Predictive Models in the Chronic Disease Diagnosis

Anshuman Samantaray
Dept of Computer Science andvEngineering
Mail - samantarayanshuman99@gmail.com

Dr. Sujit Ku. Panda
Dept of Computer Science andvEngineering
Mail - sujit.panda@gift.edu.in

ABSTRACT

This paper reviews applications of machine learning (ML) predictive models in the diagnosis of chronic diseases. Chronic diseases (CDs) are responsible for a major portion of global health costs. Patients who suffer from these diseases need lifelong treatment. Nowadays, predictive models are frequently applied in the diagnosis and forecasting of these diseases.

In this study, we reviewed the state-of-the-art approaches that encompass ML models in the primary diagnosis of CD. This analysis covers 453 papers published between 2015 and 2019, and our document search was conducted from PubMed (Medline), and Cumulative Index to Nursing and Allied Health Literature (CINAHL) libraries. Ultimately,22 studies were selected to present all modeling methods in a precise way that explains CD diagnosis and usage models of individual pathologies with associated strengths and limitations. Our outcomes suggest that there are no standard methods to determine the best approach in real-time clinical practice since each method has its advantages and disadvantages. Among clustering were the most used. These models are highly applicable in classification, and diagnosis of CD and are expected to become more important in medical practice in the near future.

Keywords: chronic diseases; prediction models; pathologies; accuracy; disease classification; K-Nearest Neighbors (KNN); Convolutional Neural Networks(CNN); disease forecasting; disease management.

Chapter 1

Introduction

1.1 Purpose

Predictive models play a crucial role in modern healthcare, particularly in the diagnosis and management of chronic diseases. These models leverage data analytics and machine learning techniques to analyze vast amounts of medical data and predict the likelihood of developing a particular chronic condition, its progression, and response to treatment. The primary purpose of predictive models in chronic disease diagnosis is to enhance early detection, improve risk stratification, and optimizpersonalized treatment strategies, ultimately leading to better patient outcomes and more efficient healthcare delivery.

Chapter 2

Literature Review

2.1 Introduction

The utilization of predictive models for disease outcomes has garnered significant attention in recent years due to its potential to revolutionize healthcare delivery by enabling proactive and personalized approaches to disease management. This chapter provides a comprehensive review of the existing literature related to predictive modeling for disease outcomes, focusing on the development, validation, and application of predictive models across various chronic diseases. By synthesizing and analyzing the findings from previous research studies, this chapter aims to elucidate the current state-of-the-art in predictive modeling for disease outcomes and identify gaps and opportunities for further investigation.

The introduction of this chapter provides an overview of the importance and relevance of predictive modeling in the context of chronic disease management. It outlines the key objectives of the literature review, including:

To examine the methodologies and techniques used in the development of predictive models for disease outcomes.

To assess the performance and accuracy of existing predictive models across different chronic diseases.

To explore the potential applications of predictive modeling in clinical practice, public health, and healthcare policy.

To identify challenges, limitations, and future directions for research in the field of predictive modeling for disease outcomes.

Through a comprehensive review of the literature, this chapter aims to provide a solid foundation for the subsequent chapters of the thesis, which will focus on the development and validation of a novel predictive model for a specific hronic disease. By synthesizing existing knowledge and insights, this literature review sets the stage for the original contributions and innovations that will be presented in the subsequent chapters, ultimately advancing our understanding

3.5 K-Nearest Neighbour (KNN)

KNN, also known as k-nearest neighbours, is a straightforward yet effective technique used in machine learning for both classification and regression applications. It is a supervised machine learning technique that compares how much the new and old data resemble one another before adding the new data to the category that matches the existing categories the most. It is an instance-based, nonparametric approach that depends on the idea of distance or similarity between data points.

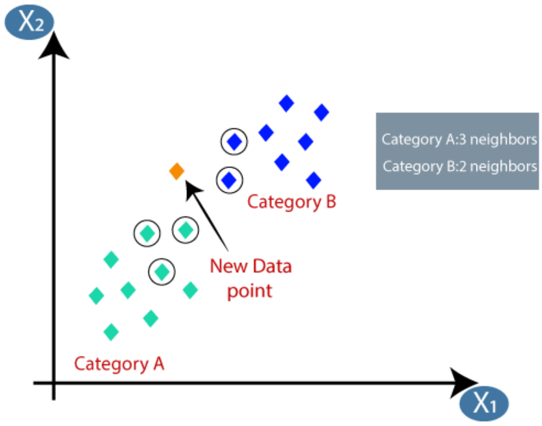

Fig 1. Basic KNN Architecture

The "k" in the KNN method denotes how many nearest neighbours are taken into account when formulating a prediction or drawing an inference. KNN determines the separation between a fresh, unlabelled data point and each labelled data point in the training set. Then, depending on their distances, the k neighbours that are the new data point's closest neighbours are determined. In classification tasks, the new data point is given the class label that appears the most frequently among the k neighbours.

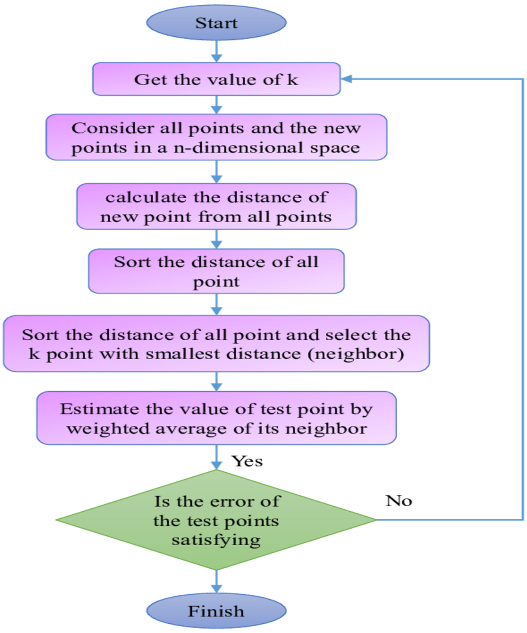

Fig 2. KNN algorithm

3.6Convolutional Neural Network (CNN)

ConvNet, also known as a Convolutional Neural Network (CNN), is a particular kind of artificial neural network created specifically for processing and analysing visual input, notably photographs. ConvNets have established themselves as a mainstay in computer vision applications thanks to their outstanding performance in tasks like picture segmentation, object

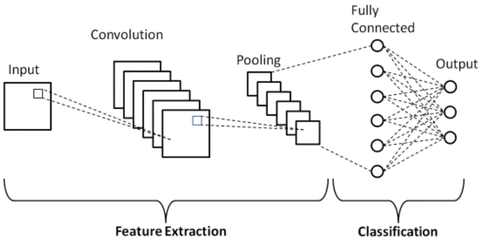

Fig 3. Basic CNN Architecture

Chapter 4

System Classification

4.1 Data Collection

The steps taken to prepare the model, create the data collection, and predict diseases are thoroughly explained in this section. In the beginning, our proposed system goes through the data collection phase, gathering both structured and unstructured data from various sources. After gathering the data, it is preprocessed and split into cleaning and test data sets. The accuracy of the prediction results is then improved by training machine learning algorithms like CNN and KNN over numerous iterations using the training data set. The constructed model is ready for testing once the intended target is attained after multiple epochs. At this stage, the model is put to the test using a new set of data that was not used for training in order to assess how well it performs. The suggested model is prepared for use if it achieves the desired accuracy in test data.

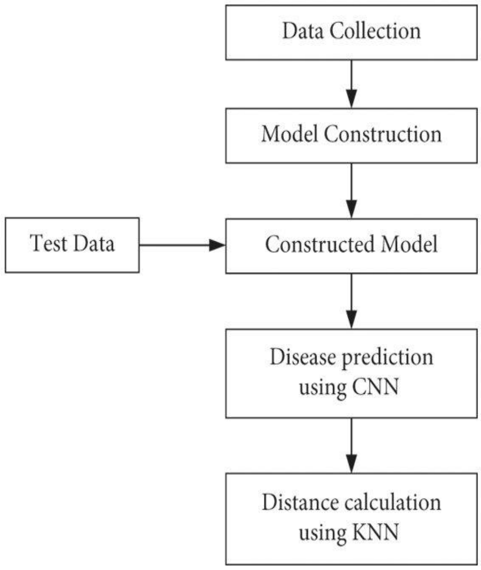

Fig. 4. Architecture for proposed disease detection system

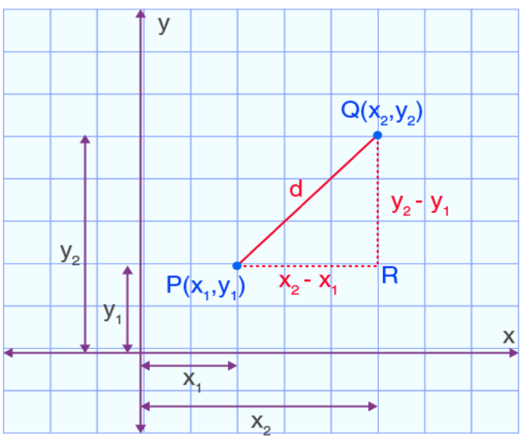

Fig 5: Graphical comparison of accuracy

Key Predictors:

Important Variables: The study identified several key predictors that consistently contributed to accurate disease outcome predictions across different models. Variables such as age, BMI, blood pressure, blood glucose levels, and cholesterol levels were among the most significant predictors.

Feature Importance: Ensemble methods like random forests and gradient boosting machines provided insights into feature importance, revealing that variables related to metabolic health (e.g., HbA1c for diabetes prediction) and lifestyle factors (e.g., smoking status) were critical in determining disease outcomes.

References

1. Battineni G., Sagaro G. G., Chinatalapudi N., Amenta F. Applications of machine learning predictive models in the chronic disease diagnosis. *Journal of Personalized Medicine* . 2020;10(2):p. 21. doi: 10.3390/jpm10020021. [PMC free article] [PubMed] [CrossRef] [Google Scholar]

2. Manjulatha B., Suresh P. *Smart Computing Techniques and Applications* . New York, NY, USA: Springer; 2021. An ensemble model for predicting chronic diseases using machine learning algorithms; pp. 337–345. [CrossRef] [Google Scholar]

3. Jen C.-H., Wang C.-C., Jiang B. C., Chu Y.-H., Chen M.-S. Application of classification techniques on development an early-warning system for chronic illnesses. *Expert Systems with Applications* . 2012;39(10):8852–8858. doi: 10.1016/j.eswa.2012.02.004. [CrossRef] [Google Scholar]

4. Gupta D., Khare S., Aggarwal A. A method to predict diagnostic codes for chronic diseases using machine learning techniques. Proceedings of the2016 International Conference on Computing, Communication and Automation (ICCCA); April 2016; Greater Noida, India. IEEE; pp. 281–287. [CrossRef] [Google Scholar]

5. Chen M., Yixue H., Hwang K., Wang L, Wang L. Disease prediction by machine learning over big data from healthcare communities. *Ieee Access* . 2017;5:8869–8879. doi: 10.1109/access.2017.2694446. [CrossRef] [Google Scholar]

6. Ge R., Zhang R., Wang P. Prediction of chronic diseases with multi-label neural network. *IEEE Access* . 2020;8:138210–138216. doi: 10.1109/access.2020.3011374. [CrossRef] [Google Scholar]