Journal of Cardiovascular Diseases & Diagnosis

Mohamed et al., J Cardiovasc Dis Diagn 2019, 7:3

Research Article · Open Access

Is *H-pylori* Infection a Factor of Acute Coronary Syndrome? – The Neural Networks based Optimization for Mining This Fact

Mohamed S[1*], Ghaleb R[1], Mansour H[1] and Hemeida A[2]

[1]*Department of Cardiology, Aswan University, Aswan, Egypt*

[2]*Department of Electrical Engineering, Aswan University, Aswan, Egypt*

[*]**Corresponding author:** Mohamed S, Department of Cardiology, Aswan University, Aswan, Egypt, Tel: + 201157166779; E-mail: Soud_elsoughier_66@yahoo.com

Received: May 06, 2019; **Accepted:** May 24, 2019; **Published:** May 31, 2019

Abstract

Background: Data mining is making impressive breakthroughs in the field of medicine. Data mining can accomplish tasks such as classifying patients, finding associations between different features and exploring hidden patterns and trends in patient data that simplify and improve medical predictions. Optimization algorithms are used along with neural networks to mine datasets and classify them.

Methods: In this study, artificial neural networks were trained and optimized using various evolutionary optimization algorithms and then applied to classify 119 Egyptian patients with symptoms of coronary artery disease CAD.

Results: The mean age of the study population was 57.9 ± 11.162. The best optimization result was obtained by the PSOGSA algorithm, which correctly classified 97.2% of patients. This high classification rate also confirms that *H. pylori* IgG can be considered a factor of acute coronary syndrome (ACS).

Conclusion: Artificial neural networks can work as classifiers for patients with CAD. The high classification rates confirm that *H. pylori* infection is indeed a strong indicator for ACS. More approaches that use data mining in medicine should be investigated.

Keywords: Data mining; *H. pylori*; ACS; Optimization; Evolutionary computation; Multilayer perceptron

Introduction

Research on data mining in medicine has grown substantially in recent years. Several studies have investigated the use of data mining to make new discoveries in medicine [1-7]. Data mining has numerous approaches that can be implemented to mine medical data, including but not limited to Naive Bayes classifiers [8], decision trees [9]and neural networks [10-12]. Coronary artery disease (CAD) is considered a major cause of mortality. One severe stage of CAD is called acute coronary syndrome (ACS). ACS factors have been studied in efforts to prevent patients from reaching that crucial stage [13-16]. Artificial neural networks (ANNs) are computer networks inspired by the way the human neural system works. ANNs were first introduced by Pitts [17]. Since then, many different types and neural network structures have been proposed, such as feed-forward neural networks [18] and recurrent neural networks [19]. Simple feed forward neural network architecture is shown in Figure 1.

Neural networks are often used to classify objects in data sets, where the input layer represents the dataset inputs. The inputs are assigned synaptic weights; then, the output layer represents the classification of these data. The objective of optimization algorithms in neural networks is to minimize the mean square error to obtain the best possible classification rate. In this work, we implemented the following six optimization algorithms to optimize the neural network: grey wolf optimizer(GWO)[20], whale optimizer (WO) [21], particles warm optimizer (PSO) [22], gravitational search algorithm (GSA)[23], particle swarm optimizer-gravitational search algorithm (PSOGSA) [24] and particle swarm optimizer-grey wolf optimizer (PSOGWO) [25].

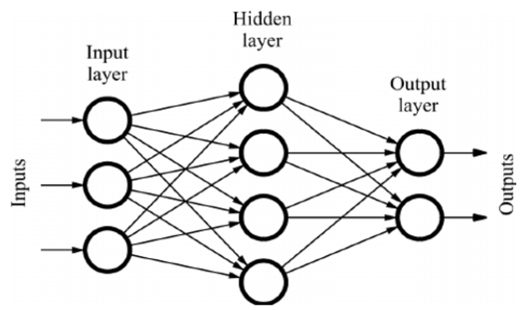

Figure 1: An ANN with structure 3-4-2.

Methods

Data collection and laboratory methods

The dataset used in this research consists of 119 Egyptian patients from Aswan City in southern Egypt who were diagnosed with typical chest pain (TCP), percutaneous coronary intervention (PCI), or both (data sampling was conducted from December 2016 to February 2017). Diabetes mellitus was detected in 46.2% of the patients (sugar level >123 mg/dl), and hypertension was detected in 42% of patients (blood pressure>140/90 mmHg). Antibody tests for infections were administered, and both IgG and IgM tests were evaluated. Other factors reported in the data set include hospital stay, family history, smoking and the existence of troponin. Before participation, all patients provided informed consent. The procedures were performed at Aswan University Hospital and at 6th of October University Hospital in Egypt.

Neural network construction and optimization

To train the neural network, approximately half of the dataset population (60 patients) was used as a training set, and the remainder (59 patients) was used as a test set. After training, the neural network was tested, and the correct classification percentage and mean square error values were calculated. We used four features to train the neural network: age, IgG, IgM and troponin. The neural network was structured as 4-9-3. The weights and biases of the neural network nodes were optimized using six optimization algorithms; then, the results were compared to find the best classifier among the tested algorithms (Figure 2). Each algorithm was executed three times, and their average classification rates and errors were calculated.

Computer specifications	
Model	HP Pavilion 15 Notebook PC
Operating System	Microsoft Windows 8, 64-bit
Processor	Intel® Core™ i5-3230M CPU@2.60 G Hz
RAM	4 GB
Graphics Card	AMD Radeon ™ HD 8670 M
MATLAB Version	R2017a

Table 1: Computer specifications for executing the mining process.

Results and Discussion

The experiments are executed on a computer with the specifications as in Table 1. Each algorithm was trained and tested three times. The best and average classification rates and MSE scores are listed in Table 2.

Algorithm	Best Classification Rate (%)	Best MSE	Average Classification Rate (%)	Average MSE
GSA	95	0.0883	95	0.090924333
GWO	90.833	0.09296	85.27776667	0.096182
PSO	95	0.0834	94.16666667	0.093162333
PSO-GSA	98.333	0.043	97.22223333	0.051051
PSO-GWO	96.33	0.0785	94.1667	0.08885
WO	88.333	0.1126	85.83333333	0.1264

Table 2: Best and average results for the optimization algorithms.

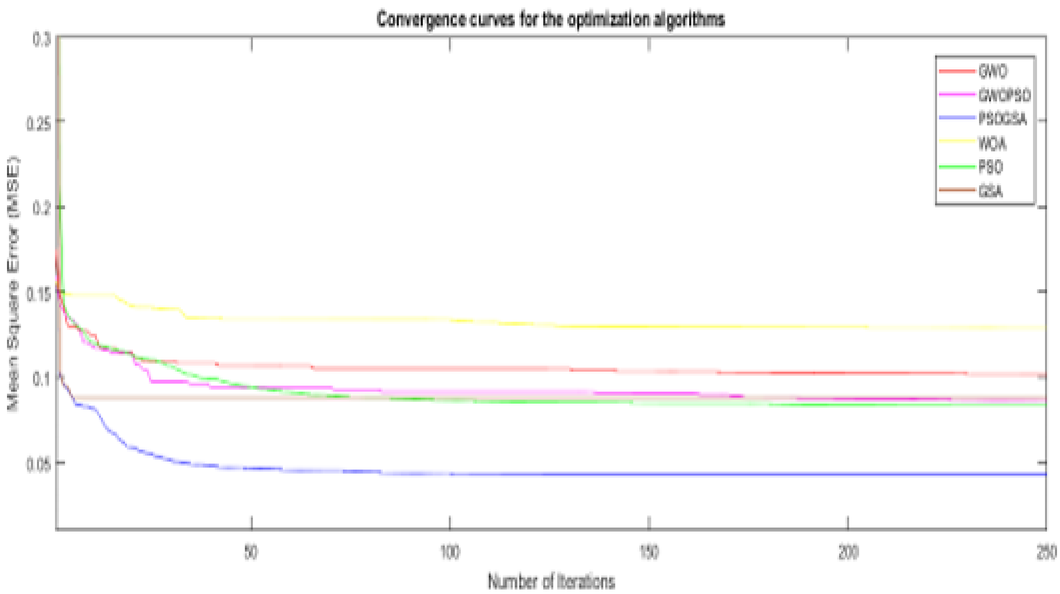

Figure 2: Converge curves for the optimization algorithms.

The hybrid PSOGSA algorithm achieved the highest classification rate and the lowest MSE score, although other algorithms provided competitive results. From these results, we can make two conclusions. First, the high classification rates demonstrate that optimized neural networks are able to address medical datasets satisfactorily. Second, the selected features helped in obtaining the high results, which in turn indicates that these parameters directly affect the results. The important features were IgG and IgM, which are found in patients infected with *H. pylori*. This result substantiates that *H. pylori* infection is indeed a risk factor for ACS, although it is not the only indicator. Research is still ongoing concerning ACS risk factors [26-30].

Limitations of Study

Our study has a number of limitations. First, selecting the optimal numbers of hidden neurons and the most appropriate network structure are challenging tasks in neural network classification because no universal method exists for selecting these parameters. Typically, these parameter values are either manually selected or found through the use of optimization algorithms or other mathematical techniques [31-35]. One issue of optimization algorithms is that they are stochastic, meaning that each execution can provide slightly different results, which is why multiple code executions should be implemented. Moreover, different optimization algorithms can result in better or worse results because according to the no-free-lunch (NFL) theorem, no single optimization algorithm is well-suited for solving all optimization problems [36]. The proposed optimization algorithms were all easily able to train the network to classify the dataset due to the nature of the dataset. A more challenging task would be to test these algorithms on larger and more diverse datasets. Additionally, the risk factors did not reveal the reasons why some of the patients in the dataset developed ACS; this issue requires further investigation.

Conclusion

Artificial neural networks are suitable for classifying patients with CAD when the appropriate features are selected. The high classification rates in this study confirm that *H. pylori* infection is indeed a strong indicator for ACS. Because data mining is capable of many more tasks than simply classification, additional approaches to incorporate data mining in the field of medicine are highly encouraged and should be very promising.

Conflicts of Interest

There are no conflicts of interest for the present study.

Availability of Data and Materials

The datasets available from the corresponding author on request.

Ethics Approval and Consent to Participate

Informed consent was obtained from each participant prior to their study participation.

References

1. Davazdahemami B, Delen D (2019) The confounding role of common diabetes medications in developing acute renal failure: A data mining approach with emphasis on drug-drug interactions. Expert Syst Appl 123: 168-177.
2. Supriyanto A, Suryono S, Susesno JE (2018) Implementation data mining using decision tree method-algorithm c4. 5 for postpartum depression diagnosis. In: E3S Web of Conferences. EDP Sciences 73: 12012.
3. Mi N, Cheng T, Li H, Yang P, Mu X, et al. (2019) Metabolite profiling of traditional Chinese medicine formula Dan Zhi Tablet: An integrated strategy based on UPLC-QTOF/MS combined with multivariate statistical analysis. J Pharm Biomed Anal 164: 70-85.
4. Gatarić B, Parojčić J (2019) Application of data mining approach to identify drug subclasses based on solubility and permeability. Biopharm Drug Dispos 40: 51-61.
5. Harahap M, Husein AM, Aisyah S, Lubis FR, Wijaya BA (2018) Mining association rule based on the diseases population for recommendation of medicine need. J Phys Conf Ser 1007: 012017.
6. Mansour H, Ghaleb R, Fakhry M, Elgohary E, Ismael A, et al. (2017) Association between Acute Coronary Syndrome and Helicobacter pylori Chronic Infection? What data mining tells us? J Cardiovascular Thoracic Surgery 2: 1-3
7. Rogers MA, Aikawa E (2018) Cardiovascular calcification: Artificial intelligence and big data accelerate mechanistic discovery. Nat Rev Cardiol 10: 1.
8. Choudhury A, Gupta D (2019) A survey on medical diagnosis of diabetes using machine learning techniques. In Recent Developments in Machine Learning and Data Analytics, Springer pp: 67-78.
9. Zhu F, Tang M, Xie L, Zhu H (2018) A classification algorithm of CART decision tree based on map reduce attribute weights. IJPE 14: 17.
10. Hannun AY, Rajpurkar P, Haghpanahi M, Tison GH, Bourn C, et al. (2019) Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network. Nat Med 25: 65-69.
11. Dgani Y, Greenspan H, Goldberger J (2018) Training a neural network based on unreliable human annotation of medical images. In: 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018) IEEE.
12. Zreik M, Lessmann N, Van Hamersvelt RW, Wolterink JM, Voskuil M, et al. (2018) Deep learning analysis of the myocardium in coronary CT angiography for identification of patients with functionally significant coronary artery stenosis. Med Image Anal 44: 72-85.
13. Zhou M, Liu J, Hao Y, Liu J, Huo Y, et al. (2018) Prevalence and in-hospital outcomes of diabetes among patients with acute coronary syndrome in China: Findings from the Improving care for cardiovascular disease in china-acute coronary syndrome project. Cardiovasc Diabetol 17: 147.
14. Yakovlev A, Metsker O, Kovalchuk S, Bologova E (2018) Prediction of in-hospital mortality and length of stay in acute coronary syndrome patients using machine-learning methods. J Am Coll Cardiol 71: A242.
15. Liu XJ, Wan ZF, Zhao N, Zhang YP, Mi L, et al. (2015) Adjustment of the GRACE score by HemoglobinA1c enables a more accurate prediction of long-term major adverse cardiac events in acute coronary syndrome without diabetes undergoing percutaneous coronary intervention. Cardiovasc Diabetol 14: 110.
16. Shemesh J, Tenenbaum A, Fisman EZ, Koren-Morag N, Grossman E (2013) Coronary calcium in patients with and without diabetes: First manifestation of acute or chronic coronary events is characterized by different calcification patterns. Cardiovasc Diabetol 12: 161.
17. McCulloch WS, Pitts W (1943) A logical calculus of the ideas immanent in nervous activity. Bull Math Biophys 5: 115-133.
18. Arulmurugan R, Anandakumar H (2018) Early detection of lung cancer using wavelet feature descriptor and feed forward back propagation neural networks classifier. In Computational Vision and Bio Inspired Computing, Springer 103-110.
19. Che Z, Purushotham S, Cho K, Sontag D, Liu Y (2018) Recurrent neural networks for multivariate time series with missing values. Sci Rep 8: 6085.
20. Mirjalili S, Mirjalili SM, Lewis A (2014) Grey wolf optimizer. Adv Eng Softw 69: 46-61.
21. Mirjalili S, Lewis A (2016) The whale optimization algorithm. Adv Eng Softw 95: 51-67.
22. Fu H, Li Z, Liu Z, Wang Z (2018) Research on big data digging of hot topics about recycled water use on micro-blog based on particle swarm optimization. Sustainability 10: 2488.

23. González B, Melin P, Valdez F, Prado-Arechiga G (2018) Ensemble neural network optimization using a gravitational search algorithm with interval type-1 and type-2 fuzzy parameter adaptation in pattern recognition applications. In Fuzzy Logic Augmentation of Neural and Optimization Algorithms: Theoretical Aspects and Real Applications, Springer 17-27.
24. Tolba MA, Diab AA, Tulsky VN, Abdelaziz AY (2018) VLCI approach for optimal capacitors allocation in distribution networks based on hybrid PSOGSA optimization algorithm. Neural Comput Appl :1-8.
25. Abdelshafy AM, Hassan H, Jurasz J (2018) Optimal design of a grid-connected desalination plant powered by renewable energy resources using a hybrid PSO – GWO approach. Energy Convers Manag 173: 331-347.
26. Okada K, Hibi K, Gohbara M, Kataoka S, Takano K, et al. (2015) Association between blood glucose variability and coronary plaque instability in patients with acute coronary syndromes. Cardiovasc Diabetol 14: 111.
27. Bęćkowski M, Gierlotka M, Gąsior M, Poloński L, Zdrojewski T, et al. (2018) Risk factors predisposing to acute coronary syndromes in young women ≤ 45 years of age. Int J Cardiol Heart Vasc 264: 165-169.
28. Ouchi S, Shimada K, Miyazaki T, Takahashi S, Sugita Y, et al. (2017) Low 1, 5-anhydroglucitol levels are associated with long-term cardiac mortality in acute coronary syndrome patients with hemoglobin A1c levels less than 7.0%. Cardiovasc Diabetol 16: 151.
29. Kennedy MW, Fabris E, Suryapranata H, Kedhi E (2017) Is ischemia the only factor predicting cardiovascular outcomes in all diabetes mellitus patients?. Cardiovasc Diabetol 16: 51.
30. Zhang S, Dai J, Jia H, Hu S, Du H, et al. (2018) Non-culprit plaque characteristics in acute coronary syndrome patients with raised hemoglobin A1c: An intravascular optical coherence tomography study. Cardiovasc Diabetol 17: 90.
31. Faris H, Mirjalili S, Aljarah I (2019) Automatic selection of hidden neurons and weights in neural networks using grey wolf optimizer based on a hybrid encoding scheme. Int J Mach Learn Cyb :1-20.
32. Sheela KG, Deepa SN (2013) Review on methods to fix number of hidden neurons in neural networks. Math Probl Eng 2013: 1-11.
33. Thomas L, Kumar M, Annappa B (2016) Discovery of optimal neurons and hidden layers in feed-forward Neural Network. In 2016 IEEE International Conference on Emerging Technologies and Innovative Business Practices for the Transformation of Societies (EmergiTech). IEEE pp: 286-291.
34. Chhachhiya D, Sharma A, Gupta M (2017) Designing optimal architecture of neural network with particle swarm optimization techniques specifically for educational dataset. In 2017 7th International Conference on Cloud Computing, Data Science & Engineering-Confluence. IEEE pp: 52-57.
35. Madhiarasan M, Deepa SN (2016) New criteria for estimating the hidden layer neuron numbers for recursive radial basis function networks and its application in wind speed forecasting. Asian Journal of Information Technology 15: 4377-4391.
36. Wolpert DH, Macready WG (1997) No free lunch theorems for optimization. IEEE Trans Evol Comput 1: 67-82.