Proceedings of the 4th International Conference on Computing and Data Science (CONF-CDS 2022)
DOI: 10.54254/2755-2721/2/20220558

Pathological prediction of heart disease based on deep width learning algorithm

Mon Hian Chew[1,3], Po Yang[1], Zhengguang Zhao[2] and Yuehui Chen[1]

[1] School of Information Science and Engineering, University of Jinan, Jinan, Shandong, 250022, China
[2] Five-year Higher Vocational College, Shandong Chemical Engineering & Vocational College, Weifang, Shandong, 261108, China

[3]hian940202@gmail.com

Abstract. As the technology has become more and more advance and massive data has been produced and pre-processed became useful information to become big data so come out a lot of technology like data mining, neural network, deep learning, etc. Each technology has its advantages, this paper put forward the width of the depth of the learning algorithm is to the advantage of two types of deep learning methods This new method is combining of Flexible Neural Tree(FNT) and Broad Learning System(BLS) to become a new deep width network. After several times of experiments, we can confirm that this new hybrid method can increase the accuracy, sensitivity, and specificity of Heart disease prediction. Therefore, having higher accuracy of prediction for heart disease means we can notice more people before having heart disease and start to prevent it and in the end, we can prevent loss of life and economy.

Keywords: heart diseases, prediction model, broad learning system, flexible neural network, hybrid system.

1. Introduction

Heart disease is a kind of the highlight disease that will be easy get by the people the old or middle age and this situation has become more and more serious and promote younger staff. This is a bad phenomenon that will lead to fatal complications that cause loss of life [1] and finally, the economy will be affected. Based on the statistics come from WHO, the total number of deaths caused by noncommunicable diseases in India, 24 percent were caused by heart disease [2-3]. Heart disease accounts for one-third of global deaths [4]. In the US and other developed countries, 50 percent of people died of a heart attack [5]. Cardiovascular disease (CVD), which kills about 17 million people worldwide each year, is widespread in Asia[6-7].

A factual database for heart disease research is Cleveland Heart Disease Database (CHDD) [8]. Factors such as smoking, poor lifestyle habits, and obesity are considered risk factors for heart disease, while diabetes and high blood pressure, which are genetic risk factors, can also contribute to heart disease. Some risks are manageable. In addition to the above factors, obesity, dietary habits, and lack of exercise are also considered major influencing factors [9-10]. In the clinic, there are many different kinds of heart disease illness, main show is congenital heart disease, coronary heart disease, myocarditis, etc.

DOI: 10.54254/2755-2721/2/20220558

Human-based pathogenic risk factors to determine the risk of heart disease are very difficult and inaccurate [11]. But deep learning algorithms can be used to predict output based on existing data. Therefore, this paper uses one new idea from deep learning which is a deep-width learning algorithm to predict heart disease from risk factors. This paper proposed the hybrid system can improve the precision of predicting the risk of heart disease.

Many methods can be used for prediction problems such as C4.5, Random forest, Multilayer perceptron, Bayesian network, Naive Bayes, etc. From those methods mentioned above, the accuracy rate for this heart disease problem is different from each other because all of those methods can be used for prediction problems but not all of them are suitable for this prediction problem in this field or because some of the algorithms of the method are not suitable to process this types of data. Because they still haven't a perfect solution or a method for this prediction problem so the new method of deep learning which is a deep-width learning algorithm that a hybrid network which the combination of BLS with FNT has been purposed.

This combination has made it become a hybrid system and the reason it is a hybrid system is because BLS [12] and FNT [13] both are deep learning algorithms but BLS is expand its nodes to become width different from other deep learning algorithms but it also considers as the type of deep learning algorithms. Consider both of them are the same type but the nodes expand differently so we would like to call this a deep-width learning algorithm and also a hybrid system. Which can have better prediction results compared to other methods.

2. Materials and method

This section will have an explanation of the materials and method. The materials are the data that we need to use for the experiment and state out where the data came from. The data will have a full explanation in the below section. The method will have an explanation of the method of the hybrid system. The below part also has the simple structure of the network to let us easy to understand the networks.

2.1. Description of the dataset

The data used in this experiment comes from the Kaggle database. Kaggle is a subsidiary of Google LLC, an online community of data scientists and machine learning practitioners. Kaggle allows users to find and publish datasets, explore and build models in a web-based data science environment, collaborate with other data scientists and machine learning engineers, and participate in competitions to solve data science challenges.

Kaggle is the source of the Cleveland database and the dataset inside the Cleveland database used for this experiment. There are 14 attributes and 1025 instances in this dataset. There are 6 numeric attributes and 8 categorical attributes. The data information of the Cleveland dataset is shown in Table 1 below.

Table 1. Basic information of Cleveland.

PropertyName	Description	Range of Values
Age	Age	29 to 77
Gender	Gender	0, 1
Cp	Chest pain type	0, 1, 2, 3
Trestbps	Patient's resting blood pressure at admission(in mm Hg on admission to the hospital)	94 to 200
Chol	Serum cholestoral outfit (mg/dl)	126 to 564
Fbs	Fasting plasma glucose (mg/dl)	0, 1
Restecg	Resting ecg results	0, 1, 2
Thalach	HRmax	71 to 202
Exang	Angina caused by exercise	0, 1

DOI: 10.54254/2755-2721/2/20220558

Table 1. (continued).

OldPeak	ST depression caused by exercise relative to rest	0 to 6
Slope	Highest sports ST segment of slope	0, 1, 2
Ca	Fluorescent color number of the main blood vessels	0 to 3
Thal	Thalassemia	0 to 3
Target	Class Attribute	0 or 1

Based on the dataset, the patient's age is from 29 to 77. The gender value 0 is represented as female patients, and gender value 1 is represented as male patients. Chest pain is a sign of heart disease and can be divided into 4 types. The trestbps is resting blood pressure values. This chol is a cholesterol level reading. The fbs is the level of fasting blood sugar, 0 means the blood sugar level is over 120mg/dl, and 1 means below. The restecg is results of resting electrocardiogram, thalach is the maximum efficiency achieved by the patient's heart, exang indicates whether exercise-induced angina, where 0 indicates no pain and 1 indicates pain. The ca is the number of major vessels colored by fluoroscop, thal is the time of exercise test in minutes and the last is num which is class attributes. Here value 0 mean normal and 1 means patients that have been diagnosed with heart disease.

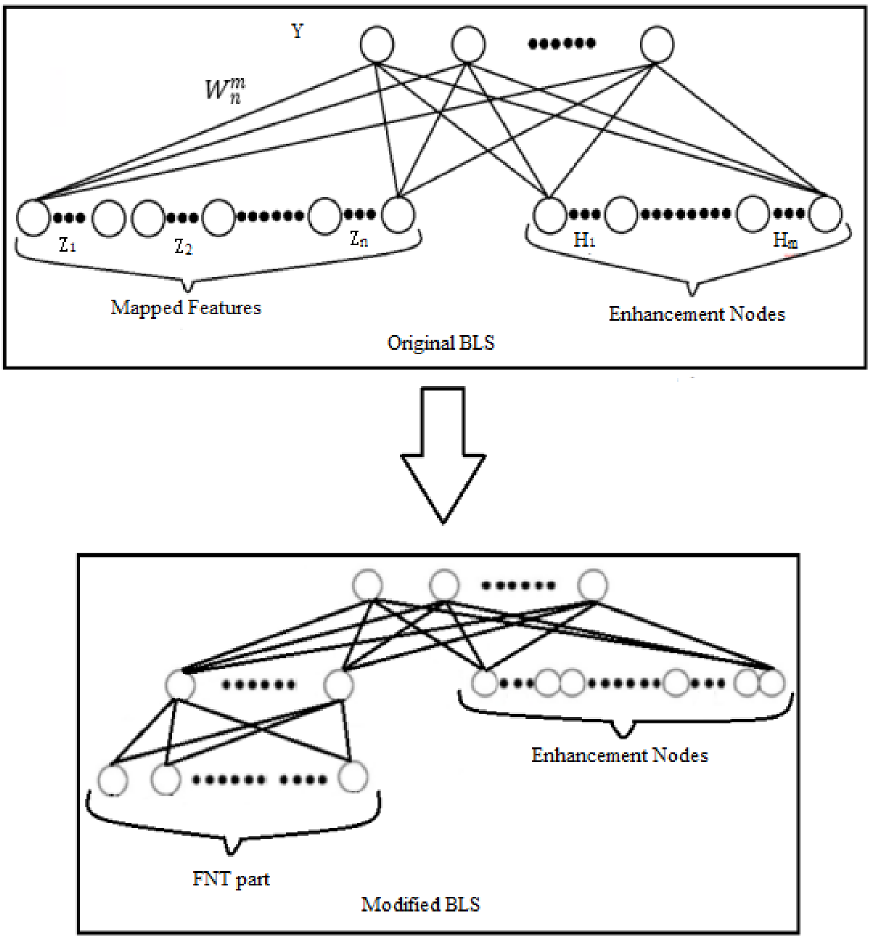

Figure 1. The change of original BLS to modified BLS.

2.2. *Deep-width learning algorithm*

This deep-width learning algorithm is the combination of FNT with BLS. Therefore, we use original BLS to do this heart disease datasets. The original BLS can do well in prediction problems, but the

DOI: 10.54254/2755-2721/2/20220558

prediction results are get based on the original feature sets so we have the idea to combine FNT and BLS. We found out that the formula of BLS in the mapped feature has a problem that we can improve. The mapped feature part weight is randomly generated and it will lower the prediction accuracy and slow convergence so we use FNT here as the initial solution to get good feature sets and improve the weight so it can get better accuracy. For the modified BLS, we replace the first half part which is the feature mapping part with FNT and the second half part of the modified BLS is remain unchanged as the original BLS.

The deep-width learning algorithm structure is shown in figure 1 above. From the figure, we can see that the first half part is a deep neural network and the next half part is a wide neural network. The second half part will use the results of the first half part to compute and then both parts are computed together and get the final result.

For the modified BLS, the changed formula is:

$$Z_i = \phi(XW_{e_i} + \beta_{e_i}), i = 1, \ldots, n \tag{1}$$

The whole formula 1 will change to a complete FNT that uses BP to optimize. It will become the first half part of the modified BLS.

The flexible neural tree model can automatically optimize the tree structure and parameters, and obtain better generalization ability with its sparse structure, but it also has problems. The flexible neural tree uses the traditional Particle Swarm Optimization (PSO) algorithm to optimize the variable parameters in the tree structure. The PSO algorithm is simple, easy to implement, and has no problems with hyperparameters. However, in the process of optimization, PSO is prone to fall into the problems of local optimum, slow convergence speed, and low precision in the later stage. To solve the above problems of the flexible neural tree model, we propose a flexible neural tree model fused with BP.

The Flexible Neural Tree-Neural Network (FNT-NN) model fused with BP uses the BP algorithm to further optimize the variable parameters in the FNT tree structure. The BP algorithm can continuously adjust the connection weights of each unit through the inverse propagation of errors until the error requirements are met. Using the BP algorithm to optimize the parameters of FNT is beneficial to improving the calculation accuracy of the FNT model.

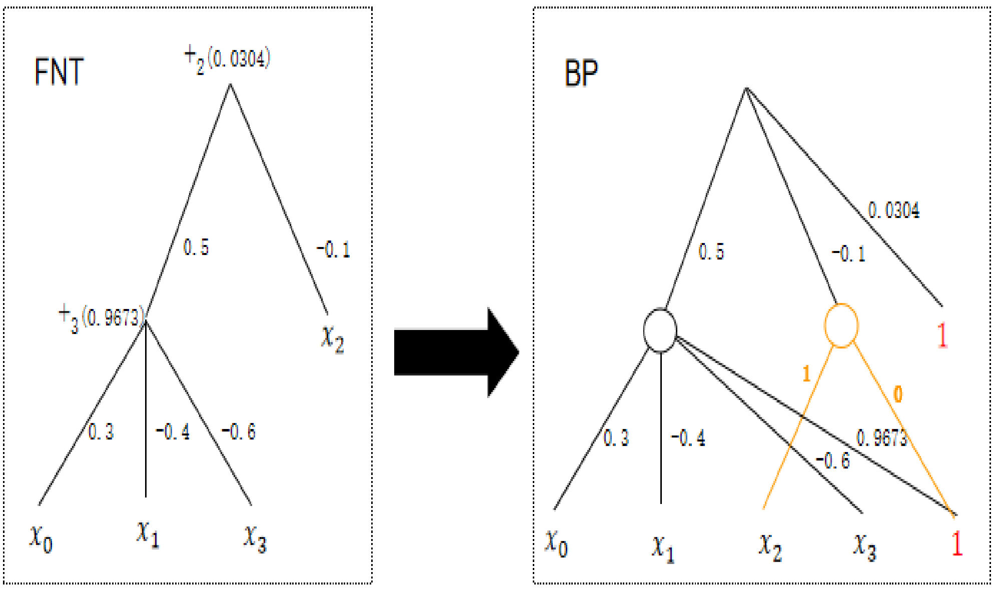

Figure 2. FNT converted to neural network structure.

DOI: 10.54254/2755-2721/2/20220558

A neural network is a fully connected network structure in which all input nodes in the network are at the same depth. On the other hand, FNT is a sparse tree structure, in which each leaf node cannot be guaranteed to be at the same depth. In order to use the BP algorithm to optimize the parameters in the flexible neural tree, it is necessary to convert the sparse tree structure into a fully connected neural network structure. In the experiments in this chapter, we convert the sparse tree structure of FNT into a fully-connected structure by using connection nodes with a bias of 0 and an input connection weight of 1 and set bias terms at each layer. For the connection weights that have been obtained in the FNT tree structure, they are directly placed in the corresponding positions of the neural network after construction. The conversion diagram of the network structure is shown in figure 2 below.

As shown in figure 3 below, given input and put into the FNT, the tree structure is designed and optimized through the GP algorithm, a neural tree is obtained by using the PSO optimization algorithm, and the corresponding neural network structure is constructed based on the obtained tree structure. The neural network performs parameter optimization to obtain a trained model.

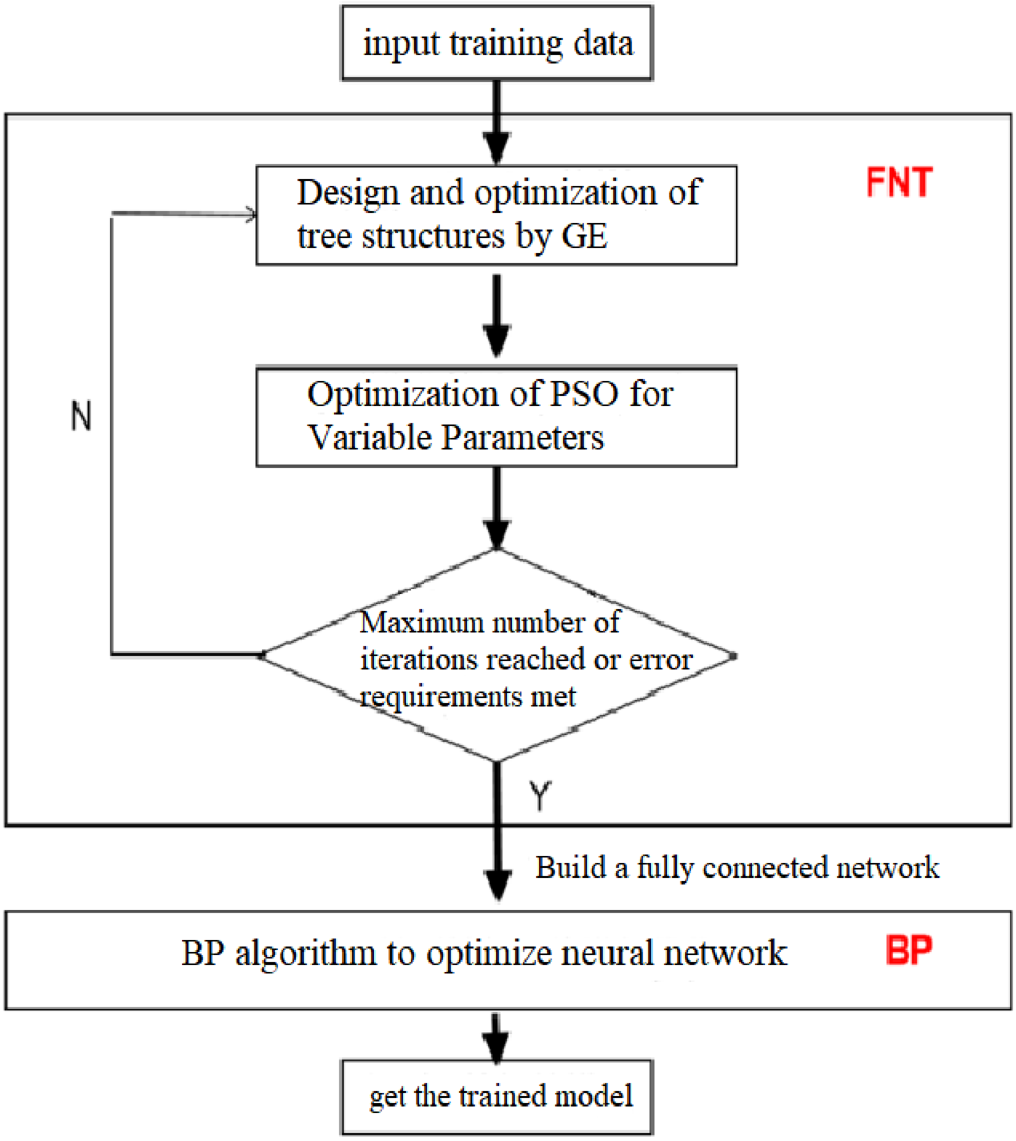

Figure 3. Flow chart of the flexible neural tree model fused with BP.

Modified BLS can do better prediction of heart disease because this modified BLS has the advantage of FNT and original BLS. Modified BLS can get the good feature sets and then mapping the low-level features containing low-category information into high-level features containing high-category information.

3. Experiment and results

3.1. Prepare work

The problem we met when combining FNT and BLS is the codes of FNT and BLS are different. The FNT is use Python to code but BLS is use Matlab to code. Therefore, we choose Python code for this deep-width algorithm. We chose Python because it was created as a generic language that is easy to read. When the code is easier to read, it is easier for us to understand the program. By changing the BLS Matlab code to Python code, we need to change the mapped feature part and replace it with FNT code.

3.2. Preprocessing of data

Although the dataset was downloaded we cannot directly use it because some of the attributes the value are too large and it will lower the prediction accuracy. This means we need to make all those attributes that have the values of more than one and preprocess them and make all of these attributes have the value between 0 to 1. For setting this problem, we just need to use normalization and then a new dataset that is suitable for the prediction model has been made.

3.3. Database setting

In this dataset, we got 1025 patients and we use half of it as the training set which has a total of 512 instances, and half of it as a testing set which has a total of 513 instances. Each instance in the training set and the testing set contains 13 input variables x and one output variable y.

3.4. Heart disease forecasting results

The hybrid method of the experimental results is shown in table 2 below. According to the results below, the hybrid system has achieved better performance. When the hybrid system is used for the heart disease prediction problem, the accuracy, specificity, and sensitivity are 93.66%, 93.17%, and 97.34%, respectively.

The heart disease prediction has been running by using 14 types of attributes. We randomly pick 50% of the patient's data for training and another 50% for the testing of the model. The accuracy of the random forest method is 85.8% [14]. The accuracy of the multilayer perceptron method is 80.9% [15]. The accuracy of Bayes net and Naive Bayes methods is 85.0% and 80.0% [16]. The result of all the methods is shown in table 3. Based on the table, we can know that the accuracy of the deep-width learning algorithm is higher than other methods, this makes it better able to deal with this kind of prediction problem.

Table 2. Hybrid system predicted results.

Prediction type	Percentage(%)
Accuracy	93.7
Specificity	93.1
Sensitivity	97.3

DOI: 10.54254/2755-2721/2/20220558

Table 3. The comparison between algorithms.

Algorithm	Accuracy(%)
C4.5	79.9
Random Forest	85.8
BLS	80.9
Multilayer Perceptron	80.9
Bayes Net	85.0
Naive Bayes	80.0
Deep Width Learning Algorithm	93.7

4. Discussions and conclusions

Heart disease prediction can help reduce the workload of doctors. If we can prevent the disease before it actually occurs, the accuracy of this prediction may save lives. Therefore, heart disease prediction has become an important topic, and that still exists and will continue to exist until now. If a predictive model can have 100% accuracy, it means that we humans can finally prevent this disease, and the workload of doctors can be saved from diagnosis.

In this article, we put forward a kind of deep-width of the hybrid learning algorithm and apply the method in the Cleveland dataset for heart disease. The accuracy, specificity, and sensitivity are 93.7%, 93.1%, and 97.3% respectively. We know that from the experiment result deep-width learning algorithm can better handle this prediction problem compared to other algorithms. And lastly, we still working on this algorithm with a different type of combination and hopefully can have another type of combination can have better results.

Acknowledgments

This work was supported by the Doctoral Foundation of University of Jinan (No.XBS1748, XBS1806), the National Natural Science Foundation of China (Grant No.61873324, 61671220, 61640218, 61201428), the Shandong Distinguished Middle -aged and Young Scientist Encourage and Reward Foundation, China (Grant No. ZR201 6FB14), the Project of Shandong Province Higher Educational Science and Technology Program, China (Grant No. J16LN07), the Shandong Province Natural Science Foundation (Grant No. ZR2018L F005, ZR201 9MF040), the Shandong Province Key Research and Development Program under Grant with No.201 8GGX101048, University Innovation Team Project of Jinan (2019GXRC015) Key Science&Technology Innovation Project of Shandong Province (2019JZZY010324).

References

[1] Heron M 2012 Deaths: leading causes for 2008 *National Vital Statistics Reports* 60(6):1-93.

[2] Mackay J and Mensah G 2004 The Atlas of heart disease and stroke *Journal of Human Hypertension* 19: 505.

[3] Chauhan S and Aeri B T 2015 The rising incidence of cardiovascular diseases in India: Assessing its economic impact *J. Prev. Cardiol.* 4(4):735-40.

[4] Uyar K and İlhan A 2017 Diagnosis of heart disease using genetic algorithm based trained recurrent fuzzy neural networks *Procedia Computer Science* 120:588-593.

[5] Patil S B and Kumaraswamy Y S 2009 Extraction of significant patterns from heart disease warehouses for heart attack prediction *IJCSNS* 9(2):228-235.

[6] Fida B, Nazir M, Naveed N and Akram S 2011 December Heart disease classification ensemble optimization using genetic algorithm *In 2011 IEEE 14th International Multitopic Conference* 19-24.

[7] Vasighi M, Zahraei A, Bagheri S and Vafaeimanesh J 2013 Diagnosis of coronary heart disease based on 1H NMR spectra of human blood plasma using genetic algorithm-based feature selection *Journal of Chemometrics* 27(10):318-322.

DOI: 10.54254/2755-2721/2/20220558

[8] Detrano R 1989 Cleveland heart disease database *VA Medical Center, Long Beach and Cleveland Clinic Foundation*.
[9] Sudhakar K and Manimekalai D M 2014 Study of heart disease prediction using data mining *International Journal of Advanced Research in Computer Science and Software Engineering* 4(1):1157-1160.
[10] Nahar J, Imam T, Tickle K S and Chen Y P P 2013 Computational intelligence for heart disease diagnosis: A medical knowledge driven approach *Expert Systems with Applications* 40(1):96-104.
[11] Khazaee A 2013 Heart beat classification using particle swarm optimization *International Journal of Intelligent Systems and Applications* 5(6):25.
[12] Chen C P and Liu Z 2017 Broad learning system: An effective and efficient incremental learning system without the need for deep architecture *IEEE Transactions on Neural Networks and Learning Systems* 29(1):10-24.
[13] Chen Y, Yang B, Dong J and Abraham A 2005 Time-series forecasting using flexible neural tree model *Information Sciences* 174(3-4):219-235.
[14] Singh Y K, Sinha N and Singh S K 2016 November Heart disease prediction system using random forest *In International Conference on Advances in Computing and Data Sciences* 613-623
[15] Kirmani M M 2017 Heart disease prediction using multilayer perceptron algorithm *International Journal of Advanced Research in Computer Science* 8(5):1169 – 1172
[16] Muibideen M and Prasad R 2020 A Fast Algorithm for Heart Disease Prediction using Bayesian Network Model *arXiv preprint* arXiv:2012.09429.