DL BASED FRAMEWORK FOR CARDIOVASCULAR DISEASES RISK PREDICTION IN IMBALANCED BIG DATA

A PROJECT REPORT

Submitted by

SARANKUMAR B

SELVA M

SUKESH B

SUNDARESWARAN N

in partial fulfillment for the award of the degree

of

BACHELOR OF TECHNOLOGY

IN

INFORMATION TECHNOLOGY

PSNA COLLEGE OF ENGINEERING AND TECHNOLOGY,

(An Autonomous Institution Affiliated to Anna University, Chennai)

DINDIGUL-624622

MAY 2024

ABSTRACT

Cardiovascular Diseases are the leading cause of death globally. Early and accurate prediction is crucial for preventive measures. Big data, containing electronic health records (EHRs) and other patient information, offers vast potential for CVD risk prediction using Deep Learning (DL) models. However, real-world medical datasets are often imbalanced, with a much lower prevalence of CVD cases compared to healthy individuals. Traditional machine learning algorithms struggle with such imbalanced data, leading to biased predictions favoring the majority class.

Existing CVD risk prediction methods often rely on traditional machine learning algorithms. These algorithms can be less effective with high-dimensional big data and may not capture the complex relationships between various risk factors. Additionally, imbalanced datasets present a significant challenge as these algorithms tend to prioritize the majority class, leading to inaccurate predictions for the minority class (i.e., CVD cases).

The proposed Long-Short Term Memory (LSTM) model in CVD prediction based on attention mechanism. The proposed model can learn the importance of each past value to the current value from the long sequence of CVD data at the past moment, which makes it possible to extract more valuable features. In order to solve the problem of low accuracy of other models.