Retraction

Retraction: Predicting heart disease using hybrid machine learning model (J. Phys.: Conf. Ser. 1916 012208)

Published 23 February 2022

This article (and all articles in the proceedings volume relating to the same conference) has been retracted by IOP Publishing following an extensive investigation in line with the COPE guidelines. This investigation has uncovered evidence of systematic manipulation of the publication process and considerable citation manipulation.

IOP Publishing respectfully requests that readers consider all work within this volume potentially unreliable, as the volume has not been through a credible peer review process.

IOP Publishing regrets that our usual quality checks did not identify these issues before publication, and have since put additional measures in place to try to prevent these issues from reoccurring. IOP Publishing wishes to credit anonymous whistleblowers and the Problematic Paper Screener [1] for bringing some of the above issues to our attention, prompting us to investigate further.

[1] Cabanac G, Labbé C and Magazinov A 2021 arXiv:2107.06751v1

Retraction published: 23 February 2022
Predicting heart disease using hybrid machine learning model

G Renugadevi¹, G Asha Priya¹, B Dhivyaa Sankari¹, R Gowthamani¹
¹ Department of Computer Science and engineering, Sri Krishna College of Engineering and Technology, Coimbatore, Tamil Nadu, India.
grenugadevi@skcet.ac.in, 17eucs019@skcet.ac.in, 17eucs045@skcet.ac.in, gowthamanir@skcet.ac.in

Abstract. Multiple Chronic disease are available especially Heart disease is the foremost reasons of death in modern world. Machine learning (ML) is useful for making conclusions and predictions based on a huge volume of data formed by the healthcare industry. The proposed approach uses machine learning techniques to find heart disease in this study. The prediction model, which employs classification techniques, is based on the Cleveland heart dataset. The Random Forest and Decision Tree machine learning techniques are used. This model for heart ailment with hybrid methodology has an accuracy level of 88.7%, according to experimental study. The boundary is determined as an input parameter from the user to predict heart disease using a Decision Tree method and Random Forest hybrid methodology.

1. INTRODUCTION

The procedure of mining is functional knowledge since huge data and using techniques like classification, clustering, and association to predict or describe the data. The accuracy of events related to heart disease are estimated using decision tree. Clinical records such as Left Bundle Branch Block (LBBB), Right Bundle Branch Block (RBBB), Atrial Fibrillation (AFIB), Normal Sinus Rhythm (NSR), Sinus Bradycardia (SBR), Atrial utter (AFL), Premature Ventricular Contraction (PVC), and Second-degree block to assess patient's specific state in correlation to heart catastrophe (BII) is used. For classification, a radial basis function network (RBFN) dataset is pre-owned., with 70% data being trained and 30% being used for classification.

Using the well-known Cleveland dataset, which can be initiate in a UCI machine learning repository, for experimental validation. The Neural Networks ML algorithm is implemented, which produces more accurate and consistent performance. For heart disease prediction, the proposed method has 13 attributes. When compared to current methods in use, the findings indicate a higher degree of efficiency. The use of neural network models, which integrate not only subsequent likelihoods but also predictable ideals from several predecessor techniques, is implemented. In the first stage, the grouping will differentiate between traffic bring about by IoT devices and traffic bring about by non-IoT devices. In second stage, each IoT device is assigned to distinct IoT device class.

The aim of this study is to upgrade the precision of heart disease prognosis results. Studies that carried out, with the result that trait limitations for computational use that have been identified. Findings of the
studies were used to spot the quality of a hybrid methodology. The results of the experiment indicate that our proposed hybrid approach is more capable of predicting heart disease than current method.

2. LITERATURE REVIEW
The PSO algorithm is used in this paper [1] to generate the best rubrics in the analysis of heart illness. The Particle Swarm Optimization (PSO) Algorithm, which is one of the most effective evolutionary algorithms, is used to produce heart disease laws. The random rules are encoded first, then optimized using the PSO algorithm based on their accuracy. Finally, the results are compared to the C4.5 algorithm.

A backpropagation neural network is used in this paper [2] to predict heart disease. A neural network is used to build a method for predicting heart disease. For heart disease prediction, the proposed method used 13 medical attributes. The proposed algorithm outperforms equivalent state-of-the-art methods, according to the results of the experiments described in this paper.

The Random Forest Classifier is used in this paper [3] to predict Coronary Heart Disease using a data mining model. This model will assist physicians in forecasting CHD and its mixed occurrences, as well as how they can be linked to different segments of the population. Angina, Acute Myocardial Infarction (AMI), Percutaneous Coronary Intervention (PCI), and Coronary Artery Bypass Graft surgery are several of the events studied (CABG). According to experimental findings, the Random Forest Classification algorithm has been effective in predicting CHD events and risk factors.

This paper [4] uses a conceptual heart illness prophecy classification using a random forest and evolutionary approach. The Chi-square trait selection compute is used to determine whether or not variables are correlated. This paper proposes a categorizing model that use a random forest as grouping, chi-square, and a genetic algorithm as attribute selection measures to anticipate heart disease. According to trial and error, our methodology improves precision as compared to other models, and can be successfully used by health care executive for foresee heart disease.

3. REQUIREMENT ANALYSIS

3.1 Data visualization
It's easier to grasp and interpret a vast volume of data when it's depicted graphically. Some employers demand that a data analyst be able to produce PowerPoint presentations, graphs, maps, and models as shown in Figure 1.

The heart disease rates are shown as a data visualization component in this method [5].

![Figure 1. Data Visualization of Heart disease rate](image)

3.2 Data Pre-Processing
Pre-processing is the method of converting raw data into a format suitable for machine learning. A data scientist can get more accurate results from a machine learning model when the data is structured and clean. Data formatting, washing, and sampling are all part of the process [6].
3.3 Dataset splitting
Machine learning datasets can be divided into three subsets: training, test, and validation sets. A training set is used to train a model and determine the best parameters for learning from data. A test set is needed for evaluating the trained model's generalization capability [7].

After being trained on training data, a model's ability to recognize trends in new, unseen data is referred to as the latter. To prevent model overfitting, or the inability to generalize, different subsets must be used for training and testing.

3.4 Model training
A model can be trained subsequently an information expert has pre-processed the composed information and divided it into train and test groups. The algorithm is "fed" with training data during this process. With predictive analysis, an algorithm can process data and produce a model that can identify a target value (attribute) in new information and give the response you want. Model development is the aim of model training.

4. PROPOSED METHODOLOGY
The proposed work is written in Python 3.6.4 and relies on libraries such as scikit-learn, pandas, matplotlib, and others. Binary heart disease classifications are included in the data. Algorithms such as decision trees and random forests are exploited in combination with the hybrid model.

Data Dictionary- The dataset included the attributes Age, Gender, Chest Pain, blood pressure level, cholesterol, fasting blood sugar, resting electrocardiographic results, maximum heart rate achieved, exercise induced angina, old peak, Slop, number of major vessels, thal, and pred

4.1 Decision Tree
A Supervised Learning Algorithm called a Decision Tree (DT) is used to solve classification problems. Both categorical and continuous input and output variables are supported. Depending on the greatest important splitter / discriminator in input variables, the sample is split into additional consistent sets in this technique.

The internal node constitutes a test on the attribute, the arm represents the outcome, and the leaf represents the decision made after evaluating the attribute in the decision tree.

Start from the root of the tree when predicting a class mark for a record in decision trees. The values of the root attribute are then compared to the values of the record's attribute.

Follow the branch that corresponds to that value and leap to the next node based on the comparison as shown in Figure 2.
4.2 Random Forest Model
1. Assume the training dataset contains \( n \) instances. Sub-samples are selected at random with substitution from these \( n \) instances. Individual trees are built using these random subsamples from the training dataset.
2. Given that there are \( k \) input variables, a number \( m \) is chosen such that \( m < k \). At each node, \( m \) variables are chosen at random from a pool of \( k \) variables. To split the node, the split that is the best of these \( m \) variables is chosen. While the forest grows, the value of \( m \) remains constant (Figure 4).
3. Each tree is allowed to grow to its full potential without being pruned.
4. The new object's class is predicted based on the majority of votes earned from all of the decision trees combined.

4.3 HYBRID MODEL
To construct a hybrid methodology, random forest probabilities are used in the combined model. The training data is combined with the random forest probabilities and cater into the decision tree model. In a similar way, decision tree prospect is frequently described and cater to test data. The predicted utility is then estimated (Figure 5).

5. EXPERIMENTAL RESULTS AND ANALYSIS
The findings indicate that both the Random Forest methodology and the hybrid methodology are successful at detecting heart disease. Random forest has an accuracy of 81 percent, Decision Tree has an accuracy of about 80 percent, and the Hybrid model has an accuracy of 80 percent as shown in Figure 3.
6. CONCLUSION AND FUTURE WORK
To improve the precision to foresee the onset of cardiac diseases at an early stage, more complex models and combinations of models are needed. The proposed scheme employs a Decision Tree method and Random Forest methodology combination to forecast heart disease. Train and test the method exploitation the Cleveland Heart Disease database to arrive at the most effective model. Several deep learning models, such as CNN or DNN algorithms, will be tested in the future for heart disease prediction. In order to determine the seriousness of the disease, it should also be described as a multi-class problem.

References
[1] Mackay J., Mensah G. 2004 Atlas of Heart Disease and Stroke Nonserial Publication, ISBN13 9789241562768 ISBN-10 9241562765.
[2] Robert Detrano 1989 Cleveland Heart Disease Database V.A. Medical Center, Long Beach and Cleveland Clinic Foundation.
[3] Yanwei Xing, Jie Wang and Zhihong Zhao Yonghong Gao 2007 Combination data mining methods
with new medical data to predicting outcome of Coronary Heart Disease Convergence Information Technology, 2007. International Conference November 2007, pp 868-872.

[4] A. Haldorai and A. Ramu, Security and channel noise management in cognitive radio networks, Computers & Electrical Engineering, vol. 87, p. 106784, Oct. 2020. doi:10.1016/j.compeleceng.2020.106784

[5] A. Haldorai and A. Ramu, Canonical Correlation Analysis Based Hyper Basis Feedforward Neural Network Classification for Urban Sustainability, Neural Processing Letters, Aug. 2020. doi:10.1007/s11063-020-10327-3

[6] Resul Das, Ibrahim Turkoglugb, and Abdulkadir Sengurb 2009 Effective diagnosis of heart disease through neural networks ensembles Expert Systems with Applications, pp 7675–7680.

[7] Robert Detrano, Andras Janosi, Johann-Jakob Schmid, Sarbjit Sandhu, Kern H. Guppy, Stella Lee, Victor Froelicher 1989 International application of a new probability algorithm for the diagnosis of coronary artery disease The American Journal of Cardiology, pp 304-310.