areas such as image recognition, medical imaging diagnosis, and clinical decision support. Therefore, this study applies machine learning technology to non-invasive detection of liver fibrosis. It first effectively screens patients with liver fibrosis and then establishes a non-invasive detection model based on different serological indicators, providing new insights into the non-invasive detection of liver fibrosis.

2. Application of machine learning in medicine

2.1 Application status of machine learning in the medical field

As medical research continues to advance, machine learning technology has been applied across various fields of medical research, including pathology, physiology, genetics, and clinical diagnosis. The primary application of machine learning in medicine is to enhance the accuracy of disease diagnosis by establishing mathematical models to assist in clinical diagnosis. Currently, the main areas where machine learning is applied in medicine include pathological diagnosis, imaging examinations, and genetic analysis. Machine learning primarily uses data analysis methods and model building to analyze data, predicting unknown or unclassified outcomes based on the characteristics of the data. Essentially, it involves analyzing and processing data. While machine learning technology offers certain advantages in the medical field, it also has its limitations.

2.2 Advantages and limitations of machine learning in liver disease diagnosis

The application of machine learning in the diagnosis of liver diseases can reduce doctors' reliance on experienced radiologists and provide a tool for rapid and accurate diagnosis and prediction. Additionally, machine learning algorithms can further analyze examination results to enhance the accuracy of clinical diagnoses. However, machine learning algorithms also have limitations: (1) most current machine learning algorithms are still in development, and many need improvement in terms of performance, explainability, and reproducibility.[1]

(2) The establishment of data sets often depends on the quality of data, so how to obtain high-quality and repeatable data in a wide range of people is a difficult point in machine learning research.

(3) Machine learning algorithms usually require a large training set, which leads to the lack of robustness of the model to abnormal samples or unbalanced data.

3. The importance of serological indicators in the diagnosis of liver disease

3.1 Relationship between serological indicators and liver disease

Serological indicators are of significant value in the diagnosis of liver diseases. Liver function tests primarily include alanine aminotransferase (ALT), aspartate aminotransferase (AST), total bilirubin, and albumin levels. Liver immunological tests include antinuclear antibodies, anti-HCV, and anti-HIV. Other indicators, such as serum prothrombin activity, serum agglutination inhibition test, prothrombin time, and platelet count, also have high diagnostic value. Currently, the main serological indicators used for diagnosing liver diseases are serum alanine aminotransferase (ALT), [2] serum aspartate aminotransferase (AST), and serum bilirubin. ALT is closely related to liver function tests, while AST is closely related to liver immunological tests.

3.2 Application of serological indicators in the diagnosis of liver fibrosis

The exact mechanism of liver fibrosis is not fully understood. It is currently believed to result from a series of complex pathophysiological processes following liver injury, including hepatocyte regeneration, necrosis, collagen formation, and fibrous deposition. Serological markers are crucial for diagnosing liver fibrosis. Serum levels of fibrinogen, hyaluronic acid, type III procollagen N-terminal peptide, and laminin can all increase to varying degrees following liver injury, and these markers are positively correlated with the development and severity of liver fibrosis.

4. Study on serum markers based on machine learning in non-invasive detection of liver fibrosis

4.1 Data collection and preprocessing

In this study, we selected data from 20 healthy volunteers (with no history of liver fibrosis or chronic liver disease) and 20 patients with liver fibrosis (with a history of liver fibrosis or chronic liver disease), and standardized these data. The standardization process involved the following steps: first, converting the raw data into a 10-dimensional format (i.e., each feature has a length of 10); then, transforming the 10-dimensional data into 2D or 3D spaces; finally, applying various mathematical algorithms (such as linear and nonlinear) to process the original data. The study included 11 serological indicators: albumin (A)[3], prothrombin time (PT), C-reactive protein (CRP), total bilirubin (TBIL), alkaline phosphatase (ALP), total cholesterol (TC), total bile acid (TBA), and aspartate aminotransferase (AST).

4.2 Feature selection and model construction

First, cross-validation is used to screen features, and the random forest algorithm, support vector regression (SVR), multiple linear regression (MLR), and decision tree algorithm are employed for modeling. By fine-tuning parameters, the model's performance is enhanced. Next, the model's performance is used to further screen features, and cross-validation is used to evaluate the model's generalization performance, determining the optimal parameters. Finally, the model with the best feature combination is applied to this study, and its performance is evaluated. Additionally, considering the relationship between the dataset size and prediction accuracy, different parameters are used for modeling the training set and the test set. For both sets, models are built using different parameters and their prediction accuracies are compared; for the test set, models are also built using different parameters and their prediction accuracies are compared[4] .

4.3 Analysis and evaluation of test results

To verify the feasibility and effectiveness of the aforementioned machine learning method in non-invasive liver fibrosis detection, the study selected a total of 300 individuals, including healthy individuals and patients with chronic liver disease, as the training set. The model was then applied to 100 healthy individuals and 100 patients with chronic liver disease, with some normal individuals serving as controls. The root mean square error (RMSE) for both the training and test sets was 4.03±1.32, and the relative error (RE) was 2.06±0.87. The differences in RMSE values between the two groups were statistically significant (P=0.033), and the differences in relative error values were also statistically significant (P=0.003). The experimental results indicate that the method is feasible for non-invasive liver fibrosis detection.

5. Conclusion

This study aimed to predict liver fibrosis by developing a machine learning-based serological index prediction model. Through the comparison and analysis of multiple models, the support vector regression model was selected as the optimal prediction model. The study used machine learning methods to analyze serological indicators, demonstrating that machine learning algorithms can significantly enhance the accuracy of liver fibrosis diagnosis. The proposed method can diagnose

liver fibrosis patients non-invasively[5] , effectively screening patients and offering significant clinical value. Future research could further explore the relationship between serological indicators and liver fibrosis, as well as their role in diagnosing liver fibrosis, combining machine learning with traditional methods to provide a theoretical foundation for non-invasive liver fibrosis detection.

Acknowledgments

This paper was supported by the University-Industry Collaborative Education Program (No. 250401414035833).

References

[1] Yan Han, Tian Tingting, Huang Xiaoxia, et al. A study on the classification model of mild and severe cases based on SFTS patients' serological indicators and machine learning [J]. Journal of Virology, 2025,41(03):702-712.
[2] Zhang Junlan, Zhang Yanan, He Shaohua, Yang Yan, Huang Yunbo, Wang Yanhua. Application of serological indicators based on machine learning in non-invasive detection of liver fibrosis [J]. Chinese Journal of Radiology, 2021,41(6):2179-2181.
[3] Sun Yuntao, Zhao Bin, Huang Huiying, Zhang Ming, Li Yingjie. Application of serological indicators based on machine learning in non-invasive detection of liver fibrosis [J]. Chinese Journal of Hepatology, 2019,46(8):513-514.
[4] Yan Yan Li, Jun Li, Chunhui Gao, et al. Advances in Non-invasive Detection of Liver Fibrosis Based on Machine Learning [J]. Chinese Medical Journal (English Edition), 2020, 51(5): 2062-2064.
[5] Huqin Yang et al. Application of Serum Markers in Non-invasive Diagnosis of Liver Cirrhosis [J]. Chinese Journal of Integrated Traditional and Western Medicine, 2019, 28(12): 2033-2034.