Diagnosis of Cardiovascular Diseases by Ensemble Optimization Deep Learning Techniques

David Opeoluwa Oyewola, Department of Mathematics and Statistics, Federal University, Kashere, Nigeria
Emmanuel Gbenga Dada, Department of Computer Science, Faculty of Physical Sciences, University of Maiduguri, Nigeria
Sanjay Misra, Department of Applied Data Science, Institute for Energy Technology, Halden, Norway*
https://orcid.org/0000-0002-3556-9331

ABSTRACT

Cardiovascular disease (CVD) is a variety of diseases that affect the blood vessels and the heart. The authors propose a set of deep learning inspired by the approach used in CVD support centers for the early diagnosis of CVD using deep learning techniques. Data were collected from patients who received CVD screening. The authors propose a prediction model to diagnose whether people have CVD or not and to provide awareness or diagnosis on that. The performance of each algorithm is compared with that of long-, short-time memory, feedforward, and cascade forward neural networks, and Elman neural networks. The results show that the ensemble deep learning classification and prediction model achieved 98.45% accuracy. Using the proposed early diagnosis model for CVD can help simplify the diagnosis of CVD by medical professionals.

KEYWORDS

Cardiovascular Disease, Cascade Forward Neural Network, Elman Neural Network, LSTM

1. INTRODUCTION

One of the noticeable infections that affect numerous people during mature age is cardiovascular disease (CVD), and in most cases, it eventually leads to deadly difficulties (Thomas et al., 2018). CVD is the disease of the heart or blood vessels. CVD caused 17.6 million deaths in 2016, which increased to14.5% from 2006 to 2016 (Naghavi et al., 2017). CVD mortality and dismal state are increasing yearly, particularly in developing regions. Re- search has indicated that roughly 80% of CVD-related deaths occur in middle-class nations. Moreover, these fatalities occur at a young age among people in high-income nations (Gersh et al., 2010). In developing nations, rapid financial change leads to ecological changes and unhealthy ways of life; Furthermore, aging of the population may increase CVD risk factors and increase the occurrence of CVD (Wu et al., 2016). Based on WHO data, 24% of deaths from nontransferable diseases in India are due to heart diseases (Mackay & Mensah, 2004;

DOI: 10.4018/IJHISI.334021
*Corresponding Author

Vanisree, 2011). A fraction of deaths in the United States and other industrialized nations are due to heart diseases (Patil & Kumaraswamy, 2009). Around 17 million people die from CVD every year around the world, and the diseases are extremely widespread in Asia (Patil & Kumaraswamy, 2009; Vasighi et al., 2013). The specific reason for CVD is unclear, but there are threat factors that are hypertension, high cholesterol levels, diabetes, obesity, age, gender, smoking, and alcohol abuse, and others (Sudhakar, 2014). There are four types of CVD, that is, coronary heart disease, strokes and transient ischemic attack, peripheral arterial ailment, and aortic ailment, but it is difficult to predict CVD chances based on risk factors (Ali, 2013).

Artificial intelligence (AI) strategies can be utilized for the prediction of heart-related problems such as Support Vector Machines (SVM), Naïve Bayes (NB), Neural Networks (NN), Regression (Lauraitis et al., 2019; Naz et al., 2021; Šalkevicius et al., 2019; Wu et al., 2019; Xing et al., 2007). Associative classifiers achieve high precision and robustness compared to conventional classifiers, even in the handling of unstructured data (Miao & Miao, 2018). An improved deep neural network (DNN) was created by (Li et al., 2019) to help people and healthcare professionals and to increase the accuracy and unwavering quality of the medical diagnosis of heart disease. DNN techniques depend on multilayer perceptron (MLP) design with regularization and dropout utilizing deep learning. The evolved DNN techniques consist of a classification model based on training data and a forecast model for identifying new patient cases utilizing information collected from the Cleveland Clinic Foundation which consists of 303 clinical occurrences from patients with coronary heart problems. To analyze the extent of CVD ailment in patients using information from 423,604 people without CVD at baseline, UK Biobank developed the AI technique to predict CVD risk based on 473 accessible factors (Alaa et al., 2019). Machine learning techniques were derived utilizing AutoPrognosis, a tool that chooses and tunes ensembles of the learning model. They contrasted their techniques with a reputable risk forecast technique based on routine CVD threat factors (Framingham score), Cox proportional hazards (PH) techniques based on recognizable threat factors (such as age, smoking, history of diabetes, systolic blood pressure and body mass index), and Cox PH techniques based on all the 473 accessible factors. The enhanced risk prediction techniques of AutoPrognosis have a low performance in terms of the area under the receiver operating characteristic curve (AUC ROC) and the CI compared to the Framingham score. The Cox PH model with traditional risk factors for AUC-ROC of 0.734 and the Cox PH model with all UK Biobank factors have a low AUC-ROC performance of 0.758. Moreover, their techniques proved risk prediction in high-risk population groups, such as people with a history of diabetes mellitus.

Their findings showed that their proposed technique has low performance in terms of sensitivity, precision, diagnostic accuracy, and diagnostic odds ratio. Therefore, there is a need to develop a high-performance, robust, and efficient method of diagnosis. In this paper, we propose an ensemble deep learning technique to predict CVD based on risk factors.

2. RELATED WORKS

Several studies have been done in the area of machine learning for cardiovascular disease diagnosis. Uddin and Halder (Uddin & Halder, 2021) proposed a multilayer dynamic system (MLDS) based on the ensemble approach, which may improve understanding in each layer. For feature extraction, the proposed model uses the Correlation Attribute Evaluator (CAE), the Gain Ratio Attribute Evaluator (GRAE), the Information Gain Attribute Evaluator (IGAE), Lasso, and the Extra Trees classifier (ETC). Lastly, the ensemble approach for classification in the model was built using Random Forest (RF), Nave Bayes (NB), and Gradient Boosting (GB) classifiers. While the base classifiers described failed to identify accurately in any layer, the K Nearest Neighbor (KNN) technique finds the test data's neighborhood data points. The study made use of a Kaggle dataset with 70,000 instances. The proposed model achieved a maximum accuracy of 94.16%. Also, it achieved a 0.94 AUC score, indicating that it has a 94 percent chance of accurately identifying positive and negative classes, with an 87.5

Related Content

Security Improvements for Better and Safer Cross-Border ePrescription and Patient Summary Services

Dimitrios G. Katehakis, George Pangalosand Andriana Prentza (2017). *International Journal of Reliable and Quality E-Healthcare (pp. 18-28).*
www.irma-international.org/article/security-improvements-for-better-and-safer-cross-border-eprescription-and-patient-summary-services/164996