International Journal of Interdisciplinary Innovative Research & Development (IJIIRD)
ISSN: 2456-236X
Vol. **11** Special Issue **02** | **2026**

AI-Based Multi Disease Prediction and Health Advisory System

Ms.Payal .V.Chopade[1], Ms.Gauri .R.Gantire[2], Ms.Sayali .R.Warade[3], Ms.Nikita .R. Khirolkar[4], Prof.Ashwini.D.Bhople[5]

[1,2,3,4] *Student, Computer Science & Engineering, Padmashri Dr V.B.Kolte College of Engineering ,Malkapur*
[5] *Assistant Professor, Computer Science & Engineering, Padmashri Dr V.B.Kolte College of Engineering, Malkapur*

DOI: 10.5281/zenodo.19286898

ABSTRACT

Chronic diseases have become a major threat to human health worldwide, particularly in developing and underdeveloped countries. Their long-term nature, high treatment costs, and frequent complications impose substantial economic and social burdens on both healthcare systems and families. As chronic diseases often coexist and evolve over time, accurate diagnosis and effective treatment planning remain challenging for clinicians. With the rapid development of medical information systems, large-scale and heterogeneous patient data—such as physiological indicators, laboratory results, and clinical records—have become increasingly available, providing new opportunities for intelligent disease prediction. Machine learning techniques are well suited for integrating and analyzing such heterogeneous data. In particular, multi-label learning methods are highly applicable to chronic disease prediction, as patients often suffer from multiple chronic conditions simultaneously. In this paper, we propose a novel multi-label neural network method (ML-NN) that combines neural network modeling with multi-label learning to predict chronic diseases. The proposed approach is trained using a cross-entropy loss function and optimized through a backpropagation algorithm, enabling effective learning of complex nonlinear relationships among multiple disease labels. To evaluate the performance of the proposed ML-NN model, extensive experiments were conducted on a real-world dataset consisting of 19,733 patients and 10 types of chronic diseases. The proposed method was compared with 14 traditional multi-label learning algorithms using five commonly used evaluation metrics. Experimental results show that ML-NN consistently outperforms the competing methods across all evaluation measures. These findings demonstrate that the proposed approach can effectively and accurately predict chronic diseases, providing valuable decision support for clinicians and contributing to improved chronic disease management and personalized healthcare.

Keywords:- ***Heart chronic diseases, machine learning, multi-label learning, neural networks, disease prediction, healthcare data, cross-entropy loss, backpropagation algorithm, clinical decision support, and performance evaluation.***

1. INTRODUCTION

Chronic diseases, also known as non-communicable diseases (NCDs), are characterized by long duration and slow progression and have become a major global health concern. Common chronic diseases include cardiovascular diseases, chronic respiratory diseases, and diabetes [1]. These conditions require long-term treatment and continuous management, which leads to high medical expenses and places a heavy economic burden on individuals, families, and society [2]. In recent years, the prevalence of chronic diseases has continued to increase, particularly in low- and
middle-income countries[3]. Chronic diseases account for a large proportion of the global disease burden and are responsible for an increasing number of deaths each year. Due to their complex causes and long-term nature, chronic diseases often develop into multiple complications over time. This complexity makes accurate diagnosis and effective treatment planning difficult for healthcare professionals [4]. With the rapid development of modern medical technologies, such as wearable devices and mobile health systems, large amounts of patient health data can now be collected continuously [5]. These data include laboratory test results, physiological measurements, and medical signals such as electrocardiograms [6]. However, patient data are often heterogeneous and complex, making manual analysis inefficient and error-prone. Therefore, there is a growing need for intelligent data analysis methods. Machine learning and data mining techniques provide effective tools for analyzing large-scale medical data. These techniques can help doctors identify disease patterns, predict health risks, and support clinical decision-making. Applying machine learning methods to chronic disease analysis has the potential to improve diagnosis accuracy and enhance patient care outcomes [7].

The Growing Burden of Chronic Diseases

Chronic diseases, also known as noncommunicable diseases (NCDs), are characterized by long duration, slow progression, and high treatment costs. They represent a major global health challenge, particularly in low- and middle-income countries, where they account for a large proportion of disease burden and an increasing number of deaths.

The Need for Data-Driven Medical Decision Support

Due to their complex causes and frequent complications, chronic diseases are difficult to diagnose and treat effectively. Although modern medical technologies generate large volumes of heterogeneous patient data, clinicians require machine learning and data mining techniques to efficiently analyze these data and support accurate diagnosis and treatment decisions.

2. LITERATURE SURVEY

Previous studies have applied various multi-label learning methods to chronic disease prediction and diagnosis. Commonly used algorithms include Binary Relevance, Classifier Chains, Label Powerset, HOMER, RAkEL, ML- kNN, Decision Trees, and Rank-SVM. These methods have shown effectiveness in identifying multiple disease labels for a single patient. However, many of these approaches treat disease labels independently or fail to fully capture correlations among diseases. In addition, medical data collected from different sources are often heterogeneous and high-dimensional, which limits the performance of traditional algorithms. Neural networks have recently gained attention in biomedical research due to their strong ability to model complex and nonlinear relationships. Despite their advantages, challenges such as limited interpretability and high computational cost still remain, motivating further research in this area.

Limitations of Traditional Multi-Label Learning Methods

Traditional multi-label learning algorithms, such as Binary Relevance and Label Powerset, have been widely used for chronic disease classification. Although these methods are relatively simple and easy to implement, they often ignore the dependencies between disease labels. Classifier Chains and ensemble-based methods attempt to model label correlations, but their performance may degrade when dealing with large-scale and heterogeneous medical data. Furthermore, distance-based and tree-based methods may struggle with missing values and noisy clinical data. These limitations reduce their effectiveness in real-world healthcare applications, where patients often suffer from multiple related chronic conditions.

Advantages and Challenges of Neural Network-Based Approaches

Neural networks have demonstrated strong performance in analyzing complex biomedical data due to their ability to learn nonlinear feature representations. They are particularly suitable for integrating multi-source patient data, such as clinical measurements and physiological signals. Neural network-based multi-label models can better capture hidden relationships among multiple chronic diseases. However, these models often require significant computational resources and large datasets for training. In addition, their black-box nature makes them difficult to interpret, which may limit their acceptance in clinical practice. Therefore, developing efficient and accurate neural network-based multi-label models remains an important research direction.

3. METHODOLOGY

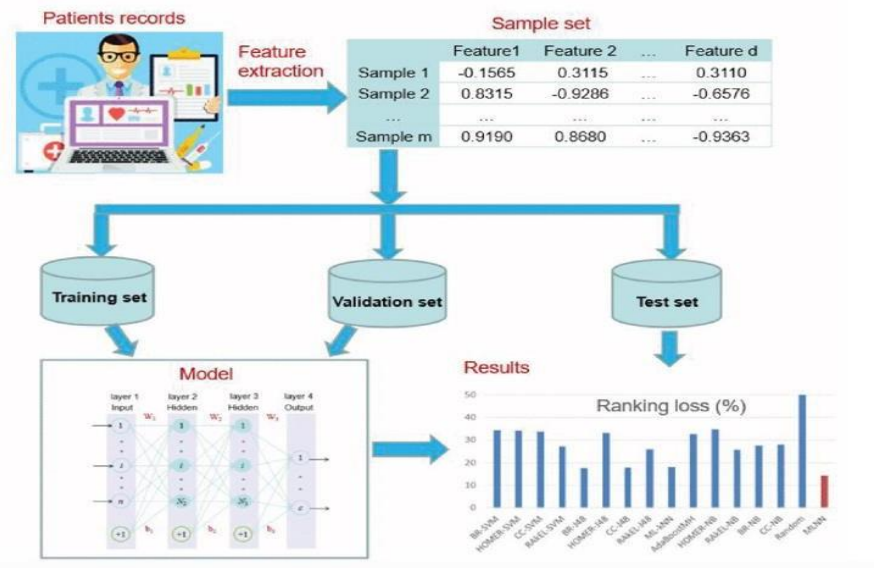

Fig:1 Chronic diseases prediction model based on multi-label neural network.

This study follows a systematic methodology to develop and evaluate a multi-label neural network model for chronic disease prediction. First, the dataset is derived from the MIMIC-II clinical database, which contains approximately 33,000 patient records. After filtering incomplete and irrelevant entries, 19,733 patients diagnosed with chronic diseases are selected for analysis. Data preprocessing is then performed to handle missing values and reduce noise. A total of 310 features representing 76 clinical attributes are extracted from the dataset. To ensure consistent feature scales, z-score normalization is applied to standardize the data.

The proposed ML-NN model is designed with two hidden layers, each containing 100 neurons, to capture complex nonlinear relationships in the data. A multi-label cross-entropy loss function is used to handle multiple disease labels simultaneously. The model parameters are optimized using gradient descent with backpropagation. For performance evaluation, the proposed model is compared with 14 traditional multi-label learning algorithms, including BR, CC, HOMER, RAkEL, ML-kNN, AdaBoostMH, SVM, J48, and Naive Bayes. Five widely used evaluation metrics— Hamming Loss, One-Error, Coverage, Ranking Loss, and Average Precision—are employed to comprehensively assess prediction performance.

Proposed System Architecture and Workflow

The workflow of the proposed system begins with the collection of patient records from the MIMIC-II clinical database. From the original dataset of approximately 33,000 records, 19,733 patients diagnosed with chronic diseases are selected for analysis. The data then undergo preprocessing to handle missing values, remove noise, and improve overall data quality. After preprocessing, 310 features representing 76 clinical attributes are extracted from heterogeneous medical data sources. To ensure consistency and avoid bias caused by different value ranges, z-score normalization is applied to standardize all features. The processed data are then used to train the proposed multi- label neural network model.

The architecture of the ML-NN model consists of an input layer, two hidden layers, and an output layer. The input layer receives the normalized feature vectors for each patient. Two fully connected hidden layers, each containing 100 neurons, are employed to learn complex and nonlinear relationships between patient characteristics and chronic disease labels. The output layer contains multiple neurons corresponding to different chronic diseases, enabling the model to predict the presence of multiple conditions simultaneously. A multi-label cross-entropy loss function is used to measure prediction error, and gradient descent with backpropagation is applied to optimize the network

parameters. Finally, the trained model is evaluated using standard multi-label performance metrics to assess its effectiveness in chronic disease prediction.

4. CONCLUSIONS

Chronic patients tend to have a variety of complications. For these patients, multi-label learning can be used to identify their complications. Comparative experiments on 10 chronic diseases show that the proposed multi-label neural network algorithm (ML-NN) in this paper is significantly better than the traditional multi-label

learning algorithms. Maybe this dataset is relatively large and the neural network can fit it well even if there are relevant features. What's more, the proposed method is also very competitive in running time, which is mainly due to the efficiency of multi-label loss function. Therefore, the proposed method can effectively assist doctors in the diagnosis and treatment of patients with chronic diseases. In the future, the interpretability of neural networks will be analyzed in order to make new discoveries and deepen the understanding of chronic diseases

5. REFERENCES

[1] R. Nugent, ''Preventing and managing chronic diseases,'' BMJ, vol. 364, p. l459, Jan. 2019.
[2] S. Mendis, S. Davis, and B. Norrving, ''Organizational update: The world health organization global status report on noncommunicable diseases 2014; one more landmark step in the combat against stroke and vascular disease,'' Stroke, vol. 46, no. 5, pp. e121–e122, May 2015.
[3] V. J. Wirtz, W. A. Kaplan, G. F. Kwan, and R. O. Laing, ''Access to medications for cardiovascular diseases in low-and middle-income countries,'' Circulation, vol. 133, pp. 2076–2085, May 24 2016.
[4] G. B. D. D. Collaborators, ''Health effects of dietary risks in 195 countries, 1990–2017: A systematic analysis for the Global Burden of Disease Study 2017,'' Lancet, vol. 393, no. 10184, pp. 1958–1972, May 11 2019.
[5] I. C. Jeong, D. Bychkov, and P. C. Searson, ''Wearable devices for precision medicine and health state monitoring,'' IEEE Trans. Biomed. Eng., vol. 66, no. 5, pp. 1242–1258, May 2019.
[6] S. J. Kabadi and A. Krishnaraj, ''Strategies for improving the value of the radiology report: A retrospective analysis of errors in formally over-read studies,'' J. Amer. College Radiol., vol. 14, no. 4, pp. 459–466, Apr. 2017.
[7] J. Béal, A. Montagud, P. Traynard, E. Barillot, and L. Calzone, ''Personalization of logical models with multi- omics data allows clinical stratification of patients,'' Frontiers Physiol., vol. 9, p. 1965, Jan. 2019.
[8] X. Guo, F. Liu, Y. Ju, Z. Wang, and C. Wang, ''Human protein subcellular localization with integrated source and multi-label ensemble classifier,'' Sci. Rep., vol. 6, no. 1, p. 28087, Jun. 2016.
[9] S. Wan, M.-W. Mak, and S.-Y. Kung, ''Sparse regressions for predicting and interpreting subcellular localization of multi-label proteins,'' BMC Bioinf., vol. 17, no. 1, p. 97, Feb. 2016.
[10] X. Wang, W. Zhang, Q. Zhang, and G.-Z. Li, ''MultiP-SChlo: Multilabel protein subchloroplast localization prediction with Chou's pseudo amino acid composition and a novel multi-label classifier,'' Bioinformatics, vol. 31, no. 16, pp. 2639–2645, Aug. 2015.