

A Novel model for diagnosing high-risk pregnancies mothers using Bayesian belief network algorithm and particle optimization

Azadeh Abkar1 , Amin Golabpour2* 

¹Department of Computer Engineering, Faculty of Computer, Karoon Institute of Higher Education, Ahvaz, Iran

²School of Allied Medical Sciences, Shahroud University of Medical Sciences, Shahroud, Iran

³*Correspondence: a.golabpour@shmu.ac.ir

Article Info

Article type:
Research

Article History:

Received: 2022-01-10

Accepted: 2022-02-09

Published: 2022-03-01

* Corresponding author:

Amin Golabpour

School of Allied Medical Sciences,
Shahroud University of Medical
Sciences, Shahroud, Iran

Email: a.golabpour@shmu.ac.ir

Keywords:

High-Risk Pregnancy
Bayesian Belief Network
Particle Optimization
Data Mining

ABSTRACT

Introduction: Diagnosis of high-risk maternal pregnancy is one of the most important issues during pregnancy and can be of great help to pregnant mothers. Also, early diagnosis can reduce mortality and morbidity in mothers.

Material and Methods: In this study, the data of 1014 pregnant mothers were used, which includes 272 people with high-risk pregnancies, 742 people with medium-risk and low-risk pregnancies. Also, the data include six independent variables. A combination of Bayesian belief network algorithms and particle optimization was used to predict pregnancy risk.

Results: For validation, the data model was divided into two sets of training and testing based on the method of 30-70. Then the proposed model was designed by training data. Then the model for training and testing data was evaluated in terms of accuracy parameters 99.18 and 98.32% accuracy were obtained, respectively. It has also performed between 0.5 and 8% better than similar work in the past.

Conclusion: In this study, a new model for designing Bayesian belief network was presented and it was found that this model can be useful for predicting maternal pregnancy risk.

Cite this paper as:

Abkar A, Golabpour A. A Novel model for diagnosing high-risk pregnancies mothers using Bayesian belief network algorithm and particle optimization. Front Health Inform. 2022; 11: 108. DOI: [10.30699/fhi.v11i1.351](https://doi.org/10.30699/fhi.v11i1.351)

INTRODUCTION

In 2010, 800 pregnant women lost their lives due to complications and complications of pregnancy (during pregnancy, during childbirth and after childbirth) worldwide, bringing the number to 287,000 per year, while Such deaths can be prevented by timely diagnosis of high-risk pregnancies and simple measures and provision of basic facilities and equipment. Statistics show that worldwide, the death rate of mothers in rural and poor areas is higher than in urban and affluent areas. 99% of maternal mortality worldwide occurs in developing countries. Overall, out of 800 maternal deaths (daily), most of them died due to severe postpartum hemorrhage, hypertension, pregnancy poisoning and unsafe abortion [1].

The term high-risk pregnancy refers to pregnant mothers who have some risky conditions and need special care during pregnancy and postpartum. There are many factors that can put pregnancy at risk. The

word high-risk may sound scary, but the term is commonly used to make sure that medical staff receive special care during pregnancy [2].

Pregnant mothers with high-risk pregnancies need special care to ensure their health and that of the fetus. These people need to see a doctor more often and may need more tests and ultrasounds during pregnancy to monitor the health of the mother and fetus. These cares will help the doctor to diagnose any medical problem quickly and take the necessary measures [3].

Early prediction of high-risk pregnancies is very useful and can sometimes help in medical decisions to continue the pregnancy or terminate the pregnancy. This can reduce the mortality and morbidity of pregnant mothers [4]. Therefore, providing a prediction model using data mining algorithms can be a good help in early diagnosis of high-risk pregnancies of mothers.

Many studies have been conducted in the field of high-risk pregnancy prediction, including in 2020, a study using the decision tree was presented for this prediction. Data for this study were collected using IoT technology and HIS software. In this study, the data of six hospitals in the period 2018 to 2020 were used and also the data include six independent variables and one dependent variable; In this study, it was shown that the decision tree algorithm has been better in predicting the model than similar methods and has reached 97% accuracy. Also, WEKA and Python software were used for implementation and it was shown that Python software has a better output than Weka [5].

In 2020, a study was presented using intelligent algorithms to predict high-risk pregnancies. In this study, the logistic regression decision tree algorithm was used [6]. The structure of this algorithm is such that a logistic regression function is executed in the roots of the decision tree instead of the final answer. The output of this research is 97.96% accurate. The data of this study include nine independent variables and one dependent variable. It was also shown that increasing the age of mothers has a direct effect on high-risk pregnancies [7].

In 2020, a study examined different algorithms for predicting high-risk pregnancies of mothers and showed that two advanced artificial neural network and convolutional neural network algorithms are suitable for predicting high-risk pregnancies and can act as a runtime for You provided a neural network-based hardware whose input is the output of IoT devices. Data is taken from the input in runtime, then the data is analyzed by a trained neural network and the output is specified. In this paper, the final output has reached 90% accuracy [8].

One of the important aspects of the study of this disease, which has not been comprehensively considered so far, is the discovery and extraction of the laws underlying

Highly accurate and interpretable data [9]. Therefore, an algorithm that can be interpreted from the physician's point of view and can use these rules without a computer must be presented. Mothers presented. To build a Bayesian belief network, an expert person or a meta-innovative algorithm must be used. In this study, using Bayern Bayesian particle optimization algorithm, this network is designed that this network can predict high-risk pregnancies of mothers.

MATERIAL AND METHODS

This study uses general UCI data. These data were

collected from several hospitals in the period 2018 to 2020 [10]. These data include 1014 records and six continuous independent variables and one rank dependent variable. There are six independent variables including Age, Diastolic Blood Pressure, Systolic Blood Pressure, Blood glucose levels, Body Temperature, HeartRate and the dependent variable has three modes of low-risk pregnancy, moderate-risk pregnancy and high-risk pregnancy. The details of this data are shown in Table 1.

Table 1: Percentage of data breakdown by type of pregnancy risk [10]

Dependent	Frequency	Percent	Cumulative Percent
Low risk	406	40.0	40.0
Mid risk	336	33.1	73.2
High risk	272	26.8	100.0
Total	1014	100.0	

As shown in Table 1, approximately 25% of the data are high risk and 75% are low and medium risk. In this study, six independent variables have been used, which are shown in Table 2, statistical details of independent data relative to the dependent variable. As can be seen in the table, there is a value difference between the means of the independent variables in both High Risk and Non-High-Risk modes. The proposed algorithm is described below.

Bayesian Belief Network

At this stage, the proposed model is expressed. Bayesian belief network can be used as a prediction model when it is designed [11]. Also, the weights of Bayesian belief network with the help of conditional functions [12] from the data are calculated, so the important issue is the design of Bayesian belief network. In this research, the design of this structure is the responsibility of the particle optimization algorithm. The structure of the particle optimization algorithm is described below.

The particle optimization algorithm must specify the structure of the study variables and what level of the Bayesian belief network each variable is located. To do this, the proposed structure is defined as:

- Each variable relates only to lower-level variables.
- It is possible that some variables in the structure of the Bayesian belief network are not defined.
- All last level variables are related to the dependent variable.

Table 2: Statistical description of dependent variables [10]

Variable	Low & Mid Risk (n=742)		High Risk (n=272)		Description
	[mean±Std]	[min,max]	[mean±Std]	[min,max]	
Age	27.55±12.88	[10,70]	36.22±13.03	[12,65]	Any ages in years when a woman during pregnant.
Diastolic BP	73.3±12.39	[49,100]	85.07±14.11	[60,100]	Lower value of Blood Pressure in mmHg, another significant attribute during pregnancy.
Systolic BP	109.02±15.9	[70,140]	124.19±20.23	[83,160]	Upper value of Blood Pressure in mmHg, another significant attribute during pregnancy.
BS	7.48±1.63	[6,18]	12.12±4.17	[6.1,19]	Blood glucose levels is in terms of a molar concentration, mmol/L.
Body Temp	98.58±1.29	[98,103]	98.9±2.41	[98,103]	Body Temperature
Heartrate	73.41±7.67	[7,88]	76.74±8.7	[60,90]	A normal resting heart rate in beats per minute.

The following is an example from the Bayesian belief network. Suppose 10 independent variables x_1, \dots, x_{10} are defined and the variable y is a dependent variable. Now suppose the variables x_2, x_3 are defined at the first level, the variables x_1, x_4, x_8 at the second level and x_9, x_5 at the last level. The basic structure of the belief network is shown in Fig 1.

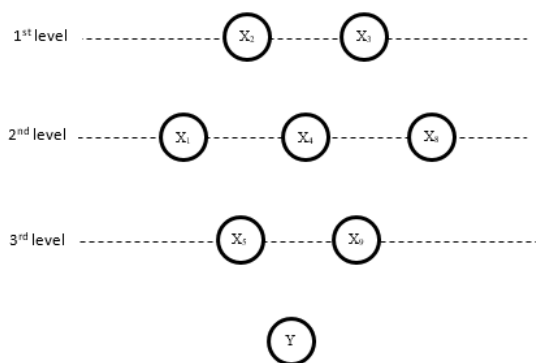


Fig 1: An example of the grading structure of belief network variables

Each variable is then linked to the next level variables. Fig 2 shows the structure of the connections in Fig 1.

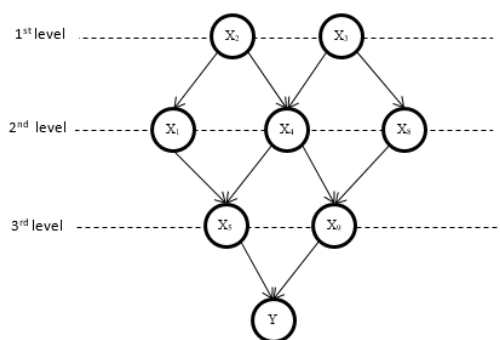


Fig 2: Completion of Bayesian belief network communications according to the example of Fig 1

The particle optimization algorithm determines how to rank the independent variables so that the

structure of the Bayesian belief network is optimally designed. The optimization parameter in this issue is considered to be the classification accuracy.

Proposed particle structure

The particle length is defined as the number of variables in the problem. Each particle points to a variable and its value is defined between zero and the number of variables. For example, if there are 10 independent variables in the problem, the particle length is defined as 10. And the values of each of its cells can be values between 0 and 10. If zero is considered, it means not considering it is a variable, and if it is five, it means that this variable is at level five. The position shown in Fig 3 is a two-particle structure proverb.

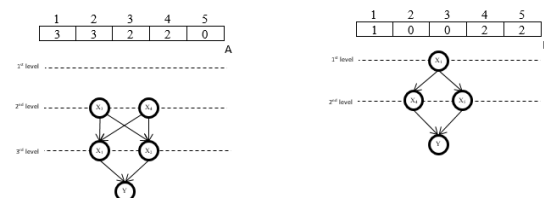


Fig 3: Drawing two belief networks with the help of artificial particles

As shown in Fig 3, some surfaces may be node-free, so those surfaces are removed. For example, in Part A of Fig 3, level one has no nodes, so this level is removed. The number of artificial particles is continuous, so by the round function the amount of continuous space enters the discrete space. Generates a particle optimization algorithm with different Bayesian networks so that the best Bayesian belief network is constructed.

Initialization of particles

The first step in the PSO algorithm is the initialization of congestion and control parameters [13]. The position of the particles is usually determined in such a way that they uniformly cover the search space. It should be noted that the particle optimization efficiency is affected by the initial variability in

congestion; That is, how much of the search space is covered and how the particles are distributed in the search space. If the optimal points are located in areas of the search space that are not covered by the initial congestion, the particle optimization algorithm will have difficulty finding them. Particle optimization finds these optimal points only if the motion of the particles directs the algorithm search process to these uncovered points.

The appropriate initialization for the position of each particle is defined as follows in Eq 1.

$$population = x_{min} + r(x_{max} - x_{min}) \quad (1)$$

Where x_{min} and x_{max} are the minimum and maximum values for each particle, respectively, and the value of x_{min} is zero and the value of x_{max} is equal to the number of independent variables, and $r \sim U(0,1)$ has a uniform distribution.

Competency function

Each particle can draw a Bayesian belief network. This Bayesian belief network is a predictive model that is evaluated by educational data and the accuracy of the Bayesian belief network is selected as the value of the competency function. Fig 4 shows the structure of the competency function.

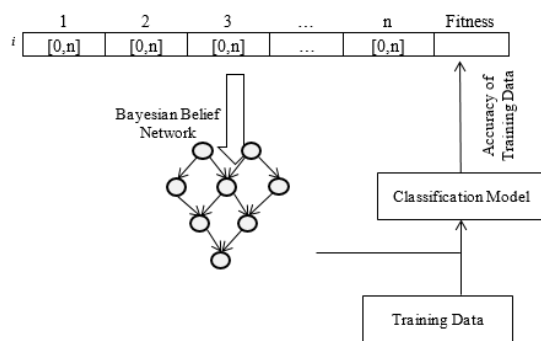


Fig 4: Proposed competency function structure

Speed control

One of the important aspects to determine the efficiency and accuracy of an algorithm is how the compromised capability between Explore and Exploit is compromised by the proposed algorithm. Explore feature is the ability of a search algorithm in different areas of the search space to find the optimal. Exploit, on the other hand, is the ability to focus the search around a potential area to improve the candidate solution. Therefore, we created a suitable solution between these two conflicting goals, which is achieved by speeding up the PSO; which is observed in Eq 2 [13].

$$v_{ij}(t + 1) = v_{ij}(t) + c_1 r_{1j}(t)[y_{ij}(t) - x_{ij}(t)] + c_2 r_{2j}(t)[\tilde{y}_j(t) - x_{ij}(t)]$$

$$v_{ij}(t + 1) = \begin{cases} v_{ij}(t + 1) & \text{if } v_{ij}(t + 1) < v_{max,j} \\ v_{max,j} & \text{if } v_{ij}(t + 1) \geq v_{max,j} \end{cases} \quad (2)$$

Where $v_{max,j}$ is the maximum velocity in the number of tables and the number of columns, the value of $v_{max,j}$ is very important. Because it speeds up the search by restraining. If the value of $v_{max,j}$ is large. The Explore feature of the algorithm will increase. While small values of this parameter enhance the local exploit capability of the algorithm. If $v_{max,j}$ is too small, the swarm may not be able to search locally well. In addition, it is possible to place the congestion in the local optimization, which will not be able to exit the algorithm. On the other hand, large values of $v_{max,j}$ carry the risk of losing good areas. Particles may jump from good solutions and search for useless areas. Large values cause the algorithm to deviate from the optimal range. In this case, the particles move faster.

Find the appropriate value of $v_{max,j}$ in order to establish two types of equilibria, which are as follows:

Move fast or slow

Explore and Exploit capability

Therefore, the value of $v_{max,j}$ is considered as a fraction of the amplitude of each dimension. Which is observed in Eq 3 [14].

$$v_{max,j} = \delta(x_{max,j} - x_{min,j}) \quad (3)$$

here $x_{max,j}$ and $x_{min,j}$ are the minimum and maximum values of the table number and column number in both dimensions, respectively, and $\delta \in (0,1)$ the value of δ is initially one and in each generation of Eq 4. The value of δ changes, the value of δ of each generation is 90% lower than the previous generation [14].

$$\delta = 0.9^i \quad i = \text{Generation number} \quad (4)$$

Algorithm stop condition

Whenever one of the following two conditions is met, the algorithm stops (this value is calculated by trial and error)

- Run 1000 generations
- 50 generations of no better answer

The structure of the proposed algorithm

It consists of the following six steps:

Step 1: Initialize the initial particle parameters according to Eq 1.

Step 2: Update the best local position of the i particle for all particles.

Step 3: Optimize the best global position of all particles.

Step 4: Calculate the New Velocity of All Particles Using Eq 2.

Step 5: Calculate the New Location of All Particles Using Eq 5.

$$x_i(t + 1) = x_i(t) + v_i(t + 1) \tag{5}$$

Step 6: Repeat steps 2 to 5 until the exit condition is met.

RESULTS

At this stage, the proposed model is evaluated. In the first stage, the data is divided into two sets of training and testing. Then, the Bayesian belief network is created by the training data, and two evaluations are performed in the training and testing sections. The data is divided into two sets of training and testing by the method of 30-70. Out of 1014 data codes, 710 records for training data and 304 records for test data are considered, then the particle optimization algorithm is executed on 710 records. And the output of Bayesian network optimization algorithm is this network is evaluated by training and testing data. The process of generating Bayesian belief network and evaluating it is performed 100 times and the average accuracy of 100 times the execution of the algorithm is considered as output. The accuracy formula is stated in Eq 6. In this regard, accuracy is defined as the amount of correct prediction to the total number of predictions. In the particle optimization algorithm, the parameter number of particles must be adjusted. This value changes from 10 to 100 with step 10. The output of the algorithm for the study data is stated below. Therefore, the boxplot diagram for training and test data is drawn with the help of five parameters: minimum output, first quarter, average, third quarter and best quarter output. Fig 5 shows the accuracy of the training data output and Fig 6 shows the accuracy of the test data.

$$Accuracy = \frac{\text{number of correctly predictoins}}{\text{Total number of predictoins}} \tag{6}$$

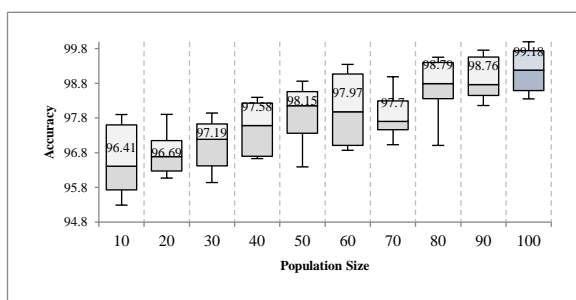


Fig 5: Evaluation of output algorithm for educational data using different initial population

Fig 5 and 6 show that as the initial population increases, the particle optimization algorithm for training and test data increases. 99.18% accuracy is obtained for educational data and 98.32% for test ata. As shown in Fig 6, the best accuracy is expressed with an initial population of 100 and an average accuracy of 98.32%. The minimum accuracy belongs to the

initial population of 10 with a value of 95.72%.

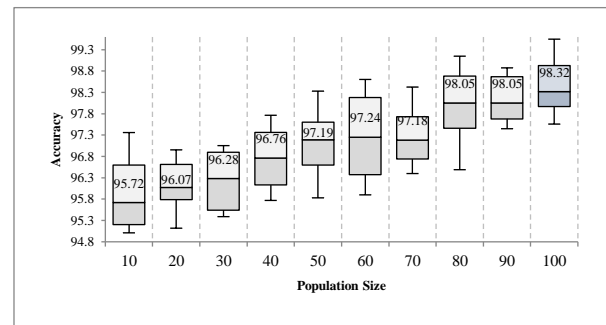


Fig 6: Evaluation of output algorithm for test data using different initial population

In the following, the proposed algorithm is examined by other methods. In Table 3, the proposed algorithm is expressed by other similar methods.

Table 3: Comparison of the proposed model with other methods

Row	Method	Accuracy
1	Decision Tree [5]	97.00%
2	Logistic Decision Tree [7]	97.96%
3	Artificial Intelligent [8]	90.00%
4	Recommended Algorithm	98.32%

As can be seen in Table 3, the proposed method performed better than the other three methods, and the accuracy of the model was 8.32% to 0.36% better.

DISCUSSION

Early and early detection of high-risk pregnancies will reduce maternal mortality and maternal morbidity during pregnancy. In this study, it was shown that data mining algorithms can be a predictive model for high-risk pregnancies. Presented. In this study, it was found that Bayesian belief network can be a good model for predicting high-risk pregnancies.

It was also found that the structure of Bayesian belief network can be designed with the help of optimization algorithms. In this study, a new structure for designing Bayesian belief network with the help of particle optimization algorithm was presented and it was found that this new structure can model Suitable for Bayesian network design.

It should be noted that the proposed hybrid algorithm is implemented with MATLAB language and another important point is that the selection of the desired factors in the input of the algorithm is very important, because the output is based on these input factors and any changes in this Factors can be applied and the output result was calculated using trial and error and its value was optimized.

In this study, an error of 1.7% was reached and

compared to similar methods, it worked between 8% and 0.5%. In the article [7] it was stated that the best method for this problem is logistic regression decision tree algorithm, but in this research, it was found that Bayesian belief network can perform better. In studies [15-18], the importance of Bayesian belief network in diagnostic fields was expressed, which was proved in this study.

CONCLUSION

According to the research presented in this article, the use of data mining technique is a good way to predict high-risk pregnancies of pregnant mothers, the results of the combined hybrid algorithm indicate this. Also, in order to do this, classification was used first and by stating the reasons, it was stated that using the combination of particle optimization and Bayesian belief network has many advantages over other classification methods and understanding the output result in it. It is far simpler and more understandable than other algorithms.

REFERENCES

- Zhu X, Chen M, Wang H, Guo Y, Chan MHK, Yan H, et al. Clinical utility of expanded non-invasive prenatal screening and chromosomal microarray analysis in high-risk pregnancy. *Ultrasound Obstet Gynecol.* 2021; 57(3): 459-65. PMID: 32198896 DOI: 10.1002/uog.22021 [PubMed]
- Holness N. High-risk pregnancy. *Nurs Clin North Am.* 2018; 53(2): 241-51. PMID: 29779516 DOI: 10.1016/j.cnur.2018.01.010 [PubMed]
- da Silva KDL, Fernandes FEM, de Lima Pessoa T, Lima SIVC, Oliveira AG, Martins RR. Prevalence and profile of adverse drug reactions in high-risk pregnancy: A cohort study. *BMC Pregnancy Childbirth.* 2019; 19(1): 199. PMID: 31185941 DOI: 10.1186/s12884-019-2321-8 [PubMed]
- Bezhenar VF, Ivanova LA, Grigoryev SG. High-risk pregnancy and perinatal losses. *Akusherstvo Ginekologiya (Russian Federation).* 2020; 2020(3): 42-7.
- Ahmed M, Kashem MA. IoT based risk level prediction model for maternal health care in the context of Bangladesh. *International Conference on Sustainable Technologies for Industry. IEEE;* 2020.
- Sumner M, Frank E, Hall M. *Speeding up logistic model tree induction. European Conference on Principles of Data Mining and Knowledge Discovery. Springer;* 2005.
- Ahmed M, Kashem MA, Rahman M, Khatun S. Review and analysis of risk factor of maternal health in remote area using the Internet of Things (IoT). *International Conference on Electrical, Control & Computer Engineering. Springer;* 2019.
- Vemuri PK, Kunta A, Challagulla R, Bodiga S, Veeravilli S, Lakshmi Bodiga V, et al. Artificial intelligence and internet of medical things based health-care system for real-time maternal stress-strategies to reduce maternal mortality rate. *Drug Invention Today.* 2020; 13(7): 1126-9.
- Ahouz F, Sadehvand M, Golabpour A. Extracting rules for diagnosis of diabetes using genetic programming. *International Journal of Health Studies.* 2019; 5(3): 1-10.
- Ahmed M. Maternal health risk data set data set [Internet]. 2020 [cited: 1 Oct 2020]. Available from: <https://archive.ics.uci.edu/ml/datasets/Maternal+H+ealth+Risk+Data+Set>
- Premchaiswadi W. *Bayesian Networks. IntechOpen;* 2012.
- Goovaerts P. Comparative performance of indicator algorithms for modeling conditional probability distribution functions. *Mathematical Geology.* 1994; 26(3): 389-411.
- Clerc M. *Particle swarm optimization. Wiley;* 2013.
- Sun J, Lai CH, Wu XJ. *Particle swarm optimisation: Classical and quantum perspectives. CRC Press;* 2016.
- Vagnoli M, Remenyte-Prescott R, Andrews J. A Bayesian belief network method for bridge deterioration detection. *Journal of Risk and Reliability.* 2021; 235(3): 338-55.
- Eyre RW, House T, Gómez-Olivé FX, Griffiths FE. Bayesian belief network modelling of household food security in rural South Africa. *BMC Public Health.* 2021; 21(1): 935. PMID: 34001089 DOI: 10.1186/s12889-021-10938-y [PubMed]
- Simsek S, Dag A, Tiahr T, Oztekin A. A Bayesian belief network-based probabilistic mechanism to determine patient no-show risk categories. *Omega;* 2021.
- Fasaee MAK, Berglund E, Pieper KJ, Ling E, Benham B, Edwards M. Developing a framework for classifying

water lead levels at private drinking water systems: A Bayesian Belief Network approach. Water Research.

2021; 189: 116641.