Article

A web-based heart disease prediction system using machine learning algorithms

Md. Mahbubur Rahman, **Morshedur Rahman Rana**, **Md. Nur-A-Alam**, **Md. Saikat Islam Khan**, **Khandaker Mohammad Mohi Uddin**

Department of Computer Science and Engineering, Dhaka International University, Dhaka-1205, Bangladesh

E-mail: mahbubur.cse@diu.ac, ranamorshedurrahman@gmail.com, munnacse44@gmail.com, saikat.cse@diu.ac, jilanicsejnu@gmail.com

Received 23 February 2022; *Accepted 11 March 2022*; *Published 1 June 2022*

Abstract

Disease diagnosis is the most critical task in the medical diagnosis system. At present, the biggest challenge is to predict heart disease very quickly; for that limitation, the number of dying people is increasing day by day. If a heart disease is diagnosed quickly, we can reduce the death rate indisputably. Thus, this research produces a manual and web-based automatic prediction system that can confer a conceptual report of clear warning of patient's heart condition. The proposed prediction system predicts heart disease using some health parameters. The system uses thirteen health parameters like age, sex, chest pain type, blood pressure, ECG, etc. Eight algorithms are used separately to diagnose heart disease accurately, namely KNN, XgBoost, Logistic Regression (LR), Support Vector Machine (SVM), Ada Boost, Decision tree (DT), Naïve Bayes, and Random Forest (RF). Decision Tree and Random Forest provide better performance than others among all methods. This research also established a website to easily check their heart condition from home instantly. The system has used 1026 individual patients' data for training and testing. It achieves higher accuracy in the different algorithms such as DT (99%), RF (99%), XgBoost (95%), KNN (89%), SVM (85%), LR (85%), Ada Boost (83%) and Naïve Bayes (82%). The experiment result provides a target value of 0 or 1 that refers to the patient's presence or absence of heart disease.

Keywords heart disease prediction; machine learning algorithms; web application; health parameters.

Network Biology
ISSN 2220-8879
URL: http://www.iaees.org/publications/journals/nb/online-version.asp
RSS: http://www.iaees.org/publications/journals/nb/rss.xml
E-mail: networkbiology@iaees.org
Editor-in-Chief: WenJun Zhang
Publisher: International Academy of Ecology and Environmental Sciences

1 Introduction

Cardiovascular Disease (CVD) is one of the leading positions holding diseases in the race of global death in developed countries. The mortality rate due to cardiovascular attack is increasing unusually in high-income regions and posing a significant challenge in the health sector (Faizal et al. 2021; Bantham et al., 2021). According to the Survey of National Surveillance System for Cardiovascular Disease (2021), the age-adjusted prevalence of all categories of heart disease is 7%, where coronary artery disease is higher in males (8.3%)

Table 4 narrates the existing methodologies with their taken attributes and accuracies. From 2006 to now, researchers have been trying to predict heart disease with a computer-aided system to reduce time and cost-efficiently.

Table 4 A comparative studies on various proposed algorithms.

Author	Proposed Methodology	Parameters/ Features	Accuracy
Alim et al. (2020)	Prediction of heart disease was measured with the help of a support vector machine and kernel equivalent to it. ***Techniques used:*** Hoeffding tree method	Not defined	86.94%
Kanwal et al. (2021)	They use Co-Active Neuro-Fuzzy implication method (CANFIS) and later combined with the genetic algorithm to identify heart disease. ***Techniques used:*** Genetic Algorithm,NN, LR,,SVM	14	96%
Khan et al. (2021)	A heart diseases prediction system (HDPS) based on the data mining approaches. ***Techniques used:*** Naïve Bayes, J48 DT, NN, RF.	14	97.70%
Acharya et al.(2015)	A proposed method on heart rate variability signals using data mining techniques where DT provides the highest accuracy. ***Techniques used:*** Naïve Bayes, KNN, DT	8	92.02%
Dwivedi (2018)	Using an artificial neural network and support vector machine, the proposed method predicts stroke patients where SVM provides the highest accuracy. ***Techniques used:*** ANN, SVM.	8	85%
Ayon et al. (2020)	Logistic Regression (LR), Support Vector Machine (SVM), Deep Neural Network (DNN), Decision Tree (DT), Nave Bayes (NB), Random Forest (RF), and K-Nearest Neighbor (K-NN) were the seven computational intelligence techniques used. ***Techniques used:*** DT, RF, NN, Logistic Regression, SVM, Naïve Bayes	9	98.15%
Beyene et al. (2018)	Propose methodology evaluate performance using tenfold cross-validation to predict heart disease. In this method, logistic regression provides the highest accuracy. ***Techniques used:*** Naïve Bayes, SVM, Classification Tree, Logistic Regression, KNN, ANN	12	Not defined
Katarya et al. (2020)	Chronic prediction system using data mining techniques where Naïve Bayes, Decision Tree provide the highest accuracy. ***Techniques used:*** Naïve Bayes, DT, SVM	10	95.56%
Motarwar et al. (2020)	Cardiovascular disease prediction using data mining techniques. Simple Cart provides the highest accuracy. ***Techniques used:*** Naïve Bayes, J48, Simple CART	Not Defined	92.2%

Farzana et al. (2020)	The proposed model predicts heart disease using classification techniques where the SVM technique is more effective and efficient than other data mining algorithms. ***Techniques used:*** Naïve Bayes, KNN Associate Rule, , SVM, ANN, DT	13	95%
Ismail et al. (2020)	A conceptual method to enhance prediction of heart disease using big data where SVM provides the highest accuracy. ***Techniques used:*** Naïve Bayes, SVM	13	90.6%
Sharma et al. (2020)	Data mining techniques predict heart diseases. ***Techniques used:*** Naïve Bayes, J48, SVM	13	90.78%
Proposed method	Heart disease prediction system from web and manual using machine learning classification. The highest accuracy is obtained from the Decision Tree and Random Forest**.** ***Techniques used:*** DT, RF, Naïve Bayes, KNN, AdaBoost, SVM, Logistic Regression, XgBoost	**13**	**99%**

6 Conclusion and Future Work

The research paper shows the overviews of existing methodologies and literature review of heart disease prediction systems which helps us to improve our method. In our approach, using the heart patients dataset from Alim et al. (2020), we analyzed different machine learning classification algorithms to predict the heart disease of any patient manually and on the web platform. The analysis shows 99% accuracy in Decision Tree and Random Forest techniques than other algorithms. Decision Tree (0.0121) has a minor classification error between these methods than Random Forest (0.0146). Further extension of this work is to get 100% accuracy to detect heart disease using more updated machine learning techniques. For enhancing user accessibility, research will be extended by developing an android app.

Abbreviations

BP, Blood Pressure; PA, Physical Exercise; RF, Random Forest; LR, Logistic Regression; HDPS, Heart Disease Prediction System; ANN, Artificial Neural Network; FP, false positive; SVM, support vector machine; TP, true positive; DT, Decision Tree.

Acknowledgement

We are grateful to Professor Dr. H. I. Lutfor Rahman Khan, Head of Cardiology, Dhaka Medical College and Hospital, for his unswerving loyalty and continuous support.

References

Acharya UR, Vidya KS, Ghista DN, Lim WJE, Molinari F, Sankaranarayanan M. 2015. Computer-aided diagnosis of diabetic subjects by heart rate variability signals using discrete wavelet transform method. Knowledge-Based Systems, 81: 56-64

Ali MM, Paul BK, Ahmed K, Bui FM, Quinn JM, Moni MA. 2021a. Heart disease prediction using supervised machine learning algorithms: performance analysis and comparison. Computers in Biology and Medicine, 136: 104672

Ali L, Bukhari SAC. 2021b. An approach based on mutually informed neural networks to optimize the generalization capabilities of decision support systems developed for heart failure prediction. Irbm, 42(5): 345-352

Alim MA, Habib S, Farooq Y, Rafay A. 2020. Robust Heart Disease Prediction: A Novel Approach Based On Significant Feature and Ensemble Learning Model. In: 2020 3rd International Conference on Computing, Mathematics and Engineering Technologies (ICoMET). IEEE

Al-Yarimi FAM, Munassar NMA, Bamashmos MHM, Ali MYS. 2021. Feature optimization by discrete weights for heart disease prediction using supervised learning. Soft Computing, 25(3): 1821-1831

Ayon SI, Islam MM, Hossain MR. 2020. Coronary artery heart disease prediction: a comparative study of computational intelligence techniques. IETE Journal of Research, 1-20

Aman RSC. 2021. Analyzing predictive algorithms in data mining for cardiovascular disease using WEKA tool. International Journal of Advanced Computer Science and Applications, 12(8)

Balakrishnan M, Christopher AA, Ramprakash P, Logeswari A., 2021. February. prediction of cardiovascular disease using machine learning. Journal of Physics: Conference Series, 1767(1): 012013

Bantham A, Ross SET, Sebastião E, Hall G. 2021. Overcoming barriers to physical activity in underserved populations. Progress in Cardiovascular Diseases, 64: 64-71

Bari Antor M, Jamil AHM, Mamtaz M, Monirujjaman Khan M, Aljahdali,S, Kaur M, Singh P, Masud M. 2021. A comparative analysis of machine learning algorithms to predict Alzheimer’s disease. Journal of Healthcare Engineering, 2021

Bergamini M, Iora PH, Rocha TAH, Tchuisseu YP, Dutra ADC, Scheidt JFHC, Nihei OK, de Barros Carvalho, MD, Staton CA, Vissoci JRN, de Andrade L. 2020. Mapping risk of ischemic heart disease using machine learning in a Brazilian state. Plos one, 15(12): e0243558

Beyene C, Kamat P. 2018. Survey on prediction and analysis the occurrence of heart disease using data mining techniques. International Journal of Pure and Applied Mathematics, 118(8): 165-174

Budholiya K, Shrivastava SK, Sharma V. 2020. An optimized XGBoost based diagnostic system for effective prediction of heart disease. Journal of King Saud University-Computer and Information Sciences. https://doi.org/10.1016/j.jksuci.2020.10.013

Builder V. 2021. Cardiovascular Pathologies and Disorders. Mosby's Pathology for Massage Professionals E-Book, Elsevier

Charbuty B, Abdulazeez A. 2021. Classification based on decision tree algorithm for machine learning. Journal of Applied Science and Technology Trends, 2(1): 20-28

Dwivedi AK. 2018. Performance evaluation of different machine learning techniques for prediction of heart disease. Neural Computing and Applications, 29(10): 685-693

Eremiasova L, Hubacek JA, Danzig V, Adamkova V, Mrazova L, Pitha J, Lanska V, Cífková R, Vitek L. 2020. Serum bilirubin in the Czech population―Relationship to the risk of myocardial infarction in males. Circulation Journal, 84(10): 1779-1785

Faizal ASM, Thevarajah TM, Khor SM, Chang SW. 2021. A review of risk prediction models in cardiovascular disease: conventional approach vs. artificial intelligent approach. Computer Methods and Programs in Biomedicine, 207: 106190

Farzana S, Veeraiah D. 2020, October. Dynamic Heart Disease Prediction using Multi-Machine Learning Techniques. In: 2020 5th International Conference on Computing, Communication and Security (ICCCS). IEEE

Ismail A, Abdlerazek S, El-Henawy IM. 2020. Big data analytics in heart diseases prediction. Journal of Theoretical and Applied Information Technology, 98(11): 15-19

Kanwal S, Rashid J, Nisar MW, Kim J, Hussain A. 2021. July. An Effective Classification Algorithm for Heart Disease Prediction with Genetic Algorithm for Feature Selection. In: 2021 Mohammad Ali Jinnah University International Conference on Computing (MAJICC). IEEE

Kavitha M, Gnaneswar G, Dinesh R, Sai YR, Suraj RS. 2021, Heart disease prediction using hybrid machine learning model. In: 2021 6th International Conference on Inventive Computation Technologies (ICICT). 1329-1333, IEEE

Katarya R, Srinivas P. 2020. Predicting Heart Disease at Early Stages Using Machine Learning: A Survey. In: 2020 International Conference on Electronics and Sustainable Communication Systems (ICESC). 302-305, IEEE

Khan M, Rahman A, Rahman M, Salehin JU, Islam M, Rabbi M. 2021. Efficient data mining techniques for heart disease prediction and comparative analysis of classification algorithms. Asian Journal of Research in Computer Science, 2021: 239056682

Liou L, Kaptoge S. 2020. Association of small, dense LDL-cholesterol concentration and lipoprotein particle characteristics with coronary heart disease: A systematic review and meta-analysis. Plos One, 15(11): e0241993

Mehmood A, Iqbal M, Mehmood Z, Irtaza A, Nawaz M, Nazir T, Masood M. 2021. Prediction of heart disease using deep convolutional neural networks. Arabian Journal for Science and Engineering, 46(4): 3409-3422

Motarwar P, Duraphe A, Suganya G, Premalatha M. 2020. Cognitive Approach for Heart Disease Prediction Using Machine Learning. In: 2020 International Conference on Emerging Trends in Information Technology and Engineering (ICETITE). IEEE

Mostafiz R, Uddin KMM, Uddin MS, et al. 2021. Diagnosis of diabetes: A machine learning paradigm using optimized features. Network Biology, 11(3): 222-240

Mohan S, Thirumalai C, Srivastava G. 2019. Effective heart disease prediction using hybrid machine learning techniques. IEEE Access, 7: 81542-81554

Parambi DGT, Unnikrishnan MK, Marathakam A, Mathew B. 2020. Demographic and Epidemiological Aspects of Aging. In: Nutrients and Nutraceuticals for Active & Healthy Ageing. Springer, Singapore

Rani KU. 2011. Analysis of heart diseases dataset using neural network approach. arXiv preprint arXiv:1110.2626

Riyaz L, Butt MA, Zaman M, Ayob O. 2022. Heart Disease Prediction Using Machine Learning Techniques: A Quantitative Review. In: International Conference on Innovative Computing and Communications. 81-94, Springer, Singapore

Samhitha BK, Priya MS, Sanjana C, Mana SC, Jose,J. 2020. Improving The Accuracy in Prediction of Heart Disease Using Machine Learning Algorithms. In: 2020 International Conference on Communication and Signal Processing (ICCSP). 1326-1330, IEEE

Shrivastava K, Jotwani V. 2021. A Comparative Analysis of Various Data Mining Techniques to Predict Heart Disease. Expert Clouds and Applications: Proceedings of ICOECA 2021. Springer

Sharma S, Parmar M. 2020. Heart diseases prediction using deep learning neural network model. International

Journal of Innovative Technology and Exploring Engineering, 9(3): 2244-2248

Singh A, Kumar R. 2020, February. Heart disease prediction using machine learning algorithms. In: 2020 International Conference on Electrical and Electronics Engineering (ICE3). 452-457, IEEE

Taneja A. 2013. Heart disease prediction system using data mining techniques. Oriental Journal of Computer Science and Technology, 6(4): 457-466

Terrada O, Hamida S, Cherradi B, Raihani A, Bouattane O. 2020. Supervised machine learning based medical diagnosis support system for prediction of patients with heart disease. Advances in Science, Technology and Engineering Systems Journal, 5(5): 269-277

Varshini KS, Uthra RA. 2021. An effectual method for disease identification in pediatric dataset. Materials Today: Proceedings. http://dx.doi.org/10.1016/j.matpr.2021.03.317

Willinger L, Brudy L, Meyer M, Oberhoffer-Fritz R, Ewert P, Müller J. 2021. Prognostic value of non-acute high sensitive troponin-T for cardiovascular morbidity and mortality in adults with congenital heart disease: A systematic review. Journal of Cardiology, 78(3): 206-212

Yadav AS, Sharm P, Swami A, Pandey G. 2018. A supply chain management of chemical industry for deteriorating items with warehouse using genetic algorithm. Selforganizology, 5(1-2): 1-9

Yadav AS, Swami A, Kher G. 2019. Blood bank supply chain inventory model for blood collection sites and hospital using genetic algorithm. Selforganizology, 6(3-4): 13-23

Yadav DC, Pal S. 2020. Prediction of heart disease using feature selection and random forest ensemble method. International Journal of Pharmaceutical Research, 12(4): 56-66

Yuyun MF, Sliwa K, Kengne AP, Mocumbi AO, Bukhman G. 2020. Cardiovascular diseases in sub-Saharan Africa compared to high-income countries: an epidemiological perspective. Global Heart, 15(1): 1-18