A model to predict low birth weight infants and affecting factors using data mining techniques

Shiva Ghaderi Ghahfarokhi[1], Jamil Sadeghifar[2*], Mosayeb Mozafari[3]

1. Department of Information Technology Management, Islamic Azad University, E-Campus, Tehran, Iran
2. Department of Health Management, Faculty of Health, Ilam University of Medical Science, Ilam, Iran
3. Department of Nursing, Faculty of Nursing and Midwifery, Ilam University of Medical Science, Ilam, Iran

***Corresponding author:**Tel: +98 8432233083 Fax: +98 -
Address: Department of Health Management, Faculty of Health, Ilam University of Medical Science, Pajuhesh's Boulevard, Ilam, Iran
E-mail: Jamil.Sadeghifar@gmail.com
Received; 2018/01/14 revised; 2018/02/9 accepted; 2018/03/17

Abstract

Introduction: Birth weight is a reliable indication of intrauterine growth and determines the child's future physical and intellectual development. The purpose of this study was to use data mining technique in identifying accurate predictors of (low birth weight) LBW.
Materials and methods: This study used secondary data from 450 medical records of newborns in the educational Hospitals affiliated to Ilam University of Medical Sciences. The birth records were reviewed from April 2015 to April 2016. The checklist used to collect data comprised of two parts: demographic and effective factors (13 factors of medical and neonatal, 4 factors of mother's lifestyle and 8 about mother factors). Data were analyzed by SPSS version 21 and WEKA software.
Results: Our findings showed that mean weight of infants was 2289 ± 864 gr. The mean gestational age was 35.2 ± 4.63 weeks. 14.9% of mothers suffer from placenta previa and 14.4% suffer from preeclampsia. The results of ANOVA showed that neonatal weight was significantly higher among mothers with weight range of 84-110 Kg. The random forest algorithm showed that gestational age less than 36 weeks is main predictor and number of fetuses, preeclampsia, and premature rupture of membrane, placenta previa, the number of pregnancies and the degree of mother education were other predictors of low birth weight.
Conclusion: This study confirmed that low birth weight is a multifactorial condition requiring a systematic and accurate program to reduce LBW. Individual and group education through mass media, repeated monitoring of pregnant mothers, activation of the referral system and pursuit of a family health care technician may reduce prevalence of LBW.

Keywords: Low birth weight, Data mining, Gestational age

Introduction

Birth weight is a valid sign of intrauterine growth and determines the future physical and mental development of children (1). According to the WHO definitions, normal neonatal weight ranges from 2500g to 4000g and a birth weight less than 2500gr

Copyright © 2018 Journal of Basic Research in Medical Science. This is an open access article distributed under the terms of the Creative Commons Attribution 4.0 International License (https://creativecommons.org/licenses/by/4.0/) which permits copy and redistribute the material, in any medium or format, provided the original work is properly cited.

[Downloaded from jbrms.medilam.ac.ir on 2026-11-10]

[DOI: 10.29252/jbrms.5.3.1]

is considered as low birth weight"(2). It has been shown that incidence of low birth weight in developing countries (16.5%) is almost twice more than developed countries (1). Despite the recent development in medical sciences, prevalence of (low birth weight) LBW showed an increasing trend over the recent years(3). The prevalence of LBW in Iran varied from5.2% to 7.3% (4, 5).

The cause of low birth weight is multifactorial including chronic maternal disease (blood pressure, kidney disease and diabetes), maternal weight and height (less than 145cm), bleeding during pregnancy, gestational age and maternal age below 20 years old (6). Preterm birth and intrauterine growth retardation are the two important neonatal causes of low birth weight (7). Birth weight is the most constant determinant of neonatal mortality (8). Low birth weight is closely associated with restricted cognitive development, high rates of neonatal morbidity and mortality as well as chronic disease in the adulthood(2). Very low birth weight (VLBW) which is defined as weight less than 1500 gr is a strong predictor for infant's death and nervous system disorders (9,10).

Besides, low birth weight has always been an important public health issue. Low birth weight infants are prone to cerebral palsy, mental retardation, sensory and cognitive impairment, neurological disabilities, respiratory illnesses, injuries due to special care, sudden death syndrome, mistreatment of children, and inadequate maternal-child attachment (6, 11). Additionally, coping ability of such children in social, psychological and physical adjustment to environment is considerably diminished. The mortality rate of children with low birth weight during the first two years of life is higher than their counterparts. There are several other factors increasing morbidity and mortality of such children comprising of biological risks associated with inadequate respiratory and cardiovascular capacities due to prematurity and social risk associated with poverty. The prevalence of congenital in low birth weight children was reported between 3 to 7% (2,12).

Therefore, every child born with a low birth weight encounters the community and health system with a high-risk person for a longtime life. On the other hand, the high prevalence of low birth weight may indirectly reflect maternal health status, social and economic well-being of the community. Hence, obtaining accurate pattern and causes of low birth weight in the community is a key factor in adopting appropriate strategies for mitigating risk factors and improving children's health status that eventually promote public health status. Previous studies mainly focused on identifying prevalence and risk factors of low birth weight. While, we could not find any evidence depicting a predicating model of low birth weight using data mining approach. Therefore, this study aimed to determine factors affecting low birth weight and to predict low birth weight using data mining techniques.

Material and methods

This study used secondary data from medical records of newborns in the educational Hospitals affiliated to Ilam University of Medical Sciences. The birth records were reviewed from April 2015 to April 2016. A checklist in three sections was developed by researchers to measure demographic, neonatal and maternal characteristics. Face and Content validity of the checklist was assessed using expert opinions. The variables in the checklist included age, height, and maternal weight, body mass index, pregnancy rate, number of fetuses, number of pregnancies, number of births, history of abortion, pre-mature rupture of membrane, preeclampsia, placenta previa, mother's education, Mother's place of residence, type of delivery, marital marriage, infant's gender and mother's occupation. In this study, the weight less than 2500 grams was considered as low-birth weight.

Low birth weight infants were selected from the list of newborns and the checklist

[Downloaded from jbrms.medilam.ac.ir on 2026-11-10]

[DOI: 10.29252/jbrms.5.3.1]

was filled by accessing their medical records. At the same time, infants with normal birth weight were also selected as controls. Nine cases were excluded due to missing information in their medical record, which yielded into total sample size of 261 infants with low birth weight. Some of the variables intended to measure within the checklist were removed from this study because of missing information. For example, there was no record on maternal haemoglobin and hematocrit, alcohol and tobacco use. At the same time, 189 subjects were finally included from initially 270 listed normal-weight infants. Data were collected during a period of three months from March 2016 to May 2016. In total, this study was conducted among 450 subjects collected at maternity wards of Shahid Mostafa Khomeini Hospital in Ilam University of Medical Sciences.

Data were analyzed by SPSS version 21 and WEKA software. Initially, data were analyzed by SPSS using frequencies, means, correlation, t-student and one-way ANOVA.Data preparation for analysis with WEKA was done as an important first step. To ensure that the results of data mining algorithms are as accurate as possible, the hidden knowledge discovery process needs to be done in the data. At this point, raw data was first cleared, integrated, selected and converted into data that could be used by data mining algorithms. Then, data was modeled using logistic regression algorithms, simple parsing, random forest, decision tree, random tree, decision table J48, and analyzed using WEKA software. The efficiency of data mining algorithms for accuracy, sensitivity, specificity, rock-level curve (AUC), F value, precision and recall were investigated. Data were analyzed using Leave-one-out, cross-validation, and randomly divided into k sections of the same distinct. The k-1 is then used to build the training and the residual data was used to test the model. Then, another section is rotated for testing, and the rest is used to make the model, and as long as this rotation continues, all parts are used at the model test stage. The accuracy of the prediction at each stage is calculated from the test data and the overall accuracy percentage is calculated as the model's validity. The random sampling and the test on the evaluated data sequence are the percentage of the test set that is properly classified. Finally, the (AUC) technique was used to evaluate and compare the performance of prediction models. The AUC from the Rock Curve can be measured by the following equation, when (characteristic -1) = t and ROC (t) are sensitive:

$$AUC = \int_0^1 ROC(t)dt$$

As the area below the ROC curve is higher, the accuracy of the prediction model will be greater in the proper resolution of the resultant values.

Results

According to the results of this study, mean weight of 450 newborns was 2289±864 gr. In addition, the average maternal BMI was overweight and the highest frequency (31.1%) attributed to mothers with BMI greater than 30 while, the least frequency (0.2%) belonged to underweight mothers with BMI less than 20.

Mean gestational age (35.2 ± 4.63 years) at which the birth was performed was less than 37 weeks. 14.9% and 14.4% of the mothers were suffering from placenta previa and preeclampsia, respectively.

Besides, the correlational analysis showed that there was a significant and negative correlation between weight of the newborn and preeclampsia, premature rupture of membrane, placenta previa, multiple fetus, higher pregnancy, childbirth, and history of abortion. Conversely, neonatal weight was significantly and directly correlated with gestational age, maternal weight

[Downloaded from jbrms.medilam.ac.ir on 2026-11-10]

[DOI: 10.29252/jbrms.5.3.1]

and BMI. On the other hand, there was no significant correlation between neonatal weight and maternal occupation, newborn gender, marital marriage and delivery type.

ANOVA between subjects was conducted to compare the effect of maternal birth weight during pregnancy on neonatal weight in maternal weight ranged from 55-66 kg, 67-72 kg, 73-77 kg, 78-84 kg and 85-110 kg. The results of ANOVA showed that neonatal weight was significantly higher among mothers with weight range of 84-110 Kg. Furthermore, neonatal weight among mothers with weight 67-72 Kg was significantly higher than that of mothers with group weight 52-66, 73-77 and 78-84kg. The random forest algorithm predicts low birth within the diagnostic index, f-measure and accuracy. While, j48 algorithm was effective in evaluating low birth weight via sensitivity and recall indices. Lastly, forest method algorithm was effective in f-measure index and accuracy. The area under ROC curve analysis confirmed that random forest algorithm is the best approach in predicting low birth weight (Figure 1).

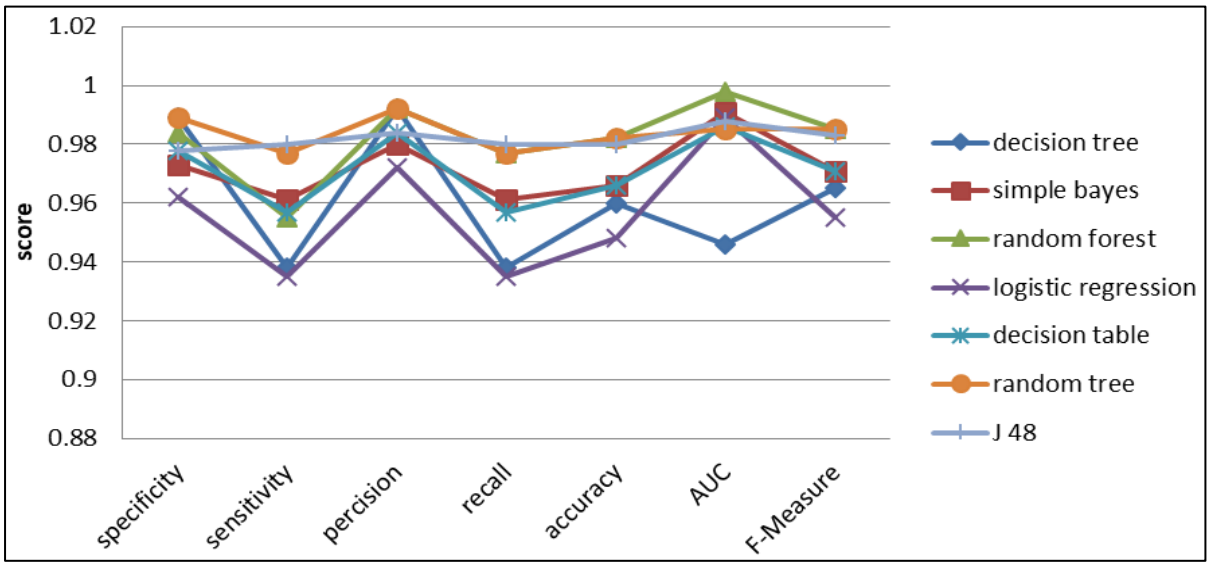


Figure 1. Comparison of the applied algorithms terms of specificity, sensitivity, precision, recall, accuracy, area under ROC curve (AUC) and F-measure.

According to the AUC, simple Bayes algorithm was placed in the second effective model in predicting low birth weight followed by logistic regression algorithm. However, the least effective method was decision tree algorithm.

The results of data mining revealed that gestational age was the most important predictor of low birth weight. It means, the likelihood of a low birth weight would be higher among pregnancies with gestational age less than 36 weeks. Accordingly, number of fetuses, preeclampsia and premature rupture of membrane, placenta previa, the number of pregnancies and the degree of mother education were predictors of low birth weight (Figure 1).

Discussion

The purpose of this study was to predict neonatal low birth weight using data mining techniques. The results showed that the most important factor in predicting LBW was gestational age. It means, the likelihood of a low birth weight would be higher among pregnancies with gestational age less than 36 weeks. Accordingly, number of fetuses, preeclampsia and premature rupture of membrane, placenta previa, the number of pregnancies and the degree of mother education were predictors of low birth weight.

One of the findings in this study was a significant relationship between maternal

[Downloaded from jbrms.medilam.ac.ir on 2026-11-10]

[DOI: 10.29252/jbrms.5.3.1]

weight and neonatal low birth weight. Similarly, Eghbalian (1) and Eshraghian et al(2)reported that higher maternal weight during pregnancy was associated with lower neonatal weight at birth. It has been shown the likelihood of low birth weight is considerably higher among mothers who gained less than 7 kg during their pregnancy (3). The amount of weight a woman gained during pregnancy indicates nutritional status which can affect neonatal weight at birth. Firouzi Jahan-Tigh et al. (2016) reported the last mother's weight before pregnancy along with age are the most important factors in predicting neonatal low birth weight (4). Similarly, Delaram (4) found the importance of maternal weight at the beginning of pregnancy and weight gain during pregnancy as the risk factors for low birth weight. There are explicit evidences favoring the existence of a meaningful relationship between maternal and infant weight (5). One explanation for such an association is the harmful effect of low Body Mass Index (BMI) or malnutrition which eventually leads to low birth weight(4). In contrast, high maternal BMI and gestational diabetes have protective effect in low birth weight (6).

Another finding from this study was the significant and inverse association between low birth weight and mother's education. Similarly, previous studies (1, 2, 4, 7) reported that the average weight of newborns in educated women was significantly higher than that in women with lower education.In contrast, Nayak et al (8), reported no relationship between maternal education and infant's weight at birth.

The current study revealed a significant relationship between preeclampsia and neonatal low birth weight. Correspondingly, Namakin et al (9) showed that the history of diseases such as diabetes, hypertension and preeclampsia increase the possibility of having a premature and low birth weight. Previous evidence showed that women with preeclampsia and hypertension had 1.8 and 4.4 times higher chance of preterm and low birth weight, respectively (10). This association could be explained by placental insufficiency caused due to hypertension which leads to preterm labor with increased chance of LBW (11). Likewise, Mosayebi et al(12) identified preeclampsia as one of the most important maternal factors that plays a role in LBW up to 46%. Findings from a prospective longitudinal study considered the pregnancy-related hypertension and placenta disorders as the main risk factors for LBW (4).

This study found a significant relationship between placenta previa and PROM with LBW. Previous findings(13)also revealed higher incidence of low birth weight among pregnancies with complications such as placenta previa and PROM. Additionally, Namakin et al(9) reported the chance of preterm birth in mothers with PROM and placenta previa were 11.9 time and 8.96 times higher than normal mothers, respectively. PROM cause intrauterine infection and oligohydramnios which consequently result in low birth weight (11).

We also found a significant relationship between the number of fetuses and LBW. The results of the Delaram (14) study showed that 79% of the twins and 100% of triplet pregnancies were categorized in low birth weight, while this rate was about 7% in monogamous pregnancies. On the other hand, the odd ratio of low birth weight in multi-fetal pregnancies was reported 16.5 times higher than single fetus.

Another highlight from this study is the significant association between LBW and history of abortion and stillbirth. The chance of LBW infant in mothers with a history of abortion has been reported 2.5 times higher than that among mothers with no history of abortion. We also found a significant relationship between gestational age and low birth weight. This finding is in line with the previous (15, 16) studies reporting a 10 times higher LBW when giving birth at gestational age less than 35 weeks.

[Downloaded from jbrms.medilam.ac.ir on 2026-11-10]

[DOI: 10.29252/jbrms.5.3.1]

Our findings fail to prove any significant relationship between maternal age and low birth weight. Conversely, previous evidences showed maternal (1, 12), especially less than 19 years old (17), is a risk factor for neonatal low birth weight. It has been shown that women younger than 20 years old are more prone to have an infant with low birth weight(18). This phenomenon could be explained due to the fact that women under 20 are still growing and the mother herself needs to receive higher amount of energy. On the other hand, women older than 35 years old are physical constrained which can affect weight of their baby (3).

This study revealed no gender differences in terms of having a low birth weight. Evidences pertaining to this findings are inconclusive since some are in favor of no gender differences (8, 19, 20)while, the others(7, 21, 22)showed that female infants are 2.5 times (4) or 1.4 times (23)more vulnerable to low birth weight compared to their counterparts. Nevertheless, the World Health Organization (WHO) reported female gender is more prone to suffer from underweight.

The limitation in this study was that we only had access to data from public hospitals and no information was available from private sector.

Conclusion

The current study used different data mining algorithms to predict infants having a LBW. Several algorithms that used in this study (including logistic regression, simple Bayes, random forest, decision tree, random tree, decision table, and J-48) indicated that the obtained results are accurate and reliable. The area under curve (AUC) confirmed that random forest algorithm was the best approach in predicting low birth weight. The area under curve provides a comprehensive assessment about precision of the screening range of threshold values for decision making. This technique is used to compare the performance of prediction models. Higher area under the ROC curve shows higher accuracy in the prediction model. According to the results of this study, the necessity of designing and implementing a systematic and accurate program to reduce LBW is essential. Several approaches could be taken to reduce LBW including individual and group education through mass media, repeated monitoring of pregnant mothers, activation of the referral system and pursuit of a family health care technician.

Acknowledgments

We would like to thank the Islamic Azad University (E-campus) for supporting this study. We would like to extend our sincere gratitude and appreciation for all of the hard work and dedication provided by the Ilam University of Medical Sciences in this study.

References

1. Shadzi SH MZ, Mostafavi F, Hassan Zadeh A. [Prevalence of low birth weight and their relation with some of maternal risk factors in Isfahan]. J Gilan Uni Medical Sci. 2000;9(33-34):55-61. (Persian)
2. Behrman RE KR, Jenson HB Obesity. In: Patrica AD. Nelson Textbook of Pediatrics. 17th ed: Philadelphia, PA Saunders; 2004.
3. Darling RD, Atav AS. Risk factors for low birth weight in New York state counties. Policy Polit Nurs Pract. 2012;13(1):17-26. doi: 10.1177/1527154412442391.
4. Hashemian Nejad N, Pejhan A, Rakhshani MH, Hoseini BL. The incidence of low birth weight (LBW) and small-for-gestational age (SGA) and its related factors in neonates, Sabzevar,

[Downloaded from jbrms.medilam.ac.ir on 2026-11-10]

[DOI: 10.29252/jbrms.5.3.1]

Iran. Int J Pediatr. 2014;2(4.2):73-78. doi: 10.22038/ijp.2014.3502.
5. Noursalehi E. [Relationship between prenatal care and low-birth-weight newborns in Guilan/Iran]. Iran J Pediatr. 2003; 13(1): 66-8. (Persian)
6. Demelash H, Motbainor A, Nigatu D, Gashaw K, Melese A. Risk factors for low birth weight in Bale zone hospitals, South-East Ethiopia : a case-control study. BMC Pregnancy Childbirth. 2015;15:264. doi: 10.1186/s12884-015-0677-y.
7. Fanaroff AAM, Fanaroff RJAA, Martin RJ. Neonatal-perinatal medicine: diseases of the fetus and infant: Mosby; 2002.
8. Mohammadian S, Vakili M, Tabandeh A. Survey of related factors in prematurity Birth. 2000.
9. Nelson E. Textbook of pediatrics/Nelson E., Richard E. Behrman MD–WB Saunders company. Philadelphia, London, Toronto. 2000:262-266.
10. Podja J, Kelley L. Administrative Committee on Coordination/Sub–Committee on Nutrition. The UN Systems Forum for Nutrition. Nutrition Policy Paper. 2000;8.
11. Stoll BJ CI. The High-Risk Infant. In: Kliegman B, Jenson, Stanton editors, editor. Nelson Text book of Pediatrics. 18th ed. Saunders Elsevier, Philadelphia USA.2007.
12. Alexander G, Wingate M, Mor J, Boulet S. Birth outcomes of Asian-Indian-Americans. Int J Gynaecol Obstet. 2007; 97(3):215-20.
13. Eghbalian F. Low birth weight causes survey in neonates. Iran J Pediatr. 2007;17(Suppl 1):27-33.
14. Eshraghian M, Abou AJ, Ghafari Z, Rajaei S. [The effects of risk factor of pregnancy period on infant's weight]. J Qazvin Uni Med Sci. 2008;11(4):60-65. (Persian)
15. Adleshoar M. [Factors predictive of underweight neonates in mothers that Referred to hospital in Rasht]. J Uni Shahid Beheshti Nurs Midwifery. 2006. 2005; 6:12-5. (Persian)
16. Firouzi Jahantigh F, Nazarnejad R, Firouzi Jahantigh M. [Investigating the Risk Factors for Low Birth Weight Using Data Mining: A Case Study of Imam Ali Hospital, Zahedan, Iran]. J Mazandaran Univ Med Sci. 2016; 25 (133):171-88. (Persian)
17. Mosayebi Z, Fakhraee SH, Movahedian AH. [Prevalence and risk factors of low birth weight infants in Mahdieh hospital, Tehran. Mosayebi Z, Fakhraee S H, Movahedian A H. Prevalence and risk factors of low birth weight infants in Mahdieh hospital, Tehran]. Feyz. 2004; 8 (2):58-67. (Persian)
18. Chen Y, Li G, Ruan Y, Zou L, Wang X, Zhang W. An epidemiological survey on low birth weight infants in China and analysis of outcomes of full-term low birth weight infants. BMC Pregnancy Childbirth. 2013; 13:242. doi: 10.1186/1471-2393-13-242.
19. Nayak RK, Metgud CS, Mallapur MD, Naik VA. Prevalence of low birth weight at Primary health centre of north Karnataka. Int J Pharm Med Bio Sci. 2013; 1:1-4.
20. Namakin K, Sharifzadeh G. [The evaluation of infants mortality causes and its related factors in Birjand]. J Isfahan Med Sch. 2009;27(95):275-82. (Persian)
21. Feresu SA, Harlow SD, Welch K, Gillespie BW. Incidence of and socio-demographic risk factors for stillbirth, preterm birth and low birthweight among Zimbabwean women. Paediatr Perinat Epidemiol. 2004;18(2):154-63.
22. Coutinho PR, Cecatti JG, Surita FG, Costa ML, Morais SS. Perinatal outcomes associated with low birth weight in a historical cohort. Reprod Health. 2011;8:18. doi: 10.1186/1742-4755-8-18.
23. Lin RX. [Maternal, medical and obstetric complications are major risk factors for low birth weight infant].

[Downloaded from jbrms.medilam.ac.ir on 2026-11-10]

[DOI: 10.29252/jbrms.5.3.1]

Zhonghua Fu Chan Ke Za Zhi. 1993; 28(1):24-6. (Chinese)
24. Ruiz RJ, Fullerton J, Dudley DJ. The interrelationship of maternal stress, endocrine factors and inflammation on gestational length. Obstet Gynecol Surv. 2003; 58(6):415-28.
25. Esmaeili MR, Haji AM, Asgardoon GH, Ghadimi R, Balegh M, Bizhani A, et al. [Effect of risk factors on low birth weight neonates]. J Babol Uni Med Sci. 2004;22(6):18-24. (Persian)
26. Delaram M, Akbari N. [Weight gain in pregnancy and its correlation with birth weight of infants]. Knowledge Health. 2008;3(2):39-43. (Persian)
27. Hajian K. [A study on mother'srisk factors for low birth weight in Babol in1999]. J Mazandaran Uni Med Si. 2001;10(26):50-4. (Persian)
28. Murphy CC, Schei B, Myhr TL, Du Mont J. Abuse: a risk factor for low birth weight? A systematic review and meta-analysis. CMAJ. 2001;164(11):1567-72.
29. Rahimi S, Moghimi R, Shadafza B, Shafaie S. The frequency of lbw and factors related to it at hospitals of tehran uniiversity of medical sciences. J Med Councile Islamic Republic of Iran (MJIRI). 2002; 20(2):83-8.
30. Ershadi A Ea, Sharif M. [Status of low birth weight infants in Kashan]. J Zanjan Uni Med Sci. 2000;8(30):54-60. (Persian)
31. Takimoto H, Yokoyama T, Yoshiike N, Fukuoka H. Increase in low-birth-weight infants in Japan and associated risk factors, 1980–2000. J Obstet Gynaecol Res. 2005;31(4):314-22.
32. Tough SC, Svenson LW, Johnston DW, Schopflocher D. Characteristic of preterm delivery and low birthweight among 113,994 infants in Alberta: 1994-1996. Can J Public Health. 2001;92(4):276-80.
33. Rafiei M. [Prevalence of low birth weight and obesity and some concomitant factors in live offspring's in 2006 and compare with 2002 result's in Arak Talleghani Hospital]. Iran J Pediatr. 2007;17(Suppl 1):47-53. (Persian)

[Downloaded from jbrms.medilam.ac.ir on 2026-11-10]

[DOI: 10.29252/jbrms.5.3.1]