http:// ijp.mums.ac.ir
Original Article (Pages: 14842-14855)

Using Bagging Neural Network to Predict the Factors Affecting Neonatal Mortality

Somayeh Heshmat Alvandi[1], Morteza Ghojazadeh[2], Mohammad Heidarzadeh[3], Saeed Dastgiri[4], Hooman Nateghian[5]

[1] Research Center for Evidence-Based Medicine, Iranian EBM Centre: A Joanna Briggs Institute Affiliated Group, Tabriz University of Medical Sciences, Tabriz, Iran.

[2] Research Development & Coordination Center, Tabriz University of Medical Sciences, Tabriz, Iran.

[3] Ministry of Health, Tehran, Iran.

[4] Tabriz Health Services Management Research Center, Tabriz University of Medical Sciences, Tabriz, Iran.

[5] Research Center for Evidence-Based Medicine, Iranian EBM Centre: A Joanna Briggs Institute Affiliated Group, Tabriz University of Medical Sciences, Tabriz, Iran.

Abstract

Background: The rate of neonatal mortality is one of the main indices of health, treatment, and development in societies. It reflects the quality of nutrition and life of mothers as well as the rate of healthcare services that mothers and children are provided with by societies. This study aimed to identify the factors affecting neonatal mortality by using a bagging neural network in Rapidminer Software.

Methods: The study was conducted on 8053 births (including 1605 death cases and 6448 control cases) all over Iran in 2015. Factors such as maternal risk factors, mother's age, gestational age, child gender, birth weight, birth order, and congenital anomalies were utilized as the predictor variables of the bagging neural network. Some criteria, including the area under the ROC curve, as well as the property and sensitivity of the bagging neural network, were compared with the neural network model. The bagging neural network with 99.24% precision rate enjoyed better results in predicting the factors affecting neonatal mortality.

Results: Our suggested method revealed that gestational age is the most significant predictor factor of a neonate's status at birth time. Besides, 1-minute Apgar, need for resuscitation, 5-minute Apgar, birth weight, congenital anomalies, and birth order, as well as diabetes and preeclampsia in mothers were identified as the most significant predicting factors after the gestational age.

Conclusion: Factors discovered in this study can be considered to decrease neonatal mortality. This can help the health of mothers' community, optimize healthcare services, and development of societies.

Key Words: Bagging Neural Network, Data Mining, Logistic Regression, Neonatal Mortality, Rapid miner.

* Please cite this article as: Heshmat Alvandi S, Ghojazadeh M, Heidarzadeh M, Dastgiri S, Nateghian H. Using Bagging Neural Network to Predict the Factors Affecting Neonatal Mortality. Int J Pediatr 2021; 9 (11): 14842-14855. DOI: **10.22038/IJP. 2021.57861.4538**

**** Corresponding Author:***

Hooman Nateghian, Research Center for Evidence-Based Medicine, Iranian EBM Centre: A Joanna Briggs Institute Affiliated Group, Tabriz University of Medical Sciences, Tabriz, Iran. Email: Hoomannt@gmail.com

Received date: May.19,2021; Accepted date:Oct.06,2021

I-NTRODUCTION

The rate of neonatal mortality (the first 28 days of birth) is one of the main indices of the quality of health and welfare of societies (1). In 2000, the United Nations declared the 15-year millennium development goals to countries. One of these goals was related to children's health and plans to decrease two-third of mortality in under-5-year children to its present rate (2). It was recognized as a criterion in classifying countries as developed or developing (3). Hence, countries were encouraged to decrease their U5MR until the year 2015. The development was achieved and the under 5-year mortality rate decreased by 52%, from an estimated rate of 90 deaths per 1000 births in 1990 to 43 deaths per 1000 births in 2015. In the same period, the yearly number of neonatal mortality decreased from 12.7 million neonates to 5.9 million ones (3, 4). In 2015, 6 million children died before reaching the age of five and 47% of these deaths were in the first 28 days of birth and the pre-neonatal period. However, the investigation of the neonatal mortality rate (NMR) in the years 1990 and 2015 revealed that the neonatal mortality rate has increased from 40% to 47% (4). We can state that the global rate of under-5-year children's mortality has decreased to 53% from 1990; hence, the neonatal mortality rate will increase from 45% in 2015 to 52% in 2030 (5). According to the reports of the World Health Organization, 4 million neonates die in the first month of their births (6). Among them, neonatal mortality in the first 24 hours of birth has the highest rate and involves 65% of neonatal mortality (7).

A wide spectrum of studies shows that the most prevalent reason for neonatal mortality in developing countries are infectious diseases, asphyxia, and congenital disorders while congenital disorders and insufficiency comprise the highest neonatal mortality rate in developed countries (8). A research report shows that numerous factors interfere with neonatal mortality such as asphyxia at birth time, low weight at birth time, prematurity, and congenital anomalies. Moreover, chronic diseases such as diabetes and preeclampsia in mothers are the main risk factors of neonatal mortality (9).

Due to the fact that decreasing neonatal mortality rate is the key priority of health in developing countries (4), predicting the reasons for neonatal mortality at the birth time in societies is a key factor in adopting proper strategies for decreasing risk factors as well as enhancing the health status of newborns, as a vulnerable group; and, finally, promoting the general status of health. The data mining of data analysis is for finding the relationships and patterns as well as discovering the hidden knowledge from the database (10). We are faced with information richness and knowledge poverty in the health domain (11). The data mining process can provide specialists in the health domain with valuable information to identify the reasons for diseases, predict, and treat them concerning governing environmental factors. Lastly, it leads to progress in health and the quality of life as well as a decrease in the general costs of health (12, 13).

In many previous studies, the prediction of neonatal mortality has been conducted by the use of logistic regression. Essomba et al. (14) employed a multivariate regression to analyze the factors related to neonatal mortality. As a result of their investigation, neonatal mortality was significantly related to prematurity. Houweling et al. (15) exploited a regression prediction model to explore neonatal mortality in countries with low and average incomes. The prediction power was 0.85 in terms of performance characteristic curve. In their approach, Fergus et al. (16) employed an artificial neural network with a precision rate of 90% to predict preterm newborns. In their research, Vincer et al. (17) used a multivariate logistic regression model with a precision rate of 60% to predict preterm neonates' mortality.

By examining previous approaches and because the consequence of neonatal mortality interferes with several factors with complex relationships, we need a powerful instrument to predict the factors influencing neonatal mortality. The examinations executed in this regard revealed that neural networks enjoy better power in discovering the complex relationships among data (18) and can identify nonlinear complex relationships or hidden patterns between dependents and independent variables (19). In this study, using a neural network with the bagging algorithm, we predicted the factors influencing the consequence of neonatal mortality more precisely than other previous studies identified the effective factors in this consequence; and the results were then compared with the neural network algorithm.

2- MATERIALS AND METHODS

2-1. The dataset

This study has a cross-sectional design and uses input data for finding the predicting factors of newborn mortality. The data used in this study is related to the medical records of pregnant mothers and neonates born in the healthcare and academic centers of Iran in 2015. The dataset of mothers and newborns was accessed via the birth registration system of the Ministry of Health.

The database of mothers and newborns comprises 1509081 records, out of which 1605 records relate to neonatal deaths all over the country. The present study has investigated the entire neonatal mortality cases all over the country. To this end, we selected a control group four times larger than the case group (6448 live births). Finally, the extracted data were divided into two groups of "live birth" and "death". The live birth group included the information of mothers whose neonates were born alive and healthy. The "death" group included those mothers whose neonates died at the first 24 hours. Stillborn records were eliminated from the study. This research made use of 26 variables to train and test the bagging neural network. They included 25 predictive variables and 1 target variable. The utilized variables included the age of mother, the weight of neonates, the gender of neonates, maternal risk factors, congenital anomalies, consanguineous marriage, the age of the gestational, birth order, number of previous deliveries, number of pregnancies, abortion, type of delivery, 1-minute Apgar, 5-minute Apgar, place of residence, neonatal resuscitation, and maternal education. We used the Rapidminer 9.2 software to design and implement the bagging neural network model.

2-2. Data Preprocessing

The data-mining operation for data processing requires appropriate data to discover and extract the hidden knowledge from data. To discover this knowledge, we should conduct some prerequisites in preparing the data. The most important data preparation operation is data cleaning. Thus, in the first stage, we investigated and preprocessed the data to eliminate noises and outlier data.

2-3. Bivariate Analysis

The purpose of cross tabulation is to show the relationship or irrelevancy among the dependent and independent variables. Therefore, we employed a T-test and Chi-square test to examine the relationship among variables and compare the control and study groups in the second stage.

2-4. Model Development

In the third stage, before starting to predict the mortality factors, binary logistic regression was used in the SPSS software (Version 22) to ensure the efficacy of the bagging neural network model. We investigated the effect of variables as well as the significant relationship of each variable (as the independent factor) to the neonates' survival statuses (as the dependent factor). In the fourth stage, the data modeling in the Rapidminer software was conducted by the use of the bagging neural network. We employed 10-fold cross-validation to train and test the BNN method. The data was divided into 10 sets, 90% of which was used for training via the BNN method and 10% was used for testing the model performance. We alternately repeated the procedure 10 times to make sure that the entire components of the dataset were chosen in the testing procedure. It is worth mentioning that, similar to the bagging neural network, the details of modeling and cross-validation in the neural network were considered in this study, as well (20).

2-5. Binary Logistic Regression Model

Logistic regression is a statistical analysis instrument employed in many epidemiologic studies to predict the relationship between predictors (e.g. diabetes, gestational age, etc.) and a predicted variable (neonatal mortality) where the dependent variable is binary. (21).

2-6. Artificial Neural Network Model

Artificial neural networks are one of the most prevalent and extensively used prediction models as well as a powerful technique for prediction and classification. In these systems, weights are adjusted by a repetitious method, so that they indicate the importance of nodes in prediction (24).

2-7. Established Neural Network Architecture

The neural network makes use of a hidden layer to compare all properties with each other in the dataset (25). This study uses a hidden layer with 16 neurons. Circles in the curve are the nodes (**Fig. 1**). The curve is developed from the left side by a node for every predictor property. Thus, there are as many nodes as the input variables in the neural network (25).

This study employs a feed-forward neural network. The network moves the information only in one direction beginning from the input nodes and via the hidden nodes to the output nodes. Thus, there is no ring in the network. In our suggested model, we used multilayer perceptron (MLP) trained by the back propagation algorithm. MLP includes several layers of nodes and every node of it uses a nonlinear activation function in the hidden layer. In many programs, this stimulus function is a sigmoid function (26). This algorithm possesses two propagation stages: from the input layer to the output layer and weights updating. In this method, to compute the rate of a network error, we compare the output with the target and this error is fed to the network once again. Then, the algorithm updates the weights of each connection to reduce the error rate. This process goes on until the network converges to a condition with a relatively slight error rate (27).

2-8. Bagging Algorithm

This algorithm is one of the most popular collaborative learning techniques able to improve the classification function of the unstable and machine-like algorithms of neural network and regression by collecting bootstraps (20,28). In the bagging method, a set of classifiers is produced from the different samples of bootstraps. Next, a final classifier is developed by the synthesis of the decisions of each classifier. In this method, the bootstrap samples are developed by the substitution from the concerned dataset and integrated by maximum voting results (28).

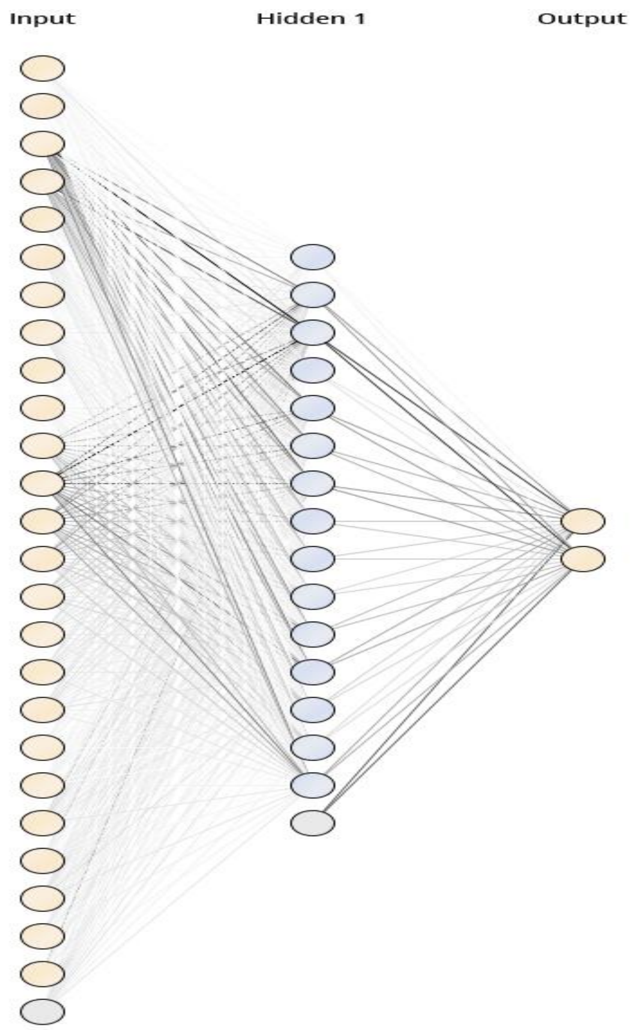

Fig. 1: The Neural Network Structure

2-9. Ensembling with Bagging Neural Networks for Neonatal Mortality Prediction

This suggested method uses a bagging classifier to improve the performance of the neural network. It employs a synthesis of several NN predictors as a base classifier to produce the final output.

2-10. Simulation in Rapid miner

The Rapid miner software is a simple platform for data input and supplies any kind of predictor model evaluation. Our previous study was done in The Rapid miner software using the bagging algorithm to enhance the predictive accuracy of neural network (20). In the present study, the neural network designed in the previous study was used to predict the factors associated with neonatal mortality.

2-11. Performance measures

The results of the study previous study (20) showed that the performance of the classifier algorithm of the bagging neural network is better than the neural network.

From these two models, the bagging neural network had higher sensitivity, specificity, precision, Kappa coefficient, and under curve area. Its precision was 99.24% (20).

The comparison of the under-curve area of the two models is presented in **Fig. 2**.

Concerning the purpose of the study to identify the factors affecting neonatal mortality at birth, we interpreted the effective factors by using the output of the bagging neural network (**Fig. 3**) (20).

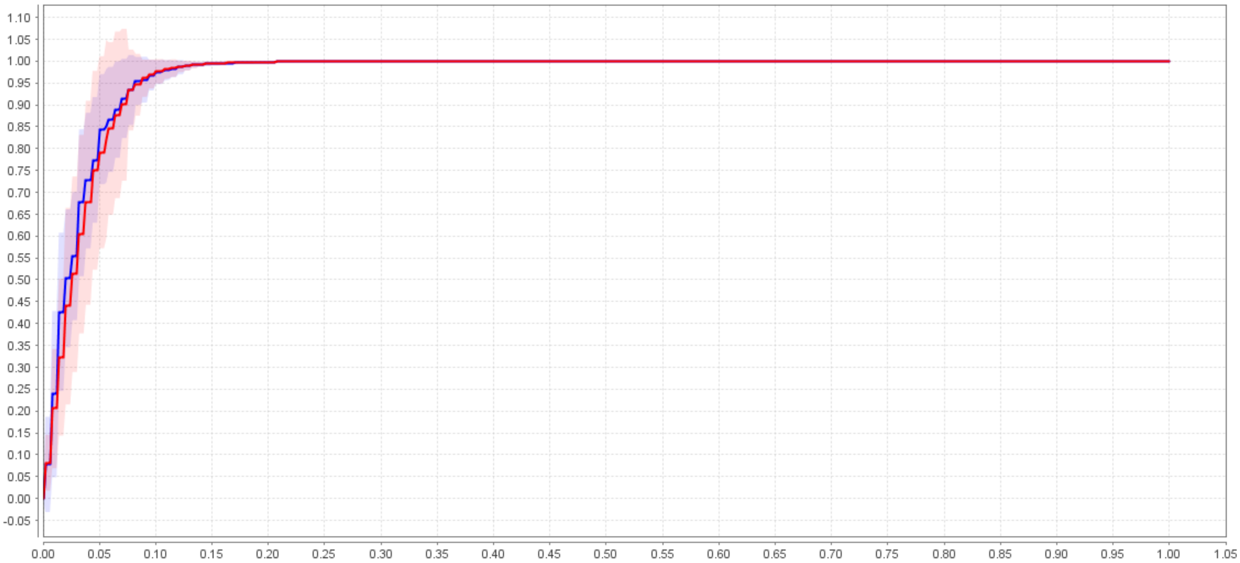

Fig. 2: The ROC Curve of the Bagging Neural Network, Neural Network

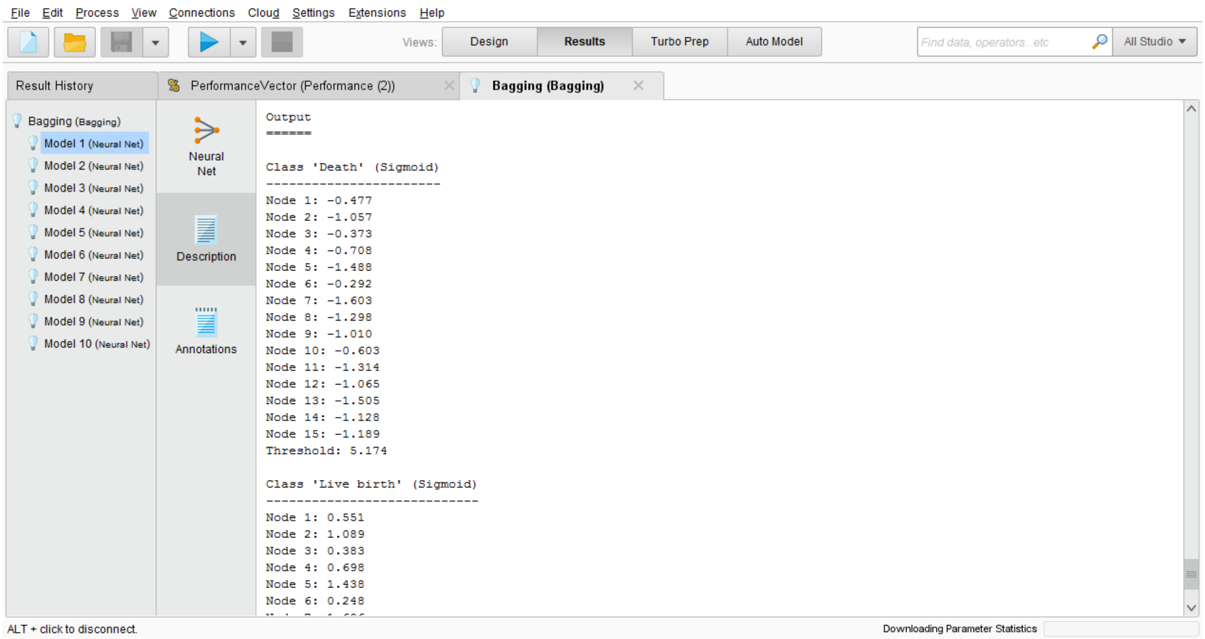

Fig. 3: A Graphical view of BNN for the neonatal mortality prediction model

3- RESULTS

3-1. Bivariate Analysis Result

8053 neonates were considered in this study. Out of 8053 born neonates, 4159 (51.8%) were male and 54.1% of neonatal mortality was among male newborns. According to the results of this study, the weighted mean of newborns was 2885.500±862.090 gr. The weighted mean among the dead newborns' group was 1689.300±1043.871 gr. The age mean of mothers was 28.032 ±6.211 years, 60% of whom were below 30 years. The age mean of mothers in the dead neonates' group was 28.245 ±6.347. The age mean of gestational was 37.100±4.398. The age mean of the mothers' gestational age in the dead neonates' group was 30.470±5.927. Besides, 16.3% of the mothers had academic education and 31.2% of them belonged to rural areas of the country. 9.7% of the mothers suffered from risk factors: 1.9% was related to diabetes, 0.4% was related to cardiac diseases, 0.1% was related to chronic hypertension, 2.2% was related to thyroid diseases, 1.6% was related to preeclampsia, 0.0% was related to the risk factors of AIDS and Syphilis in mothers, and 2.2% was related to other diseases (**Table 1).**

In this study SPSS software (version 22) was used to examine the effect of variables like mother's age, birth weight, type of delivery, congenital anomalies, mother's education, gestational age, gender, and 1-minute Apgar on neonatal mortality. Results of univariate analysis of factors associated with neonatal mortality are shown in **Table 2**.

Based on these findings and considering a $p<0.2$ as a significant result, congenital anomalies (odds ratio; OR= 74.035, confidence interval; CI= 51.697- 106.026), Neonatal Resuscitation (OR= 9.471, CI= 8.069- 11.116), Birth order (OR= 8.736, CI= 6.294- 12.124), Preeclampsia (OR= 3.378, CI= 2.384- 4.788), Diabetes (OR= 2.853, CI= 2.059- 3.953), Other diseases: (OR= 2.509, CI= 1.843- 3.417), Chronic hypertension (OR= 2.681, CI= 1.698- 4.234), and Type of delivery (OR= 2.138, CI= 1.370- 3.337) are significantly related to neonatal mortality.

3-2. Binary Logistic Regression Analysis

To examine the simultaneous effect of the variables and excluding the confounding variables, variables with a p-value less than 0.25 were included in the multivariate regression. The results are shown in **Table 3**.

The analysis of the data through logistic regression and considering a $p<0.05$ as a significant result showed that the variables of congenital anomalies (OR= 21.381, CI= 7.751-58.979), need for neonate resuscitation (OR= 4.417, CI= 3.111- 6.271), weight at the birth time (OR= 0.999, CI= 0.999-1.000), gestational age (OR= 0.695, CI= 0.624-0.774), 5-minute Apgar (OR= 0.639, CI= 0.546-0.747), and 1-minute Apgar (OR= 0.511, CI= 0.412- 0.634) had significant relationships with neonatal mortality; and they were the most effective factors in this consequence.

3-3. Prediction with Bagging Neural Network

In this study, we presented the predictor model through the bagging neural network algorithm in the Rapidminer software. The suggested model predicted the factors affecting neonatal mortality in Iran at the first 24 hours of birth and with 99.24% precision. The factors affecting neonatal mortality were investigated with respect to the bagging neural network output and were compared with the neural network (**Table 4**) (20).

According to the results of this study and through the use of the bagging neural network, the gestational age is the most important predictor of a neonate's status at birth. After that, the factors of 1-minute Apgar, need for resuscitation, 5-minute

Apgar, birth weight, congenital anomalies, birth order, mother's diabetes, and preeclampsia were identified as the most important predicting factors. In a neural network, the factors of gestational age, 1-minute Apgar, birth weight, 5-minute Apgar, need for resuscitation, birth order, and congenital anomalies were identified as the most important predicting factors.

Table-1: Descriptions of the Variables

Characteristic		Frequency	Percent	Neonatal Mortality(Yes)	Neonatal Mortality(No)
Birth weight	<2500	1485	18.4%	1155(71.9%)	330(5.1%)
	2500 - 4750	6366	79.1%	433(27.0%)	5933(92.0%)
	>4750.0	202	2.5%	18(1.1%)	184(2.9%)
				1689.300±1043.871	3183.483±460.149
Child gender	Boy	4169	51.8%	869(54.1%)	3300(51.2%)
	Girl	3845	47.7%	698(43.5%)	3147(48.8%)
	Unknown	39	0.5%	39(2.4%)	0(0.0%)
Maternal risk factors	Cardiac diseases	32	0.4%	10(0.6%)	22(0.3%)
	Diabetes	154	1.9%	63(3.9%)	91(1.4%)
	Chronic hypertension	78	1.0%	31(1.9%)	47(0.7%)
	Preeclampsia	131	1.6%	59(3.7%)	72(1.1%)
	Anemia	36	0.4%	10(0.6%)	26(0.4%)
	HIV/AIDS	3	0.0%	1(0.1%)	2(0.0%)
	Syphilis	2	0.0%	1(0.1%)	1(0.0%)
	Thyroid diseases	179	2.2%	42(2.6%)	137(2.1%)
	Other diseases	177	2.2%	67(4.2%)	110(1.7%)
Congenital anomalies	Yes	476	5.9%	443(27.6%)	33(0.5%)
	No	7577	94.1%	1163(72.4%)	6414(99.5%)
Consanguineous marriage	Yes	2196	27.3%	509(31.7%)	1686(26.2%)
	No	5857	72.7%	1907(68.3%)	4760(73.8%)
Gestational age	-	-		30.470±5.927	38.750±1.318
Mother age	-	-		28.245 ±6.347	27.980±6.176
Birth order	-	-		1.084±0.317	1.008 ±0.0912
Number of previous deliveries	-			1.110±1.368	1.031±1.207
Number of pregnancies	-			2.410±1.551	2.222±1.355
History of abortion	Yes	6704	83.2%	356(22.2%)	993(15.4%)
	No	1349	16.8%	1250(77.8%)	5454(84.6%)
Type of delivery	Vaginal	208	2.6%	23(1.4%)	185(2.9%)
	C-section	7845	97.4%	1583(98.6%)	6262(97.1%)
1-minute Apgar	-	-		2.027± 2.069	8.938. ± 0.390
5-minute Apgar	-	-		1.331± 2.548	9.910±0.472
Place of residence	Rural	2514	31.2%	528(32.9%)	1986(30.8%)
	City	5539	68.8%	1078(67.1%)	4461(69.2%)
Maternal education	Yes	1312	16.3%	241(15.0%)	1071(16.6%)
	No	6741	83.7%	1365(85.0%)	5376(83.4%)
Neonatal Resuscitation	0 (Not Resuscitate)	7002	87.0	744(46.4%)	6260(97.1%)
	1 (Initial steps)	309	3.8	126(7.8%)	183(2.8%)
	2 (Ventilation)	219	2.7	214(13.3%)	5(0.1%)
	3 (Chest compressions)	167	2.1	167(10.4%)	0(0.0%)
	4 (Drug Prescription)	354	4.4	354(22.1%)	0(0.0%)

Table-2: Results of univariate analysis of factors associated with neonatal mortality using a two-level logistic regression model

	P-value	O.R	95% C.I	
			Lower	Upper
Diabetes	0.000	2.853	2.059	3.953
Cardiac diseases	0.114	1.831	0.865	3.874
Other diseases	0.000	2.509	1.843	3.417
Chronic hypertension	0.000	2.681	1.698	4.234
Thyroid diseases	0.249	1.228	0.866	1.743
Preeclampsia	0.000	3.378	2.384	4.788
Congenital anomalies	0.000	74.035	51.697	106.026
Neonatal Resuscitation	0.000	9.471	8.069	11.116
Consanguineous marriage	0.000	1.311	1.164	1.477
Type of delivery	0.001	2.138	1.370	3.337
Number of pregnancies	0.017	1.053	1.009	1.099
Number of previous deliveries	0.000	1.098	1.058	1.139
Abortion	0.000	1.383	1.263	1.515
Gestational age	0.000	0.524	0.504	0.544
Birth weight	0.000	0.998	0.997	0.998
1-minute Apgar	0.000	0.173	0.149	0.201
5-minute Apgar	0.000	0.255	0.226	0.287
Birth order	0.000	8.736	6.294	12.124
Mother age	0.125	1.007	0.998	1.016
Maternal education	0.105	0.978	0.952	1.005
Anemia	0.242	1.548	0.745	3.217
Place of residence	0.120	0.912	0.811	1.024
VDRL	0.326	4.016	0.251	64.244
HIV	0.569	2.009	0.182	22.166
sex	0.738	0.982	0.882	1.093

Table-3: Results of multivariate analysis of factors associated with neonatal mortality using a two-level logistic regression model

	P-value	O.R	95% C.I	
			Lower	Upper
Diabetes	0.348	3.696	0.241	56.658
Cardiac diseases	0.459	0.134	0.001	27.259
Other diseases	0.900	1.188	0.081	17.472
Chronic hypertension	0.524	0.285	0.006	13.559
Thyroid diseases	0.757	0.664	0.049	8.928
Preeclampsia	0.551	0.398	0.019	8.260
Congenital anomalies	0.000	21.381	7.751	58.979
Neonatal Resuscitation	0.000	4.417	3.111	6.271
Consanguineous marriage	0.232	0.681	0.773	2.730
Type of delivery	0.291	0.769	0.472	1.252
Number of pregnancies	0.413	0.812	0.492	1.338
Number of previous deliveries	0.125	1.408	0.909	2.180
Abortion	0.769	0.827	0.233	2.933
Gestational age	0.000	0.695	0.624	0.774
Birth weight	0.026	0.999	0.999	1.000

	P-value	O.R	95% C.I	
			Lower	Upper
1-minute Apgar	0.000	0.511	0.412	0.634
5-minute Apgar	0.000	0.639	0.546	0.747
Birth order	0.260	0.336	0.050	2.244
Mother age	0.329	0.976	0.929	1.025
Maternal education	0.182	0.904	0.779	1.048
Anemia	0.599	0.108	0.000	431.059
Place of residence	0.641	0.872	0.489	1.278

Table-4: The Comparison of the Effects of Variables' in BNN and NN Based on the Weights of Variables

BNN Model		NN Model	
Factor	Weight	Factor	Weight
Gestational age*	2.928	Gestational age**	2.903
1-minute Apgar*	2.585	1-minute Apgar**	2.495
Neonatal Resuscitation*	2.349	Birth weight **	1.601
5-minute Apgar*	1.909	5-minute Apgar**	1.542
Birth weight*	1.835	Neonatal Resuscitation**	0.937
Congenital anomalies*	1.584	Birth order**	0.827
Birth order*	1.125	Congenital anomalies**	0.824
Diabetes*	0.944	Anemia**	0.820
Preeclampsia*	0.858	Syphilis**	0.815
Syphilis*	0.689	HIV/AIDS**	0.802
Anemia*	0.659	Cardiac diseases**	0.795
HIV/AIDS*	0.641	Chronic hypertension**	0.628
Cardiac diseases*	0.613	Type of delivery **	0.463
Chronic hypertension*	.482	Preeclampsia**	0.426
Type of delivery*	0.444	Other diseases**	0.212
Mother age*	0.435	Thyroid diseases**	0.178
Number of pregnancies*	0.415	Diabetes**	0.146
Consanguineous marriage*	0.383	Place of residence**	0.119
Maternal education*	0.351	Mother age**	0.084
Other diseases*	0.260	Child gender**	0.062
abortion*	0.235	Consanguineous marriage**	0.025
Place of residence*	0.214	Number of pregnancies**	0.013
Number of previous deliveries*	0.209	* Bagging Neural Network ** Neural Network	
Thyroid diseases*	0.167		
Child gender*	0.157		

The rules extracted from the bagging neural network output

if BW ≤ 2030 then death

if BW > 2445 then live birth

if BW > 2235 and MOTH_AGE ≤ 36 and APG1 > 8.5 then live birth

if BW > 2505 and NR ≤ 1 and APG1 > 6.5 then live birth

if BW ≤ 2405 and APG2 < 8 and GA ≤ 37 then death

if APG2 > 7.5 then live birth

if APG1 ≤ 6.5 then death

else live birth

The rules extracted from the bagging neural network indicate that the gestational age less than 37 weeks, 1-minute Apgar score less than 7 and 5-minute Apgar score less than 8, need for resuscitation, and the weights less than 2500 Kg are effective factors in predicting neonatal mortality at the first 24 hours of birth.

4- DISCUSSION

This study aimed to identify factors of neonatal mortality among deceased neonates at the first 24 hours of birth in Iran through a data mining technique. It employed the crucial mother and neonate factors to predict the factors affecting neonatal mortality.

We exploited a bagging neural network with 99.24% precision to predict factors affecting neonatal mortality. Besides, to prove the improvement in function, we compared the results of the suggested method with a neural network with 99.21% precision. Moreover, we employed a binary logistic regression model to investigate the correctness of estimating factors affecting neonatal mortality via a bagging neural network. As it was shown in the results section of this study, the identified factors affecting neonatal mortality via the bagging neural network corresponded to the identified factors in the neural network and binary logistic regression models.

According to the two predictor classes of death and live birth as well as the extracted rules from the bagging neural network in this study, we showed that the gestational age of less than 37 weeks is the most effective factor of neonatal mortality at the first 24 hours of birth. Furthermore, this variable has the highest effect on predicting the neonate's live birth, in the 37-41 weeks interval. The Apgar score is the best method to examine neonates' health at birth time. The 1-minute Apgar indicates that neonates need resuscitation stages and the 5-minute Apgar specifies the possibility of death and neural indispositions (29). In this study, the 1-minute Apgar score less than 7 and 5-minute Apgar scores less than 8 had a significant effect on the prediction of neonatal mortality. Besides, the 1-minute Apgar score equal to or more than 7 and 5-minute Apgar score equal to or more than 8 had a significant effect on the prediction of the live birth of neonates (30). Concerning the need for resuscitation variable, as observed in the extracted rules, when the value of this variable is less than 1, there is no need for resuscitation. When the value of this variable increases, the death probability of neonates increases, as well. Congenital anomalies (digestive, cardiac, hand and foot abnormalities, etc.) are the important factors of neonatal mortality in developing countries (9). They were also identified as the factors affecting neonatal mortality in this study. In different studies, low weight at birth time is one of the most important factors of neonatal mortality, especially in developing countries (31). In this study, low weight at the birth time was the main factor of neonatal mortality, and the effect of the weight less than 2500 Kg on neonatal mortality was obviously observed. Moreover, chronic diseases such as diabetes and preeclampsia in mothers were identified as effective risk factors in neonatal mortality.

In addition, in our study, the neonates born from mothers with a low educational level, living in rural areas were more exposed to the risk of mortality. In this research, the data mining technique on the data of Iran's Ministry of Health is presented to evaluate the factors determining neonatal mortality. Such findings contribute to the therapeutic interventionist programs, especially decreasing the risk of neonatal mortality in societies. This model can be employed as the primary warning in respect to the identification of neonates exposed to the risk of death.

5- CONCLUSION

This study identified factors of neonatal mortality at the first 24 hours of birth in Iran. We presented a predicting model through the bagging neural network in the Rapidminer software. The suggested model predicted the factors affecting neonatal mortality with 99.24% precision. Our results show that the gestational age less than 37 weeks, 1-minute Apgar scores less than 7, need for resuscitation, 5-minute Apgar score less than 8, birth weights less than 2500 Kg, congenital anomalies, and birth order are effective factors in predicting neonatal mortality at the first 24 hours of birth. In addition, maternal risk factors such as mother's diabetes and preeclampsia, are also associated with neonatal mortality.

Our findings aim to decrease neonatal mortality and the results are compatible with the findings of countries with low or average income. The results of this study can help neonatal mortality reduction and improve mothers' health in societies. This model might also be used in predicting many neonatal clinical consequences occurring in neonatal intensive care units.

6- ABBREVIATIONS

U5MR: Under-five Mortality Rate, NMR: Neonatal Mortality Rate, BNN: Bagging Neural Network, MLP: Multi-Layer Perceptron, NN: Neural Network, O.R: Odds Ratio, C.I: Confidence Interval, BW: Birth Weight, ROC: Receiver Operating Characteristic

7- CONFLICT OF INTERESTS

Hereby, the authors declare that this study has no conflict of interest with other organizations and individuals.

8- ACKNOWLEDGMENT

The authors would like to express their sincere gratitude to professors, the Ministry of Health Organization, mothers, and the personnel of healthcare centers because of their worthy cooperation during the study conduction.

9- REFERENCES

1. Ghorat F, Ghafarzadeh R, Esfehani RJ. Perinatal mortality and its associated risk factors: A study in the north-east of Iran. Iran J Neonatol 2016; 7(1):47-51.

2. Hangoma P, Aakvik A, Robberstad B. Explaining changes in child health inequality in the run up to the 2015 Millennium Development Goals (MDGs): The case of Zambia 2017;12(2):e0170995.

3. Tesfaye B, Atique S, Elias N, Dibaba L, Shabbir SA, Kebede M. Determinants and development of a web-based child mortality prediction model in resource-limited settings: A data mining approach. Comput Methods Programs Biomed 2017;140:45–51.

4. Schmidt S, Bounnack S, Hoehn T. Neonatal mortality and morbidity in the post-implementation period of a neonatal teaching program in provincial hospitals in Laos. Public Health 2018;154:123–9.

5. Rinta-Koski OP, Särkkä S, Hollmén J, Leskinen M, Andersson S. Gaussian

process classification for prediction of in-hospital mortality among preterm infants. Neurocomputing 2018;298:134–41.

6. Yamey G, Horváth H, Schmidt L, Myers J, Brindis CD. Reducing the global burden of Preterm Birth through knowledge transfer and exchange: A research agenda for engaging effectively with policymakers. Reprod Health 2016;13(1):26.

7. Kibria A, Burrowes V, Choudhury A, Sharmeen A, Ghosh S, Mahmud A, et al. Determinants of early neonatal mortality in Afghanistan : an analysis of the Demographic and Health Survey 2015. Global Health. 2018; 1; 14(1):47.

8. Debelew GT, Afework MF, Yalew AW. Determinants and causes of neonatal mortality in jimma Zone, Southwest Ethiopia: A multilevel analysis of prospective follow up study. PLoS One 2014; 9(9):e107184.

9. Babaei H, Dehghan M, Pirkashani LM. Study of Causes of Neonatal Mortality and its Related Factors in the Neonatal Intensive Care Unit of Imam Reza Hospital in Kermanshah, Iran during (2014-2016). Int J Pediatr 2018; 6(5):7641-9.

10. Aghajani S, Kargari M. Determining Factors Influencing Length of Stay and Predicting Length of Stay Using Data Mining in the General Surgery Department. Hosp Pract Res 2016; 1(2):53-8.

11. Kim J, Groeneveld PW. Big Data, Health Informatics, and the Future of Cardiovascular Medicine. J Am Coll Cardiol 2017; 69:899–902.

12. Dilsizian SE, Siegel EL. Artificial intelligence in medicine and cardiac imaging: Harnessing big data and advanced computing to provide personalized medical diagnosis and treatment. Curr Cardiol Rep 2014; 16(1):441-8.

13. Wang Y, Kung LA, Wang WYC, Cegielski CG. An integrated big data analytics-enabled transformation model: Application to health care. Inf Manag 2018; 55(1):64-79.

14. Essomba N, Koum D, Odile N, Ngwe I, Ndanga M, Ndombo PK, et al. Factors Associated With Early Neonatal Morbidity and Mortality in an Urban District Hospital. Int J Latest Res Sci Technol 2015; 5(3):43-9.

15. Houweling TAJ, Van Klaveren D, Das S, Azad K, Tripathy P, Manandhar D, et al. A prediction model for neonatal mortality in low- and middle-income countries: An analysis of data from population surveillance sites in India, Nepal and Bangladesh. Int J Epidemiol 2019; 48(1):186-98.

16. Fergus P, Idowu I, Hussain A, Dobbins C. Advanced artificial neural network classification for detecting preterm births using EHG records. Neurocomputing 2016; 188:42–9.

17. Vincer MJ, Armson BA, Allen VM, Allen AC, Stinson DA, Whyte R, et al. An Algorithm for Predicting Neonatal Mortality in Threatened Very Preterm Birth. J Obstet Gynaecol Canada 2015; 37(11):958-65.

18. Chang C-L, Liao C-S. Parameter Sensitivity Analysis of Artificial Neural Network for Predicting Water Turbidity. Int J Environ Chem Ecol Geol Geophys Eng 2012; 6(10):657-60.

19. Zhang L, Suganthan PN. A survey of randomized algorithms for training neural networks. Inf Sci (Ny) 2016; 364:146-55.

20. Alvandi SH, Pourhaji AK, Ghojazadeh M Heidarzadeh M, Dastgiri S. The Precision of Neonatal Birth Outcomes Prediction Using the Bagging Neural Network. Depiction Heal 2019; 10(2):129-43.

21. Ahmadi P, Majd HA, Khodakarim S, Tapak L, Kariman N, Amini P, et al. Prediction of low birth weight using Random Forest: A comparison with Logistic Regression. J Paramed Sci 2017; 8(3):36-43.

22. Pesantez-Narvaez J, Guillen M, Alcañiz M. Predicting motor insurance claims using telematics data—XGboost versus logistic regression. Risks. 2019; 7(2).

23. Sur P, Candès EJ. A modern maximum-likelihood theory for high-dimensional logistic regression. Proc Natl Acad Sci U S A. 2019; 116(29):14516–25.

24. Ishthaq AK, Akthar S. Survey on Artificial Neural Network Learning Technique Algorithms. Int Res J Eng Technol 2016; 33(02):36-9.

25. Çelik U, Başarır Ç. The Prediction of Precious Metal Prices via Artificial Neural Network by Using RapidMiner. Alphanumeric J 2017; 5(1):45-54.

26. Jishan ST, Rashu RI, Haque N, Rahman RM. Improving accuracy of students' final grade prediction model using optimal equal width binning and synthetic minority over-sampling technique. Decis Anal 2015; 2(1):1–25.

27. Geetha A, Nasira G. Artificial neural networks' application in weather forecasting – Using RapidMiner. Int J Comput Intell Informatics 2014; 4(3):177-82.

28. Kim MJ, Kang DK. Ensemble with neural networks for bankruptcy prediction. Expert Syst Appl 2010; 37(4):3373-9.

29. Sabol BA, Caughey AB. Acidemia in neonates with a 5-minute Apgar score of 7 or greater – What are the outcomes? Am J Obstet Gynecol 2016; 215(4):486-e1.

30. Veloso FC, Kassar LD, Oliveira MJ, de Lima TH, Bueno NB, Gurgel RQ, et al. Analysis of neonatal mortality risk factors in Brazil: a systematic review and meta-analysis of observational studies. Jornal de Pediatria (Versão em Português). 2019; 95(5):519-30.

31. Almeida MFB, Kawakami MD, Moreira LMO, Santos RM, Anchieta LM, Guinsburg R. Óbitos neonatais precoces associados à asfixia perinatal em neonatos ≥2.500 g no Brasil. J Pediatr (Rio J). 2017; 93(6):576–84.