Provide a diagnostic model using a combination of two neural network algorithms and a genetic algorithm

Farshad Minaei¹, Hassan Dosti², Ebrahim Salimi Turk³, Amin Golabpour³* (E)

¹Karoon Institute of Higher Education, Ahvaz, Iran
²Department of Mathematics and Statistics Macquarie University, Sydney, Australia
³Assistant Professor, Shahroud University of Medical Sciences, Shahroud, Iran

Introduction: Improvement of technology can increase the use of machine learning algorithms in predicting diseases. Early diagnosis of the disease can reduce mortality and morbidity at the community level.

Material and Methods: In this paper, a clinical decision support system for the diagnosis of gestational diabetes is presented by combining artificial neural network and meta-heuristic algorithm. In this study, four meta-innovative algorithms of genetics, ant colony, particle Swarm optimization and cuckoo search were selected to be combined with artificial neural network. Then these four algorithms were compared with each other. The data set contains 768 records and 8 dependent variables. This data set has 200 missing records, so the number of study records was reduced to 568 records.

Results: The data were divided into two sets of training and testing by 10-Fold method. Then, all four algorithms of neural-genetic network, ant-neural colony network, neural network-particle Swarm optimization and neural network-cuckoo search on the data The trainings were performed and then evaluated by the test set. And the accuracy of 95.02 was obtained. Also, the final output of the algorithm was examined with two similar tasks and it was shown that the proposed model worked better.

Conclusion: In this study showed that the combination of two neural network and genetic algorithms can provide a suitable predictive model for disease diagnosis.

INTRODUCTION

Diabetes is an important disease in which the body is unable to control blood sugar due to insufficient levels of the hormone insulin [1]. Diabetes during pregnancy is known as the most common medical problem and affects 3.3% of pregnant women. No matter what the type of diabetes, gestational diabetes is one of the most important disorders of pregnancy, which is defined as glucose intolerance during pregnancy. Mothers who are most at risk are often those who have had an abortion for no apparent reason, or are overweight, or have a family history of diabetes [2].

 Mothers who suffer from pregnancy complications are more likely to develop overt diabetes after giving birth. Diabetes may also recur in later pregnancies. At the same time, these mothers will not give birth to normal children. According to the results of extensive studies around the world, it is possible to control this disease with low cost and diet and prevent its complications. Prior to the discovery of insulin, pregnancy was low due to mortality in diabetic women and impaired fertility in diabetics, but with the discovery of insulin and the survival of diabetic women, an increase in fertility has been observed. Research has shown that there is a direct relationship between fetal survival and maternal blood sugar, so that with careful monitoring of maternal glucose, pregnancy complications and abnormal birth can be prevented [3]. Gestational diabetes is the most common metabolic disorder during pregnancy, with a prevalence of between 17.7 and 0.15 percent. It is 1 to 6 percent, of which 15 to 25 percent is insulin-dependent diabetes and 75 to 85 percent is non-insulin dependent. In Iran, it seems that the total
prevalence of diabetes is 3 to 4% and the independent condition is more than 7 to 8% and the prevalence of glucose tolerance test is the same, which can turn into overt diabetes in physiological conditions such as pregnancy or disease [4].

Early prediction of gestational diabetes causes timely treatment [5]. This study tries to present a prediction model using artificial neural network algorithm. To optimize the neural network, meta-innovative algorithms are used. Four meta-heuristic algorithms, genetic algorithm, particle Swarm optimization algorithm, ant colony algorithm and cuckoo search algorithm are presented to optimize the artificial neural network. All four meta-heuristic algorithms are compared in terms of accuracy and the best meta-heuristic algorithm for Neural network optimization was determined.

**MATERIAL AND METHODS**

Table 1: Gestational diabetes data structure [7]

| Row | Variables       | Type       | Scale  | Lower | Upper | Mean  | Variance | Description                      |
|-----|----------------|------------|--------|-------|-------|-------|----------|----------------------------------|
| 1   | Pregnancies    | Dependent  | Ordinal | 0     | 7     | 3.84  | 3.37     | Number of pregnancies            |
| 2   | PG Concentration| Dependent  | Ordinal | 0     | 99    | 120.88| 31.99     | Two hours glucose tolerance      |
| 3   | Diastolic BP   | Dependent  | Ordinal | 0     | 22    | 69.1  | 19.37     | Diastolic blood pressure         |
| 4   | Tri Fold Thick | Dependent  | Ordinal | 0     | 99    | 25.52 | 5.96      | Blood sugar is calculated regardless of the last meal |
| 5   | Serum Ins      | Dependent  | Ordinal | 0     | 846   | 79.83 | 15.32     | Serum insulin                    |
| 6   | BMI            | Dependent  | Ordinal | 8.2   | 2.42  | 32    | 7.88      | Body mass                        |
| 7   | DP Function    | Dependent  | Ordinal | 0.078 | 2.42  | 0.47  | 0.33      | It is possible that this patient was inherited from the previous generation |
| 8   | Age            | Dependent  | Ordinal | 21    | 81    | 33.18 | 8.64      | Patient age                      |
| 9   | Diabetes       | Nominal    | Ordinal | 0     | 1     | 0.65  | 0.48      | It includes two sick and healthy states, of which 268 are sick and 500 are healthy |

**Algorithm preprocessing**

Independent variables have different intervals. In order for the intervals of all independent variables to be the same, the data transfer formula to the interval [0,1] is used. In (Eq. 1) the structure of data conversion from each interval to interval [0,1] is expressed [8].

\[
Value_{new} = \frac{Max-Value_{old}}{Max-Min} 
\]

(1)

Runs on all independent variables (Eq. 1) and all variables are converted to an interval of [0,1].

Proposed model

The proposed model consists of two parts. In the first part, the proposed model is expressed using artificial neural network and in the second stage, artificial neural network is optimized using four meta-innovative algorithms: particle Swarm optimization, ant colony and cuckoo search. Which optimization algorithm works best?

**Problem solving with the help of artificial neural network algorithm**

The prediction algorithm used in this research is an artificial neural network. The data is divided into two sets of training and testing [9]. The artificial neural network provides a predictive model by the training data and this model is evaluated by the training and test data. Fig 1 evaluates the structure of the artificial neural network.

![Fig 1: Model design structure with the help of artificial neural network](image)

**Neural network optimization using meta-heuristic algorithms**

In this part, neural network learning is done by meta-innovative algorithms. One of the problems of neural network structure is optimal training, because the neural network uses derivative-based methods such as gradient descent method. In this study, the proposed model is implemented which is a combination of neural network algorithm and meta-heuristic algorithm.
as gradients in the standard mode. They have difficulty determining the optimal neural network weights, so the neural network adapts well to the training data but does not work well for the test data. To solve this problem, neural network training is placed on a meta-heuristic algorithm and the amount of neural network weights is determined by the meta-heuristic algorithm, given that the neural network weights and bias values are continuous.

Therefore, in this research, four meta-innovative algorithms in the continuous field have been used, which include genetic algorithms, particle Swarm optimization, ant colony, and cuckoo search. The considered neural network has three layers, which first calculate the number of neural network variables including neural network weight and bias. The neural network defined in this issue is three layers and the number of neural network inputs is equal to the number of independent variables. In this study, the number of variables is considered to be eight. The number of neurons in the middle layer is considered to be 10. Neural network dependent variables have an output. The number of middle layer neurons is calculated from (Eq. 2), where n1 is the number of input layer neurons, n3 is the number of output layer neurons, and ni is the number of layer i neurons [11].

\[ n_2 = n_1 + n_3 + 1 \]  

(2)

According to (Eq. 2), the value of n1 is equal to 8 (number of independent variables) and the number of n3 is equal to 1 (number of dependent variables), so n2 is the number of middle layer neurons, which is calculated according to the value of (Eq. 2) 10. In the next step, the number of unknowns in the artificial neural network, which includes weights and biases, must be determined. If the number of neural network layers is k and ni represents the number of neurons in the neural network of layer i, the number of unknowns is calculated from (Eq. 3) (the number of unknowns is denoted by m) [11].

\[ m = \sum_{i=1}^{k} (n_i + 1) n_{i+1} \]  

(3)

According to (Eq. 3) and the number of neurons in each layer, the number of network weights and biases in the neural network is 101. Therefore, a 101-dimensional problem is defined, which indicates the complexity of the problem and the length of the meta-heuristic algorithm. The lower and upper limits in this case are -1 and +1 because the weights of the neural network are between -1 and +1. The response structure of the meta-heuristic algorithm is shown in Fig. 2.

Fig 2: Structure of a solution of meta-heuristic algorithms

Each structure in the meta-heuristic algorithms of an Individual is said to be used by each Individual to create a neural network. The structure of neural network production with the help of Individual is shown schematically in Fig 3.

![Fig 3: An example of converting an Individual to a neural network and vice versa](image)

For all four genetic algorithms, particle Swarm optimization, ant colony, and cuckoo search algorithm, the structure of Fig 1 is established. In any meta-heuristic algorithm, there are three steps:

- **Primary population**
- **Competency function**
- **Algorithm end condition**

All three of the above steps are the same problem in all four algorithms.

### Primary population

After this step, the population size must be determined. That is, how many populations should be available. Population size is an important factor in the efficiency of the algorithm. If the population size is too small, a small portion of the answer space will be searched and the answer will quickly and most likely converge to a local optimization. And if the population size is very large, a lot of calculations that are disproportionate to the result obtained will be performed, resulting in a very long execution time [12]. The initial population number for all four algorithms is 100. The value is calculated by trial and error. Normal distribution is used to produce the initial population.

### Calculate the competency function

Each Individual is a set of weights of an artificial neural network that can be used to draw the structure of a neural network. The neural network derived from the Individual must be evaluated. Gestational diabetes data is used to assess the artificial neural network. Gestational diabetes data is divided into two groups of training and test data. To evaluate the neural network, only training data is used and test data is not used. Test data would be used for the final evaluation of the neural network. The neural network evaluation parameter is
considered correct. Fig 4 shows the calculation structure of the competency function.

**Algorithm end condition**

The condition for the end of all four meta-heuristic algorithms is to produce 50 generations without changing the competency function, this value has been calculated by trial and error. The following is a general flowchart of neural network optimization using meta-heuristic algorithm.

As can be seen in Fig 5, the general structure of artificial neural network optimization with meta-heuristic algorithm is such that the meta-heuristic algorithm forms the best neural network after generating the initial answer, then the neural network with the case accuracy parameter. The evaluation value is sent back to the meta-heuristic algorithm to produce a better Individual meta-heuristic algorithm. This process is repeated to optimize the neural network algorithm using a meta-heuristic algorithm.

In the following, the four meta-heuristic algorithms used in this research are compared to determine which meta-heuristic algorithm produces the most optimal neural network. Fig 6 shows a comparison of four meta-heuristic algorithms. First, the whole data is divided into two sets of training and testing based on the K-Fold method. Then, on the training data, the neural network algorithm is implemented with four learning models based on meta-heuristic algorithms. After generating these four learning models, these models are evaluated with test data based on the accuracy parameter to determine which learning algorithm worked best.

**RESULTS**

Now the parameters of all four genetic algorithms, particle Swarm optimization, ant colony and cuckoo search are expressed. All values are calculated by trial and error. Table 2 shows the parameters of these four algorithms. The initial population size is assumed to be 100 in all four algorithms.
Table 2: Parameters of four optimization algorithms

| Parameter name | Value | Parameter name | Value | Parameter name | Value | Parameter name | Value | Parameter name | Value |
|----------------|-------|----------------|-------|----------------|-------|----------------|-------|----------------|-------|
| Initial Population | 100 | Initial Population | 100 | Initial Population | 100 | Initial Population | 100 | Initial Population | 100 |
| Minimal Spawning | 2 | Sample Size | 200 | C_1 | 2 | Crossover | 0.8 | Mutations | 0.3 |
| Maximum Spawning | 5 | |

Table 3: Output of proposed algorithms

| Row | Algorithm | Training data | Test data | |
|-----|-----------|---------------|-----------|-----|
|     |           | Sensitivity   | Specificity | Accuracy | Positive prediction | Negative prediction | F-measure | Sensitivity | Specificity | Accuracy | Positive prediction | Negative prediction | F-measure |
| 1   | ANN       | 83.21         | 80.60      | 80.30    | 93.27            | 81.15            | 82.84    | 82.18       | 77.18       | 73.75    | 81.54             | 58.19                | 81.55    |
| 2   | Cuckoo search algorithm and ANN | 97.14 | 85.28 | 89.87 | 98.43 | 93.49 | 92.58 | 99.49 | 72.76 | 83.25 | 84.49 | 79.04 | 91.20 |
| 3   | Particle swarm optimization and ANN | 94.55 | 81.25 | 96.43 | 92.28 | 81.00 | 82.36 | 93.48 | 79.84 | 85.26 | 89.38 | 76.12 | 91.29 |
| 4   | Ant colony algorithm and ANN | 98.89 | 98.25 | 98.20 | 98.89 | 93.75 | 94.15 | 98.34 | 86.96 | 92.35 | 89.19 | 82.31 | 93.42 |
| 5   | Genetic algorithm ANN | 98.34 | 96.96 | 97.35 | 99.19 | 92.31 | 93.42 | 98.89 | 87.75 | 95.02 | 89.70 | 83.59 | 94.04 |

DISCUSSION

The combination of genetic algorithm and artificial neural network algorithm has the best output. As shown in Table 4, the output of the proposed algorithm performed better than two similar algorithms.

Table 4: Comparison of the proposed algorithm with similar tasks

| Row | Comparison | Accuracy |
|-----|------------|----------|
| 1   | Genetic algorithm and ANN | 95.02% |
| 2   | [13]       | 93.05%   |
| 3   | [14]       | 94.65%   |

Extraction of predictive model in medical data with two indicators of accuracy and high interpretability is very important in increasing the accuracy and speed of diagnosis by specialists [15]. In this paper, a combined model of genetic algorithm and neural network is presented to determine the factors affecting the diabetes data set and produce a predictable model with high interpretability. In this study, it was shown that the neural network algorithm is a suitable model for disease diagnosis.

The proposed model is used for the following reasons: (1) The proposed model, by using data mining algorithms, significantly reduces the cost of diagnosing diseases. (2) In most cases, the doctor cannot say with complete certainty whether people have gestational diabetes or not, but the proposed model can be a great guide for the doctor so that he can comment more clearly based on an intelligent response.

Given the importance of diagnosing gestational diabetes, today advances in technology and innovations have led to significant results in the health sector, and researchers using the various algorithms mentioned in this article, with the help of physicians in detecting and controlling diseases were useful.

It is important to diagnose diabetes as soon as possible. In intelligent systems, one of the important applications of data mining techniques is related to the diagnosis of diseases. Automatic diagnostic systems, with the help of medical data, can shorten the diagnosis time, reduce the possible error of experts and make the details more accurate. In this paper, a combined neuro-genetic model is presented to provide a diagnostic model from medical data. The goal was to achieve a predictable model with high interpretability.

The effectiveness of the proposed method on the diabetes dataset was evaluated. This dataset contains 176 records containing information on people with diabetes and 401 records containing information on non-diabetics. Despite the imbalance between positive and negative samples of this dataset, the best output on the test data achieved 95.02% accuracy, 98.80% sensitivity, 87.75% specificity, PPV99.7%, NPV 83.05% and f-measure 94.04%. The proposed model has reached this output using 8 features. Also, according to the proposed model and
the reliability of differentiation between healthy and sick people according to NPV and PPV criteria, the proposed model can be used in the construction of medical decision-making systems.

**CONCLUSION**

Gestational diabetes can become a chronic disease if not diagnosed early. Diagnosing gestational diabetes in the early stages of the disease can dramatically prevent mortality. Definitive treatment of cancer is possible with early detection. So far, various methods based on meta-innovative algorithms, artificial neural network for diagnosing gestational diabetes have been proposed. The best model for diagnosing this disease is the one with the highest accuracy. In this study, a model for diagnosing gestational diabetes was presented based on artificial neural network algorithm and genetics. Evaluations were performed on the UCI Pima dataset. According to the results, the proposed model was more accurate than similar algorithms and other models.

**AUTHOR’S CONTRIBUTION**

All authors contributed to the literature review, design, data collection and analysis, drafting the manuscript, read and approved the final manuscript.

**CONFLICTS OF INTEREST**

The authors declare no conflicts of interest regarding the publication of this study.

**FINANCIAL DISCLOSURE**

No financial interests related to the material of this manuscript have been declared.

**REFERENCES**

1. Kakooei S, Afzali S, Parirokh M, Kakooei S, Mostafavii M, Nekouei A. The knowledge and attitude of diabetic patients regarding oral and dental disorders in Kerman diabetes clinics. J Dent (Shiraz). 2020; 21(3): 195-201. PMID: 33062813 DOI: 10.30476/DENTJODS.2020.77878.0 [PubMed]

2. Gestational diabetes [Internet]. 2021 [cited: 15 Jun 2021; updated: 1 Jun 2021]. Available from: https://en.wikipedia.org/w/index.php?title=Gestational_diabetes&oldid=899796366

3. Artzi NS, Shilo S, Hadar E, Rossman H, Barbash-Hazan S, Ben-Haroush A, et al. Prediction of gestational diabetes based on nationwide electronic health records. Nat Med. 2020; 26(1):71-76. PMID: 31932807 DOI: 10.1038/s41591-019-0724-8 [PubMed]

4. Vounzoulakis E, Khunti K, Ahner SC, Tan BK, Davies MJ, Gillies CL. Progression to type 2 diabetes in women with a known history of gestational diabetes: Systematic review and meta-analysis. BMJ. 2020; 369: m1361. PMID: 32404325 DOI: 10.1136/bmj.m1361 [PubMed]

5. McIntyre HD, Catalano P, Zhang C, Desaye G, Mathiesen ER, Damm P. Gestational diabetes mellitus. Nat Rev Dis Primers. 2019; 5(1): 47. PMID: 31296866 DOI: 10.1038/s41572-019-0098-8 [PubMed]

6. Kile C, Wong M. Risk of gestational diabetes among Pima Indians: Categorical analysis data in SAS. California State University; 2020.

7. Gnanadass I. Prediction of gestational diabetes by machine learning algorithms. IEEE Potentials. 2020; 39(6): 32-7.

8. Yang H. Data preprocessing. Citeseer; 2018.

9. Sammut C, Webb GI. Encyclopedia of machine learning and data mining. Springer; 2017.

10. Witten IH, Frank E, Hall MA, Pal C. Data mining: Practical machine learning tools and techniques. Elsevier; 2017.

11. Wang SC. Interdisciplinary computing in Java programming. Springer; 2012.

12. Hu J, Morais H, Sousa T, Lind M. Electric vehicle fleet management in smart grids: A review of services, optimization and control aspects. Renewable and Sustainable Energy Reviews. 2016; 56: 1207-26.

13. Srivastava S, Sharma L, Sharma V, Kumar A, Darbari H. Prediction of diabetes using artificial neural network approach. In: Ray K, Sharan SN, Rawat S, Jain SK, Srivastava S, Bandyopadhyay A (eds). Engineering vibration, communication and information processing. Springer; 2019.

14. El Jerjawi NS, Abu-Naser SS. Diabetes prediction using artificial neural network. International Journal of Advanced Science and Technology. 2018; 121: 55-64.

15. Haux R. Medical informatics: Past, present, future. Int J Med Inform. 2010; 79(9): 599-610. PMID: 20615752 DOI: 10.1016/j.ijmedinf.2010.06.003 [PubMed]