Journal of Health Administration

Winter 2022, Vol 25, Issue 4

Orginal Article

A Neonatal jaundice prediction system based on the support vector machine algorithm

Tahereh Jafari [1], Somayeh Nasiri [2], Mohammadjavad Sayadi [3], Hassan Emami [4], Saman Mohammadpour [5*]

[1] *M.Sc, School of Health Management and Information Sciences, Iran University of Medical Sciences, Tehran, Iran.*
[2] *Assistant Professor, Department of Health Information Management, School of Health Management and Information Sciences, Iran University of Medical Sciences, Tehran, Iran.*
[3] *Ph.D Student, School of Health Management and Information Sciences, Iran University of Medical Sciences, Tehran, Iran.*
[4] *Assistant Professor, Department of Health Information Management and Technology, School of Paramedical Sciences, Shahid Beheshti University of Medical Sciences, Tehran, Iran.*
[5] *Assistant Professor, Department of Health Information Management and Technology, School of Paramedical Sciences, Shahid Beheshti University of Medical Sciences, Tehran, Iran.*

ARTICLE INFO

Corresponding Author:
Saman Mohammadpour
e-mail addresses:
Samanmohammadpour90 @gmail.com

Received: 21/Mar/2022
Modified: 14/Jun/2022
Accepted: 19/Jun/2022
Available online: 07/Dec/2022

Keywords:
Jaundice
Neonatal
Support Vector Machine

ABSTRACT

Introduction: Jaundice is one of the most common problems in the neonatal period, affecting about 60% of full-term and 80% of premature infants in their first week of life. The present study aimed to develop a system for predicting neonatal jaundice within the first 24 to 72 hours post-delivery by using the Support Vector Machine (SVM) algorithm.
Methods: This applied-developmental study employed a quantitative method. First, based on a literature review, a questionnaire containing effective factors for predicting jaundice in newborns was designed. Data analysis was performed using descriptive statistics, and factors that were recognized as necessary by at least 50% of the experts were included in the model. Then, data from 1178 newborns delivered at Lolagar hospital in Tehran were extracted from birth records, and several machine learning algorithms were used to predict neonatal jaundice.
Results: The findings of this research showed that the proposed model based on the SVM algorithm is the best output due to the distance between classes. Therefore, the final model of the SVM algorithm was created using the Gaussian kernel, with a sigma value of 1.2360605. Thirty percent of the samples (354 cases) were tested, and 321 cases were correctly predicted. In the proposed SVM model, parameters such as precision, the area under the Receiver Operating Characteristic (ROC), and F1 score were 92.7%, 93%, and 88% respectively.
Conclusion: Incorporating SVM into a system for predicting jaundice in newborns can aid doctors with timely prediction of jaundice in newborns and provide the possibility of taking preventive measures and preventing possible risks caused by jaundice in newborns.

Extended Abstract

Introduction

Jaundice is one of the most common problems in infancy and one of the most important causes of hospitalization in the first week of life, affecting about 60% of full-term (term) and 80% of premature (pre-term) infants. The first main symptom of this disease is the yellowing of the skin and conjunctiva due to the accumulation of bilirubin in the skin, and in most cases, it is considered a benign condition. [1-4] Untreated increases in blood bilirubin levels cause serious damage to the brain. [5] Jaundice must therefore be promptly treated to prevent dangerous complications such as encephalopathy and kernicterus, which can result in lifelong disabilities [6,7] or even be fatal and cause mental retardation and neurological disorders in infants. [8,9] The rapid prediction and treatment of neonatal jaundice is crucial and helps doctors make timely decisions regarding the appropriate treatment. [7,10] In recent years, machine learning algorithms have emerged as a highly promising and reliable tool for the accurate diagnosis and a wide range of diseases. [11-13] Studies have shown that machine learning algorithms can increase the precision and speed of classification, as compared to other existing methods. Specifically, in the context of neonatal jaundice, various algorithms, such as the k-nearest neighbor, Bayesian network, neural network, and decision tree, have been employed for disease diagnosis and prediction. [1,4,14] Meanwhile, Support Vector Machines (SVM) is one of the supervised machine learning algorithms that is used as a classifier for linear and non-linear problems, and using machine learning theories increases the precision of prediction. SVM is used for pattern recognition and classification of objects in certain classes, and it is important from the point of view that by using this algorithm, the type and severity of the disease can be recognized, and experts can be used to predict and timely treatment of diseases helped. [15-18] The present study was conducted to develop a system for predicting neonatal jaundice within the first 24 to 72 hours after birth.

Methods

The current research was of an applied-developmental type, which was carried out with a descriptive-quantitative method in two stages. The first stage was to determine the factors affecting neonatal jaundice, and the research population at this stage included six neonatologists and 13 obstetricians and gynecologists at Lolagar hospital in Tehran, and due to their small number, sampling was not done and the entire research population was examined. The data collection tool at this stage was a researcher-made questionnaire, which was designed based on a literature review on a two-choice scale (necessary or not). This questionnaire included effective factors for predicting neonatal jaundice (26 factors) and effective factors suggested by experts. The validity of the questionnaire was confirmed by four neonatologists and seven obstetricians. The questionnaire was distributed to expert doctors through e-mail so that they could specify their opinions regarding the necessity or non-necessity of each of the factors affecting neonatal jaundice. The data were analyzed in the first stage by descriptive statistics (frequency and percentage) using Excel software. Only the factors that at least 50% of the expert doctors deemed necessary were included in the next stage to create the model. [19] The second stage of the research was carried out in order to implement and evaluate the proposed model for neonatal jaundice prediction. First, newborn data were extracted from mother and baby files at Lolagar hospital, resulting in 1,328 records. The analyzed data were collected in Excel software and transferred to SQL Server

software. Then, pre-processing and data cleaning operations were performed in order to remove outliers, redundant, and noisy data. In doing so, records containing noisy and missing samples were replaced with real and correct data if possible, and those that could not be corrected were removed from the data set in order to ensure data quality and subsequently model reliability. Finally, 1,178 records were left to run the algorithm, of which 516 were positive (presence of jaundice) and 662 were negative (absence of jaundice), creating a balanced data set for modeling. The SVM algorithm, Visual Studio 2017 software, and Matlab programming language were used for data modeling. SVM is one of the classification methods for linear and non-linear functions, in which the hyperplane with the largest margin of the samples on the border between the two classes is used to separate the samples of the two classification classes. If the samples are non-linear and have a complex structure, the separation of the data in the new space is mapped through the mathematical functions of the kernel with more dimensions. [19,20] In the classification method, one of the features of the data is the class label or the target variable, which should be considered as the output field, and other features should be included in the input field as predictor variables. In this research, jaundice presence or absence in the newborn was considered the output field. Also, laboratory, clinical, and demographic factors were considered model inputs. SVM data were divided into training and testing subsets for model evaluation, with 70% of the data used for training and 30% used for testing. Model performance was evaluated by the following parameters: precision, accuracy, sensitivity, specificity, the area under the Receiver Operating Characteristic (ROC) curve, and the F1 score.

Results

Determination of effective factors for predicting neonatal jaundice: To identify the effective features for predicting the presence or absence of jaundice in newborns, we calculated the correlation of independent variables with the target column. Based on expert evaluations in the field, a total of 25 factors were identified as being effective for predicting neonatal jaundice. The list of variables selected for the modeling process is presented in Table 1.

Table 1. Effective factors for predicting neonatal jaundice

Number	Feature
1	Neonatal sex
2	Vaginal bleeding
3	Maternal hypertension
4	Gestational diabetes
5	Premature rupture of fetal membranes
6	Maternal urinary tract infection
7	Vaginal delivery
8	ABO blood group incompatibility
9	Neonatal G6PD enzyme deficiency
10	A history of jaundice in the baby's sibling(s)
11	Breastfeeding
12	Neonatal race
13	Maternal age > 25 years
14	Maternal epilepsy
15	Maternal addiction
16	Previous maternal fertility history (miscarriage and premature birth)
17	Maternal hyperthyroidism
18	Neonatal weight >2500
19	Indirect bilirubin > 12 mg/dL
20	Rh incompatibility
21	Neonatal septicemia
22	Neonatal cephalohematoma
23	Maternal gestational age < 42 weeks
24	Pyloric stenosis
25	Infant Crigler-Najjar syndrome

Implementation and evaluation of the proposed neonatal jaundice prediction model: After effective factors for predicting neonatal jaundice were determined, several machine learning algorithms were used. The results of comparing the evaluation criteria showed that the SVM algorithm was the best algorithm for this modeling. Then, according to the results of the initial evaluation and the selection of the SVM algorithm, various settings were made for the SVM kernel algorithm, and the

proposed neonatal jaundice prediction model was created using the SVM algorithm in MATLAB software. The final model was selected using this algorithm and the Gaussian kernel with a sigma of 1.2360605 as it achieved the best performance. Table 2 shows the evaluation metrics of the SVM algorithm with different Gaussian, Laplace, and Curve kernels.

Table 2. Precision, accuracy, and sensitivity evaluation metrics based on SVM algorithm parameters

Parameter c	Kernel parameter	Alpha	Sigma	Kernel Type	Precision	Accuracy	Sensitivity
2.9665	1.1091	-	-	Gaussian	0.927	0.8323	0.880
2.8798	1.4833	-	-	Gaussian	0.879	0.8283	0.8675
2.4548	2.3891	-	-	Gaussian	0.8751	0.8184	0.8558
2.3873	2.6372	-	-	Gaussian	0.8707	0.8106	0.8549
2.3314	2.6409	-	-	Gaussian	0.8655	0.7865	0.8371
2.1873	2.6449	-	-	Gaussian	0.8492	0.7768	0.8355
2.1587	2.668	-	-	Gaussian	0.8477	0.7638	0.8249
2.0857	3.0443	-	-	Gaussian	0.8356	0.7502	0.7906
2.0702	3.1021	-	-	Gaussian	0.8039	0.7384	0.7724
1.9886	3.1818	-	-	Gaussian	0.7963	0.735	0.7603
1.8823	3.4731	-	-	Gaussian	0.7845	0.6927	0.7539
1.5621	-	-	3.5919	Laplace	0.7816	0.6400	0.6041
1.5609	-	-	3.7779	Laplace	0.7656	0.6575	0.6563
1.2109	-	-	3.903	Laplace	0.7585	0.6551	0.5876
1.2106	-	32.7938	-1.2791	Curve	0.757	0.757	0.6162
1.13	-	31.126	-2.3442	Curve	0.7461	0.7513	0.6013
1.1255	-	30.8828	-1.4961	Curve	0.7212	0.6861	0.5966
1.1119	-	30.711	-2.5724	Curve	0.6958	0.6469	0.5763
1.1104	-	30.1216	-2.0562	Curve	0.6928	0.6442	0.5613

In the final evaluation and testing of the model, 354 samples (30% of the data) were applied, resulting in 33 cases of misdiagnosis. Table 3 displays the confusion matrix of the final model. Evaluation measures, including precision (92.7%), sensitivity (88%), specificity (92%), area under the ROC curve (93%), and F1 score (88%), were calculated for the test data set. The ROC curve is shown in Figure 1.

Table 3. Confusion matrix of the final neonatal jaundice prediction model

Actual class	**Predicted class**	
	Jaundice presence	Jaundice absence
Jaundice presence	125 (True positive)	17 (False negative)
Jaundice absence	16 (False positive)	196 (True negative)

The results of the model evaluation with the training data set were also used to check the presence of overfiting, and the results were not significantly different from model evaluation with the test data set.

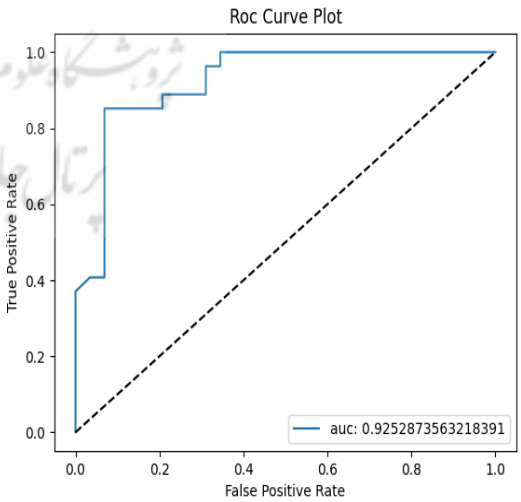


Figure 1. ROC curve for the neonatal jaundice Prediction model

After the proposed model was developed by using the same data set and labeled features, the data were run on other similar algorithms to compare and evaluate the

outputs. The comparison results of the selected algorithm (the proposed model based on the SVM algorithm) against the implemented algorithms are shown in Table 4. According to the findings in Table 4, a larger area under the ROC curve indicates a higher precision of the prediction model. The research results showed that the proposed SVM algorithm-based model achieved a higher classification precision than did other algorithms and was therefore recognized as the best output due to the distance between classes.

Table 4: Comparison of the selected algorithm with other machine learning algorithms

Algorithm	Precision	The area under the ROC curve	F1 score
SVM (proposed model)	92.7%	93%	88%
SVM regression (SMO)	91.9%	91%	86%
Simple logistic regression	92.2%	92%	83%
Decision table	92.1%	90%	85%
JRip rule-based classifier	91.5%	89%	85%
J48 classification tree	91.6%	90%	84%
Multilayer Perceptron	92.5%	93%	85%
AdaBoost	90.6%	86%	81%
Baseline	90.6%	89%	80%
Naïve Bayes	89.8%	85%	80%
Random Forest	46.0%	76%	54%

Discussion

Pediatricians have always prioritized the monitoring and treatment of jaundice in infants. [2,4] This study focuses on early prediction of jaundice in newborns, so that doctors and families can be warned in time and prepared to take appropriate measures to eliminate the risk if jaundice is predicted before or shortly after birth. Therefore, the importance of understanding these factors coupled with the need for low-cost and reliable methods can encourage researchers to employ more computational methods, such as machine learning algorithms. In most diseases, there are numerous risk factors that not only interact with one another but also have non-linear relationships. [16] The SVM algorithm is one of the machine learning algorithms that offers great capability and flexibility for solving high-dimensional problems and modeling various data in the form of linear and non-linear relationships, and therefore can be used to predict jaundice in newborns. [21,22] Most of the characteristics investigated in the present study are in line with those studied by Adebayo Idowu et al. [23] and Umer et al. [24], which indicates the importance of these characteristics in predicting neonatal jaundice. However, other characteristics, including religion, biliary tract complications, hypothermia, sepsis, asphyxia, hemolysis, being treated for jaundice, place of delivery, multiple births, anemia, prematurity, Serum Glutamic Oxaloacetic Transaminase (SGOT), and Serum Glutamic Pyruvic Transaminase (SGPT), have been used in other neonatal jaundice prediction systems, which is not consistent with our study. A study by Sussma et al. [25] showed bilirubin to be an important predictor of neonatal jaundice. Safdari et al. [14] used more features to predict neonatal jaundice, including total bilirubin levels (on the day of jaundice

occurrence and 24 hours after the initial onset of jaundice), direct bilirubin levels (on the day of jaundice occurrence, 24 hours after the initial onset of jaundice, two days after the initial onset of jaundice, and before discharge), and maximum total bilirubin levels. In another study by Ferreira et al. [1], bilirubin levels measured through skin testing (from birth to discharge, with a time interval of up to eight hours) were also considered one of the important features in the prediction model. However, the current research is not in line with the findings of the aforementioned studies [1, 14], indicating the strong effect of indirect bilirubin on neonatal jaundice prediction.

The present study demonstrated that the SVM algorithm is more efficient than other machine learning algorithms. Notably, the precision (92.7%), the area under the ROC curve (93%), and F1 score (88%) obtained in this study were higher than those achieved by other classification methods. Consistently with the present study, two previous studies have used the SVM algorithm to predict neonatal jaundice. [23,25] Umer et al. [24], who collected data from 550 medical records of 28-day-old infants with jaundice, identified 19 characteristics as effective factors. Then, in order to construct a prediction model, they utilized several algorithms, including the SVM algorithm. The SVM algorithm employed in their study resulted in a precision of 97.5% and an F1 score of 97%, which are comparable to the findings of our study. Furthermore, the study by Umer et al. [24] employed three SVM kernel functions, with the Radial Basis Function (RBF) showing the highest accuracy (98.8%) compared with other linear and polynomial functions with accuracies of 93.6% and 87.5%, respectively. Conversely, the performance of SVM algorithm parameters in our study showed that the Gaussian and the Laplace kernels delivered the best performance with the lowest error percentage. This is attributed to the SVM algorithm's underlying principle, whereby a larger distance between the classification classes leads to better outcomes. In other words, the neonatal jaundice prediction system needs a large distance between the classified classes to produce better results in complex conditions. However, in similar situations, the SVM algorithm with a curved kernel did not yield satisfactory results.

References

1. Ferreira D, Oliveira A, Freitas A. Applying data mining techniques to improve diagnosis in neonatal jaundice. BMC Med Inform Decis Mak. 2012;12:1-6.
2. Nelson WE, Kliegman R. Nelson textbook of pediatrics. 19th ed. Philadelphia: Elsevier Saunders; 2011.
3. Onyearugha CN, Chapp-Jumbo A, George IO. Neonatal jaundice: Evaluating the knowledge and practice of expectant mothers in aba, Nigeria. Journal of Health Science Research. 2016;1(2):42-7.
4. Mansor MN, Yaacob S, Muthusamy H, Nisha Basah S, Ahmad Jamil SHFS, Mohd Khidir ML, et al. PCA-based feature extraction and K-NN algorithm for early jaundice detection. International Journal of Soft Computing and Software Engineering. 2011;1(1):25-9.
5. Mansor MN, Yaacob S, Hariharan M, Basah SN, Ahmad Jamil SHFS, Mohd Khidir ML, et al. Jaundice in newborn monitoring using color detection method. Procedia Eng. 2012;29:1631-5.
6. Shukla M, Agarwal M. Knowledge of mothers regarding neonatal jaundice attending immunisation clinic at a tertiary care hospital of Lucknow. International Journal of Applied Research. 2016;2(6):297-9.
7. Arulmozhi A, Ezhilarasi M. Maximal information compression index (MICI) and PSO based detection of jaundice. Journal of Multiple-Valued Logic and Soft Computing. 2015;24(5-6):583-97.
8. Ramachandran A. Neonatal hyperbilirubinaemia. Paediatr Child Health. 2016;26(4):162-8.
9. Lauer BJ, Spector ND. Hyperbilirubinemia in the newborn. Pediatr Rev. 2011;32(8):341-9.
10. Boskabadi H, Zakeri Hamidi M, Goudarzi M. Investigating the effect of maternal risk factors in incidence of neonatal jaundice. Iranian journal of obstetrics, gynecology, and infertility. 2013;15(34):1-6. [In Persian]

11. Taylor RA, Pare JR, Venkatesh AK, Mowafi H, Melnick ER, Fleischman W, et al. Prediction of in-hospital mortality in emergency department patients with sepsis: A local big data driven, machine learning approach. Acad Emerg Med. 2016;23(3):269-78.

12. Sayadi M, Varadarajan V, Sadoughi F, Chopannejad S, Langarizadeh M. A machine learning model for detection of coronary artery disease using noninvasive clinical parameters. Life. 2022;12(11):1-10.

13. Mohammadzadeh N, mosayebi Z, Beigy H, Shojaeinia M. Prediction of sepsis due to acinetobacter infection in neonates admitted to nicu. Payavard Salamat. 2021;14(6):497-505. [In Persian]

14. Safdari R, Kadivar M, Tabari P, Shawky Own H. Comparison of data classification algorithms to determine the type of neonatal jaundice. Payavard Salamat. 2018;11(5):541. [In Persian]

15. Mueller M, Almeida JS, Stanislaus R, Wagner CL. Can machine learning methods predict extubation outcome in premature infants as well as clinicians? J Neonatal Biol. 2013;2:1-18.

16. Rezaee Kh, Haddadnia J, Rasegh Ghezelbash M. A novel algorithm for accurate diagnosis of hepatitis B and its severity. International Journal of Hospital Research. 2014;3(1):1-10.

17. Pal M, Foody GM. Feature selection for classification of hyperspectral data by SVM. IEEE Trans Geosci Rem Sens. 2010;48(5):2297-307.

18. Dormohammadi S, Alizadeh S, Asghari M, Shami M. Proposing a prediction model for diagnosing causes of infertility by data mining algorithms. Journal of Health Administration. 2014;17(57):46-57. [In Persian]

19. Farajollahi B. Presenting of prediction model for successful allogenic hematopoietic stem cell transplantation in adults with acute myeloid leukemia [master's thesis]. Theran: Iran University of Medical Sciences; 2022. [In Persian]

20. Hashemian AH, Manochehri S, Afshari D, Manochehri Z, Salari N, Shahsavari S. Prognosis of multiple sclerosis disease using data mining approaches random forest and support vector machine based on genetic algorithm. Tehran University Medical Journal. 2019;77(1):33-40. [In Persian]

21. Setareh S, Zahiri M, Zare Bandamiri M, Raeesi A, Abbasi R. Using data mining for survival prediction in patients with colon cancer. Iranian Journal of Epidemiology. 2018;14(1):19-29. [In Persian]

22. Cruz JA, Wishart DS. Applications of machine learning in cancer prediction and prognosis. Cancer Inform. 2006;2:59-77.

23. Adebayo Idowu P, Egejuru NC, Ademola Balogun J, Ajibola Sarumi O. Comparative analysis of prognostic model for risk classification of neonatal jaundice using machine learning algorithms. Computer Reviews Journal. 2019;3:122-46.

24. Umer A. Early prediction of neonatal jaundice using machine learning [master's thesis]. Ethiopia: Bahir Dar University; 2021.

25. Sussma S, Srivignesh S, Kishore VS, Marimuthu M. Jaundice prediction using machine learning approach . International Journal of Advance Research, Ideas and Innovations in Technology. 2022;7(6):493-5.

پرتال جامع علوم انسانی

نشریه مدیریت سلامت

زمستان ۱۴۰۱، دوره ۲۵، شماره ٤

مقاله اصیل

سیستم پیش‌بینی زردی نوزادان بر اساس الگوریتم ماشین بردار پشتیبان

طاهره جعفری [۱]، سمیه نصیری [۲]، محمد جواد صیادی [۳]، حسن امامی [۴]، سامان محمدپور [۵*]

[۱] کارشناس ارشد، دانشکده مدیریت و اطلاع رسانی پزشکی، دانشگاه علوم پزشکی ایران، تهران، ایران.

[۲] استادیار گروه مدیریت اطلاعات سلامت، دانشکده مدیریت و اطلاع رسانی پزشکی، دانشگاه علوم پزشکی ایران، تهران، ایران.

[۳] دانشجوی دکتری، دانشکده مدیریت و اطلاع رسانی پزشکی، دانشگاه علوم پزشکی ایران، تهران، ایران.

[۴] استادیار، گروه مدیریت و فناوری اطلاعات سلامت، دانشکده پیراپزشکی، دانشگاه علوم پزشکی شهید بهشتی، تهران، ایران.

[۵] دانشجوی دکتری، گروه مدیریت و فناوری اطلاعات سلامت، دانشکده پیراپزشکی، دانشگاه علوم پزشکی شهید بهشتی، تهران، ایران.

اطلاعـــات مقاله

نویسنده مسئول:
سامان محمدپور
رایانامه:
Samanmohammadpour90@gmail.com

وصول مقاله: ۱۴۰۱/۰۱/۰۱
اصلاح نهایی: ۱۴۰۱/۰۳/۲۴
پذیرش نهایی: ۱۴۰۱/۰۳/۲۹
انتشارآنلاین: ۱۴۰۱/۰۹/۱۶

واژه‌های کلیدی:
زردی
نوزادان
ماشین بردار پشتیبان

چکیـــــده

مقدمه: زردی یکی از مشکلات شایع دوران نوزادی است که حدود ۶۰ درصد از نوزادان رسیده و ۸۰ درصد از نوزادان نارس در هفته اول زندگی به آن مبتلا می‌شوند. مطالعه حاضر، به‌منظور ایجاد سیستمی برای پیش‌بینی زردی نوزادان در ۲۴ تا ۷۲ ساعت اول پس از تولد با بکارگیری الگوریتم ماشین بردار پشتیبان انجام شد.

روش‌ها: این مطالعه از نوع کاربردی -توسعه‌ای بود که با روش کمی انجام شد. ابتدا بر اساس بررسی متون، پرسشنامه‌ای حاوی عوامل مؤثر در پیش‌بینی زردی نوزادان طراحی شد. تحلیل داده‌ها با استفاده از آمار توصیفی انجام شد و عواملی در مدل لحاظ شد که حداقل ۵۰ درصد از متخصصان آن را ضروری تشخیص دادند. سپس، داده‌های ۱۱۷۸ نوزاد متولدشده در بیمارستان لولاگر تهران از پرونده زایمان استخراج شد و جهت پیش‌بینی زردی نوزادان، از چندین الگوریتم یادگیری ماشین استفاده شد که در این میان با توجه به نتایج حاصله جهت مدلسازی نهایی، الگوریتم ماشین بردار پشتیبان استفاده و با سنجه‌های مختلف ارزیابی گردید.

یافته‌ها: یافته‌های حاصل از این پژوهش نشان داد که مدل پیشنهادی با الگوریتم SVM به دلیل ایجاد فاصله بین کلاس‌ها به‌عنوان بهترین خروجی انتخاب شد. بنابراین، مدل نهایی الگوریتم SVM با استفاده از هسته گوسی و با سیگمای ۱/۲۳۶۰۶۰۵ ایجاد شد که ۳۰ درصد از نمونه‌ها (۳۵۴ مورد) آزمون شدند و از این تعداد ۳۲۱ مورد به درستی پیش‌بینی شد. در این مدلسازی سنجه‌های دقت، سطح زیر نمودار ROC و معیار F1 به ترتیب ۹۲/۷ درصد، ۹۳ درصد و ۸۸ درصد بدست آمد.

نتیجه‌گیری: استفاده از SVM در ایجاد سیستم پیش‌بینی زردی نوزادان می‌تواند به پزشکان در پیش‌بینی به‌موقع زردی نوزادان کمک نماید و امکان انجام اقدامات پیشگیری و جلوگیری از خطرات احتمالی ناشی از زردی نوزادان را فراهم نماید.

مقدمه

زردی (Jaundice) یکی از مشکلات شایع دوران نوزادی و از مهم‌ترین علت بستری در اوایل هفته زندگی نوزاد است که حدود ۶۰ درصد از نوزادان سالم رسیده (ترم) و ۸۰ درصد از نوزادان نارس (پری ترم) به آن مبتلا می‌شوند. اولین علامت اصلی این بیماری، زرد شدن رنگ پوست و ملتحمه چشم ناشی از تجمع بیلی‌روبین (Bilirubin) در پوست است و در اغلب موارد وضعیتی خوش‌خیم محسوب می‌شود. [۴-۱] به‌طورمعمول، زردی در نوزادانی شایع است که قبل از ۳۸ هفته بارداری به دنیا می‌آیند و در برخی از نوزادانی که از شیر مادر تغذیه می‌کنند. [۵] در شرایط طبیعی، میزان بیلی‌روبین غیرمستقیم در سرم خون بند ناف یک تا سه میلی‌گرم در دسی لیتر است و علت آن شکسته شدن گلبول‌های قرمز (Hemolysis) است [۶،۷]؛ اما اگر بیلی‌روبین سرم خون بند ناف، از پنج میلی‌گرم در دسی لیتر بالاتر رود و زردی بیش از هفت روز طول بکشد، به آن زردی پاتولوژیک (Pathological jaundice) گفته می‌شود. [۸] زردی زودرس نوزادی (Early neonatal jaundice)، معمولاً پاتولوژیک بوده و شروع آن در کمتر از ۲۴ ساعت اول پس از تولد رخ می‌دهد. [۹] علاوه بر بیلی‌روبین، سایر عوامل خطر که ممکن است منجر به زردی نوزادی شوند، عبارتند از [۹،۱۰]: ناسازگاری‌های گروه خونی و Rh، کمبود آنزیم گلوکز فسفات دی هیدروژناز (Glucose 6 Phosphate Dehydrogenase (G6PD))، عفونت مادرزادی (Congenital infection) مانند توکسوپلاسموز (Toxoplasmosis)، سرخجه، سیتومگالوویروس (Cytomegalovirus)، هرپس سیمپلکس و سیفلیس، عفونت پس از تولد، سندرم گیلن باره (Guillain-Barre syndrome)، سابقه زردی در فرزندان قبلی، فشارخون حاملگی، دیابت بارداری، تالاسمی مینور در مادر، ترخیص زودهنگام نوزادان، عفونت ادراری و سفال هماتوم (Cephalohematoma). اگر سطح بیلی‌روبین خون افزایش یابد و به‌موقع درمان نشود، صدمات جدی به مغز وارد می‌شود. [۱۱] بنابراین، در صورت عدم درمان به‌موقع، زردی می‌تواند به عوارض خطرناکی مانند انسفالوپاتی (Encephalopathy) و کرنیکتروس (Kernicterus) منجر شود که معلولیت تمام عمر را به دنبال خواهد داشت [۱۲،۱۳]؛ این امر نیز می‌تواند کشنده و عقب‌ماندگی‌های ذهنی و اختلالات عصبی را در نوزادان به همراه داشته باشد. [۱۴،۱۵] امروزه در کشورهای آسیایی و از جمله ایران، زردی به‌عنوان یکی از مشکلات شایع دوران نوزادی مطرح می‌شود. [۹] ازآنجایی‌که عوارض ناشی از این بیماری عمدتاً قابل درمان نیست [۸]؛ شناسایی و درمان زودهنگام نوزادان در معرض خطر افزایش بیلی‌روبین از اولویت بالایی برخوردار است و می‌تواند باعث جلوگیری از افزایش بیماری و عوارض بعدی شود. [۹،۱۶] اگرچه پیشرفت‌های گسترده‌ای در حوزه پزشکی صورت گرفته و درمان‌هایی برای زردی نوزادان نظیر نوردرمانی (Phototherapy)، تزریق آلبومین یا انتقال خون وجود دارد؛ اما آسیبی که توسط بیلی‌روبین به مغز حادث می‌شود، همیشه قابل جبران نیست و می‌تواند پیامد آن فلج مغزی، ناشنوایی یا کم شنوایی باشد. [۱۷] بنابراین، پیش‌بینی و درمان سریع زردی نوزادی از اهمیت زیادی برخوردار است و به پزشکان در تصمیم‌گیری به‌موقع برای انتخاب نوع درمان کمک بسیار می‌نماید. [۹،۱۳] امروزه روش‌های متعددی در پیش‌بینی بیماری‌ها بکار رفته می‌شوند که برخی از این روش‌ها مانند الگوریتم‌های یادگیری ماشین، به‌عنوان ابزاری مناسب و قابل‌اعتماد در تشخیص دقیق بسیاری از بیماری‌ها استفاده می‌شود. [۱۸-۲۰] مطالعات نشان داده است که الگوریتم‌های یادگیری ماشین می‌تواند دقت و سرعت طبقه‌بندی را نسبت به سایر روش‌های موجود افزایش دهد. در این راستا، مطالعاتی برای تشخیص و پیش‌بینی زردی نوزادی با استفاده از الگوریتم‌هایی نظیر کای نزدیک‌ترین همسایه، شبکه بیزین، شبکه عصبی و درخت تصمیم انجام‌شده است. [۱،۴،۱۷] این در حالی است که ماشین بردار پشتیبان (Support Vector Machines (SVM)) یکی از الگوریتم‌های نظارت‌شده یادگیری ماشین است که به‌عنوان

یک طبقه بند (Classifier) برای مسائل خطی و غیرخطی بکار می رود و با استفاده از تئوری‌های یادگیری ماشین، دقت پیش‌بینی را افزایش می‌دهد. SVM برای شناسایی الگو و طبقه‌بندی اشیا در کلاس‌های خاص استفاده می‌شود و از این نظر حائز اهمیت است که با استفاده از این الگوریتم، می‌توان نوع و شدت بیماری را تشخیص داد و متخصصان را در امر پیش‌بینی و درمان به‌موقع بیماری‌ها یاری نمود. [۲۴-۲۱] بر اساس مطالب یادشده، با توجه به شیوع بالای زردی و همچنین دخالت عوامل متعدد در بروز این بیماری، تشخیص زودهنگام این بیماری دشوار به نظر می‌رسد. بنابراین، مطالعه حاضر به‌منظور ایجاد سیستمی برای پیش‌بینی زردی نوزادان در ۲۴ تا ۷۲ ساعت اول پس از تولد انجام‌شده است. در این مطالعه در ابتدا از الگوریتم‌های مختلف یادگیری ماشین استفاده شد که در این میان با بررسی سنجه‌های ارزیابی، الگوریتم SVM انتخاب شد و با به‌کارگیری این الگوریتم مدل مورد انتظار با سنجه‌های ارزیابی مورد قبول پیاده‌سازی گردید تا به پزشکان در پیش‌بینی به‌موقع زردی نوزادان کمک نماید و امکان انجام اقدامات پیشگیری و جلوگیری از خطرات ناشی از زردی نوزادان را فراهم نماید.

روش‌ها

پژوهش حاضر از نوع کاربردی - توسعه‌ای بود که با روش کمی-توصیفی در دو مرحله انجام شد. مرحله اول، تعیین عوامل مؤثر بر زردی نوزادان بود که جامعه پژوهش در این مرحله شامل شش نفر متخصص نوزادان و ۱۳ نفر متخصص زنان و زایمان بیمارستان لولاگر شهر تهران بود که به دلیل تعداد کم آن‌ها نمونه‌گیری انجام نشد و کل جامعه پژوهش بررسی شد. ابزار گردآوری داده در این مرحله، پرسشنامه پژوهشگر ساخته بود که بر اساس مرور متون در مقیاس دوگزینه‌ای (ضرورت دارد یا ندارد) طراحی شد. این پرسشنامه شامل عوامل مؤثر در پیش‌بینی زردی نوزادی (۲۶ عامل) و عوامل مؤثر پیشنهادی توسط متخصصین بود. روایی پرسشنامه توسط چهار نفر از متخصصین نوزادان و هفت نفر از متخصصین زنان و زایمان تائید شد. پرسشنامه از طریق پست الکترونیک در اختیار پزشکان خبره قرار گرفت تا آن‌ها نظرات خود را نسبت به ضرورت یا عدم ضرورت هرکدام از عوامل مؤثر بر زردی نوزادی مشخص نمایند که در صورت ضرورت داشتن امتیاز یک و در صورت ضرورت نداشتن امتیاز صفر اختصاص داده شد. پس از اخذ نظرسنجی از متخصصین، تحلیل داده مرحله اول به روش آمار توصیفی (فراوانی و درصد) با استفاده از نرم‌افزار اکسل انجام شد. عواملی که حداقل ۵۰ درصد پزشکان متخصص ضروری تشخیص دادند، در مرحله بعد برای ایجاد مدل لحاظ شد. [۲۵] مرحله دوم پژوهش به‌منظور پیاده‌سازی و ارزیابی مدل پیشنهادی سیستم پیش‌بینی زردی نوزادان انجام شد. ابتدا با مراجعه به بیمارستان لولاگر، داده‌های نوزادان متولدشده از پرونده مادر و نوزاد استخراج شد که شامل ۱۳۲۸ رکورد بود. داده‌های بررسی‌شده در محیط نرم‌افزار اکسل جمع‌آوری و به نرم‌افزارSQL Server انتقال داده شد. سپس، عملیات پیش‌پردازش و پاکسازی داده به‌منظور حذف داده‌های پرت، افزونه و نویز دار انجام شد. در این راستا برای حفظ کیفیت داده جهت رسیدن به یک مدل قابل اعتماد، رکوردهای حاوی نمونه‌های نویزدار و گمشده در صورت امکان با داده‌های واقعی و درست جایگزین شد و مواردی هم که قابل اصلاح نبودند از مجموعه داده حذف شد. در نهایت ۱۱۷۸ رکورد برای اجرای الگوریتم باقی ماند که از این تعداد نمونه، ۵۱۶ مورد مثبت (وجود زردی) و ۶۶۲ مورد منفی (نبود زردی) بود و شرایط مجموعه داده برای مدلسازی متعادل بود. مدلسازی داده با استفاده از الگوریتم SVM و نرم‌افزار Visual Studio نسخه ۲۰۱۷ و زبان برنامه‌نویسی متلب (Matlab) انجام شد. SVM یکی از روش‌های طبقه‌بندی برای توابع خطی و غیرخطی است که در آن از ابر صفحه (Hyperplane) با بیشترین حاشیه از نمونه‌های موجود در مرز بین دو کلاس برای جداسازی نمونه‌های دو کلاس طبقه‌بندی استفاده می‌شود. درصورتی‌که نمونه‌ها به صورت غیرخطی و ساختار پیچیده باشد، تفکیک داده‌ها در فضای جدید از طریق توابع ریاضی هسته یا کرنل (Kernel) با ابعاد بیشتر نگاشت پیدا می‌کنند. [۲۵،۲۶] در روش طبقه‌بندی یکی از ویژگی‌های داده‌ها، همان برچسب کلاس یا متغیر هدف است

که به‌عنوان فیلد خروجی باید در نظر گرفته شود و سایر ویژگی‌ها باید در فیلد ورودی به‌عنوان متغیرهای پیش‌بینی کننده قرار گیرند. در این پژوهش وجود زردی و نبود زردی نوزاد به‌عنوان فیلد خروجی در نظر گرفته شد. همچنین، عوامل آزمایشگاهی، بالینی و جمعیت شناختی، به‌عنوان ورودی مدل لحاظ شد. پس از مدلسازی برای ارزیابی، SVM داده‌ها به دو زیرمجموعه آموزش و آزمون تقسیم شد که از داده‌های یادگیری برای آموزش (۷۰ درصد) و از داده‌های آزمون برای آزمایش مدل (۳۰ درصد) استفاده شد. سپس، ارزیابی عملکرد مدل‌ها از طریق سنجه‌های دقت (Precision)، صحت (Accuracy)، حساسیت (Sensitivity)، ویژگی (Specifity)، سطح زیر منحنی مشخصه عملکرد گیرنده (Receiver Operating Characteristic(ROC)) و معیار F1 محاسبه شد که در فرمول‌های زیر نشان داده‌شده است.

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad \text{(فرمول یک)}$$

$$\text{Accuracy} = \frac{TP + TN}{TP + FN + TN + FP} \quad \text{(فرمول دو)}$$

$$\text{Sensitivity (TPR)} = \frac{TP}{TP + FN} \quad \text{(فرمول سه)}$$

$$\text{Specifity (TNR)} = \frac{TN}{TN + FP} \quad \text{(فرمول چهار)}$$

بدست آوردن مقادیر بکار رفته در فرمول‌های ذکرشده براساس معیارهای زیر انجام می‌شود:

- مثبت واقعی (TP): نوزادانی که زردی داشته‌اند و سیستم طراحی‌شده به درستی پیش‌بینی کرده است.
- منفی واقعی (TN): نوزادانی که مبتلابه زردی نبوده‌اند و سیستم طراحی‌شده به درستی پیش‌بینی کرده است.
- مثبت کاذب (FP): نوزادانی که مبتلابه زردی نبوده‌اند؛ اما سیستم طراحی‌شده ابتلا به زردی را برای آن‌ها پیش‌بینی نموده است.
- منفی کاذب (FN): نوزادان که زردی داشته‌اند؛ اما سیستم طراحی‌شده مبتلا نبودن به زردی را برای آن‌ها پیش‌بینی کرده است.

یافته‌ها

تعیین عوامل مؤثر در پیش‌بینی زردی نوزادان:

برای شناسایی ویژگی‌های مؤثر در پیش‌بینی وجود یا فقدان زردی نوزاد از روش محاسبه همبستگی متغیرهای مستقل و ستون هدف استفاده شد. در نهایت ۲۵ عامل مؤثر در پیش‌بینی زردی نوزادی انتخاب شد که به تائید متخصصان این حوزه رسید. فهرست متغیرهای انتخاب‌شده جهت مدلسازی در جدول یک ارائه‌شده است.

جدول ۱: عوامل مؤثر در پیش‌بینی زردی نوزادان

ردیف	ویژگی
۱	جنسیت نوزاد
۲	خون‌ریزی واژینال
۳	فشارخون بالای مادر
۴	دیابت بارداری
۵	پارگی زودرس کیسه آب
۶	عفونت ادراری مادر
۷	زایمان طبیعی
۸	ناسازگاری گروه خونی ABO
۹	کمبود آنزیم G6PD نوزاد
۱۰	سابقه زردی در خواهر یا برادر نوزاد
۱۱	تغذیه نوزاد با شیر مادر
۱۲	نژاد نوزاد
۱۳	سن بالای ۲۵ سال مادر
۱۴	صرع مادر
۱۵	اعتیاد مادر
۱۶	سابقه باروری قبلی مادر (سقط و زایمان زودرس)
۱۷	پرکاری تیروئید مادر
۱۸	وزن نوزاد <۲۵۰۰
۱۹	بیلی‌روبین غیرمستقیم بیشتر از ۱۲ میلی‌گرم بر دسی لیتر
۲۰	ناسازگاری RH
۲۱	سپتی سمی نوزاد
۲۲	سفال هماتوم نوزاد
۲۳	سن حاملگی مادر کمتر از ۴۲ هفته
۲۴	تنگی پیلور نوزاد (Pyloric stenosis)
۲۵	سندرم کریگر نجار نوزاد (Crigler–Najjar syndrome)

پیاده‌سازی و ارزیابی مدل پیشنهادی سیستم پیش‌بینی زردی نوزادان: پس از تعیین ویژگی‌ها برای پیش‌بینی زردی نوزادی، از چندین الگوریتم یادگیری ماشین استفاده شد. نتایج حاصل از مقایسه سنجه‌های ارزیابی نشان داد که الگوریتم SVM بهترین الگوریتم برای این مدلسازی بوده است. سپس، با توجه به نتایج ارزیابی اولیه و انتخاب الگوریتم ماشین بردار پشتیبان، تنظیمات مختلف برای هسته الگوریتم SVM انجام شد و مدل پیشنهادی سیستم پیش‌بینی زردی نوزادی با استفاده از الگوریتم SVM در نرم‌افزار متلب ایجاد شد. بنابراین، یک مدل نهایی با استفاده از این الگوریتم و هسته گوسی با سیگمای ۱/۲۳۶۰۶۰۵ به عنوان بهترین هسته انتخاب شد. سنجه‌های ارزیابی الگوریتم SVM با هسته‌های مختلف گوسی، لاپلاس و منحنی در جدول دو نشان داده شده است.

جدول ۲: سنجه‌های ارزیابی دقت، صحت و حساسیت براساس پارامترهای الگوریتم SVM

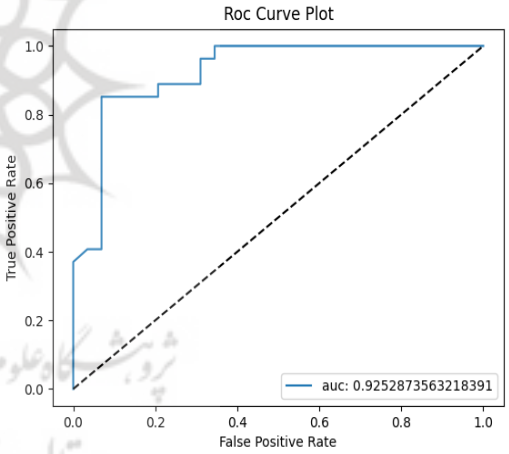
پارامتر C	پارامتر هسته	آلفا	سیگما	نوع هسته	دقت	صحت	حساسیت
۲/۹۶۶۵	۱/۱۰۹۱	-	-	گوسی	۰/۹۲۷	۰/۸۳۲۳	۰/۸۸۰
۲/۸۷۹۸	۱/۴۸۳۳	-	-	گوسی	۰/۸۷۹	۰/۸۲۸۳	۰/۸۶۷۵
۲/۴۵۴۸	۲/۳۸۹۱	-	-	گوسی	۰/۸۷۵۱	۰/۸۱۸۴	۰/۸۵۵۸
۲/۳۸۷۳	۲/۶۳۷۲	-	-	گوسی	۰/۸۷۰۷	۰/۸۱۰۶	۰/۸۵۴۹
۲/۳۳۱۴	۲/۶۴۰۹	-	-	گوسی	۰/۸۶۵۵	۰/۷۸۶۵	۰/۸۳۷۱
۲/۱۸۷۳	۲/۶۴۴۹	-	-	گوسی	۰/۸۴۹۲	۰/۷۷۶۸	۰/۸۳۵۵
۲/۱۵۸۷	۲/۶۶۸	-	-	گوسی	۰/۸۴۷۷	۰/۷۶۳۸	۰/۸۲۴۹
۲/۰۸۵۷	۳/۰۴۴۳	-	-	گوسی	۰/۸۳۵۶	۰/۷۵۰۲	۰/۷۹۰۶
۲/۰۷۰۲	۳/۱۰۲۱	-	-	گوسی	۰/۸۰۳۹	۰/۷۳۸۴	۰/۷۷۲۴
۱/۹۸۸۶	۳/۱۸۱۸	-	-	گوسی	۰/۷۹۶۳	۰/۷۳۵	۰/۷۶۰۳
۱/۸۸۲۳	۳/۴۷۳۱	-	-	گوسی	۰/۷۸۴۵	۰/۶۹۲۷	۰/۷۵۳۹
۱/۵۶۲۱	-	-	۳/۵۹۱۹	لاپلاس	۰/۷۸۱۶	۰/۶۴۰۰	۰/۶۰۴۱
۱/۵۶۰۹	-	-	۳/۷۷۷۹	لاپلاس	۰/۷۶۵۶	۰/۶۵۷۵	۰/۶۵۶۳
۱/۲۱۰۹	-	-	۳/۹۰۳	لاپلاس	۰/۷۵۸۵	۰/۶۵۵۱	۰/۵۸۷۶
۱/۲۱۰۶	-	۳۲/۷۹۳۸	-۱/۲۷۹۱	منحنی	۰/۷۵۷	۰/۷۵۷	۰/۶۱۶۲
۱/۱۳	-	۳۱/۱۲۶	-۲/۳۴۴۲	منحنی	۰/۷۴۶۱	۰/۷۵۱۳	۰/۶۰۱۳
۱/۱۲۵۵	-	۳۰/۸۸۲۸	-۱/۴۹۶۱	منحنی	۰/۷۲۱۲	۰/۶۸۶۱	۰/۵۹۶۶
۱/۱۱۱۹	-	۳۰/۷۱۱	-۲/۵۷۲۴	منحنی	۰/۶۹۵۸	۰/۶۴۶۹	۰/۵۷۶۳
۱/۱۱۰۴	-	۳۰/۱۲۱۶	-۲/۰۵۶۲	منحنی	۰/۶۹۲۸	۰/۶۴۴۲	۰/۵۶۱۳

در مدل نهایی برای ارزیابی و آزمون روی ۳۰ درصد از داده‌ها اعمال شد که شامل ۳۵۴ نمونه بود و از این تعداد در ۳۳ مورد تشخیص اشتباه حاصل شد. جدول سه، ماتریس درهم‌ریختگی (Confusion matrix) مدل نهایی را نشان

می‌دهد که در این مدلسازی سنجه‌های ارزیابی دقت (۹۲/۷ درصد)، حساسیت (۸۸ درصد)، ویژگی (۹۲ درصد)، سطح زیر نمودار ROC (۹۳ درصد) و معیار F1 (۸۸ درصد) روی مجموعه داده آزمون محاسبه گردید. نمودار ROC در شکل یک ارائه‌شده است. از نتایج ارزیابی مدل روی مجموعه داده آموزش نیز جهت بررسی وجود بیش برازش (Overfitting) استفاده شد که نتایج حاصله تفاوت آشکاری با ارزیابی مدل روی مجموعه داده آزمون نداشت.

جدول ۳: ماتریس درهم‌ریختگی مدل نهایی تشخیص زردی نوزاد

پیش‌بینی‌شده / واقعی	وجود زردی	نبود زردی
وجود زردی	۱۲۵ (مثبت واقعی)	۱۷ (منفی کاذب)
نبود زردی	۱۶ (مثبت کاذب)	۱۹۶ (منفی واقعی)



شکل ۱: نمودار ROC در مدل تشخیص زردی نوزاد

پس از ایجاد مدل پیشنهادی بر مبنای مجموعه داده یکسان و ویژگی‌های برچسب‌دار، داده‌ها بر روی سایر الگوریتم‌های مشابه اجرا شد تا خروجی‌ها با یکدیگر مقایسه و ارزیابی شد. نتایج مقایسه الگوریتم انتخابی (مدل پیشنهادی الگوریتم SVM) با الگوریتم‌های اجراشده در جدول چهار نشان داده‌شده است. با توجه به یافته‌های جدول چهار، هر چه مساحت زیر سطح منحنی ROC بیش‌تر باشد میزان دقت مدل پیش‌بینی نیز بیشتر است. براین اساس، یافته‌های پژوهش نشان داد که دقت طبقه‌بندی مدل پیشنهادی با الگوریتم SVM نسبت به سایر الگوریتم‌ها بالاتر بوده است. بنابراین، مدل پیشنهادی با الگوریتم SVM به دلیل ایجاد فاصله بین کلاس‌ها به‌عنوان بهترین خروجی شناخته شد.

جدول ۴: مقایسه الگوریتم انتخابی با سایر الگوریتم‌های یادگیری ماشین

الگوریتم	دقت	سطح زیر نمودار ROC (درصد)	معیار F1 (درصد)
SVM (Proposed model)	۹۲/۷	۹۳	۸۸
SVM regression (SMO)	۹۱/۹	۹۱	۸۶
Simple logistic regression	۹۲/۲	۹۲	۸۳
Decision table	۹۲/۱	۹۰	۸۵
JRip rule-based classifier	۹۱/۵	۸۹	۸۵
J48 classification tree	۹۱/۶	۹۰	۸۴
Multilayer Perceptron	۹۲/۵	۹۳	۸۵
AdaBoost	۹۰/۶	۸۶	۸۱
Baseline	۹۰/۶	۸۹	۸۰
Naïve Bayes	۸۹/۸	۸۵	۸۰
Random Forest	۴۶/۰	۷۶	۵۴

بحث و نتیجه‌گیری

زردی نوزادان یکی از مواردی است که همواره از لحاظ نحوه پیگیری و درمان مدنظر متخصصان اطفال بوده است. ازآنجاکه این بیماری یکی از بیماری‌های شایع نوزادان در طول هفته اول زندگی آن‌هاست. بنابراین، با توجه به شیوع بالای بیماری در نوزادان، شناخت عوارض آن و نیز بررسی راهکارهای درمانی در این زمینه از اهمیت ویژه‌ای برخوردار است. [۲،۴] در این پژوهش پیش‌بینی زودهنگام زردی نوزادان مدنظر است تا در صورت پیش‌بینی زردی، قبل از تولد و یا اندکی پس از تولد نوزاد، بتوان به پزشکان و خانواده هشدار به‌موقع و آمادگی لازم برای انجام اقدامات بعدی تا هنگام رفع خطر را داد. ازاین‌رو، شاید بتوان گفت که اهمیت شناخت این موارد از یک‌سو و نیاز به روش‌های ارزان و قابل‌اعتماد از سوی دیگر، می‌تواند محققان را به استفاده هر

چه بیشتر روش‌های محاسباتی، از جمله استفاده از الگوریتم‌های یادگیری ماشین ترغیب نماید. از سویی در بیشتر بیماری‌ها، تعداد زیادی از عوامل خطر وجود دارد که افزون برداشتن اثرات متقابل با یکدیگر، دارای روابط غیرخطی نیز هستند. [۲۲] SVM به‌عنوان یکی از الگوریتم‌های یادگیری ماشین، توانایی و انعطاف خوبی در حل مسائل با ابعاد بالا و مدل‌سازی داده‌های گوناگون به صورت روابط خطی و غیرخطی دارد. بنابراین، امکان استفاده از آن در پیش‌بینی زردی نوزادان وجود دارد. [۲۷،۲۸] در الگوریتم‌های یادگیری ماشین، آموزش باید بر مبنای اطلاعات صحیح انجام پذیرد. بنابراین، در این مطالعه برای پیش‌بینی زردی نوزادان، از نظر پزشکان متخصص و بررسی متون معتبر استفاده شد. از ۲۶ عامل مؤثر که در پیش‌بینی زردی نوزادان شناسایی شدند، پس از جمع‌آوری پرسشنامه‌ها و تحلیل نتایج، ۲۵ ویژگی مهم (شامل عوامل بالینی، آزمایشگاهی و جمعیت شناختی) برای سیستم پیش‌بینی استخراج شد که در دو کلاس نوزادان بیمار مبتلابه زردی و فاقد زردی دسته‌بندی شد. بسیاری از ویژگی‌های بررسی‌شده در مطالعه حاضر با مطالعات ادبایو ایدوو و همکاران [۲۹] و اومر [۳۰] همراستا است که بیانگر اهمیت این ویژگی‌ها (جنسیت نوزاد، سن مادر، سن حاملگی مادر، G6PD، ناسازگاری گروه خونی و RH، وزن نوزاد، تغذیه با شیر مادر، سندرم کریگر نجار، مشکل تیروئید، عفونت مجاری ادراری، سابقه زردی در خواهر و برادر نوزاد و روش زایمان) در زمینه پیش‌بینی زردی نوزادی است. از طرف دیگر در میان مطالعات مذکور از ویژگی‌های دیگری نظیر مذهب، عارضه مجاری صفراوی، هیپوترمی، سپسیس، خفگی، همولیز، تحت درمان زردی، محل زایمان، چندقلوزایی، کم‌خونی، نارس بودن، SGPT و SGOT برای سیستم پیش‌بینی زردی نوزادی استفاده‌شده است که با پژوهش حاضر همخوانی ندارد. یافته‌های مطالعه ساسما و همکاران [۳۱] نشان داد که بیلی‌روبین ویژگی مهمی است که بر پیش‌بینی زردی نوزادی تأثیرگذار بوده است. مطالعه صفدری و همکاران [۱۷] نیز از ویژگی‌های بیشتری برای پیش‌بینی زردی نوزادی استفاده نموده‌اند که عبارتند از: میزان بیلی‌روبین توتال (Total bilirubin) (در روز بروز زردی و ۲۴ ساعت بعد از بروز اولیه زردی)، میزان بیلی‌روبین مستقیم (Direct bilirubin) (در روز بروز زردی، ۲۴ ساعت بعد از بروز اولیه زردی، دو روز بعد از بروز اولیه زردی و قبل از ترخیص) و حداکثر میزان بیلی‌روبین توتال. همچنین، مطالعه فریرا و همکاران [۱] مقدار سطح بیلی‌روبین از طریق پوست (هنگام تولد تا زمان ترخیص با فاصله زمانی حداکثر هشت ساعت) را به عنوان یکی از ویژگی‌های مهم در مدل پیش‌بینی در نظر گرفته است. این در حالی است که پژوهش حاضر با یافته‌های مطالعات مذکور [۱،۱۷] همراستا نیست و حاکی از اثرگذاری بالای ویژگی بیلی‌روبین غیرمستقیم (Indirect bilirubin) بر مدل پیش‌بینی زردی نوزادی است. یافته‌های پژوهش حاضر نشان داده است که SVM نسبت به سایر الگوریتم‌های یادگیری ماشین کارایی بیشتری داشته است. به‌طورکلی دقت (۹۲/۷ درصد)، سطح زیر منحنی (۹۳ درصد) و معیار F1 (۸۸ درصد) بیشتر از سایر روش‌های طبقه‌بندی بوده است. همانند مطالعه حاضر، دو مطالعه از الگوریتم SVM به‌منظور پیش‌بینی زردی نوزادی استفاده نموده‌اند. [۳۰،۳۲] در مطالعه اومر [۳۰] داده‌ها از پرونده ۵۵۰ نوزاد ۲۸ روزه مبتلابه زردی با ۱۹ ویژگی (عامل مؤثر) جمع‌آوری و در فایل اکسل ذخیره شد. سپس، به‌منظور ایجاد مدل پیش‌بینی از الگوریتم‌های متعدد از جمله SVM استفاده‌شده است که مقایسه میزان دقت و معیار F1 در الگوریتم SVM در مطالعه مذکور به ترتیب ۹۷/۵ و ۹۷ درصد بود که در مقایسه با مطالعه حاضر بیشتر است. از طرف دیگر در مطالعه اومر [۳۰] از سه تابع هسته SVM استفاده شد که از میان آن‌ها تابع پایه شعاعی (Radial Basis Function (RBF)) با بیشترین صحت (۹۸/۸ درصد) به‌عنوان بهترین هسته نسبت به سایر توابع خطی (Linear) و چندجمله‌ای (Polynomial) به ترتیب با صحت ۹۳/۶ و ۸۷/۵ درصد انتخاب شد. درحالی‌که عملکرد پارامترهای الگوریتم SVM در مطالعه حاضر نشان داد که از بین الگوریتم‌های اجراشده بر روی مجموعه داده‌های یکسان، بهترین هسته مربوط به

گوسی (دقت ۹۲/۷ و صحت ۸۳/۲ درصد) و سپس لاپلاس (دقت ۷۸/۱ و صحت ۶۴ درصد) بوده است که بهترین خروجی و کمترین درصد خطا را داشته است. این امر به دلیل وجود فاصله بین کلاس‌هاست که براساس الگوریتم SVM هرچه فاصله بیشتری بین کلاس‌های طبقه‌بندی انتخاب شود، نتیجه مطلوب‌تری به دست می‌آید. به‌عبارت‌دیگر سیستم پیش‌بینی زردی نوزادی برای اینکه نتیجه بهتری در شرایط پیچیده داشته باشد به فاصله زیادی بین کلاس‌های طبقه‌بندی‌شده نیاز دارد. درحالی‌که الگوریتم SVM با هسته منحنی در شرایط مشابه نتایج مطلوبی نداشت. فیروزی جهانتیغ و همکاران [۳۳] با استفاده از SVM و سایر الگوریتم‌های طبقه‌بندی به شناسایی عوامل خطر و پیش‌بینی نوزادان کم‌وزن در زمان تولد پرداختند. در این مطالعه اطلاعات ۱۸۹ مادر و نوزاد کم‌وزن (کمتر از ۲۵۰۰ گرم) در بخش اطفال بیمارستان امام علی (ع) زاهدان از طریق مصاحبه با مادران باردار بعد از زایمان و پرونده پزشکی جمع‌آوری شد و نوزادان مرده به دنیا آمده از مطالعه حذف شد. نتایج پژوهش مذکور نشان داد که میزان دقت الگوریتم SVM (۷۳/۱ درصد) و معیار F1 (۸۳/۹ درصد) در مقایسه با یافته‌های پژوهش حاضر پائین تر بوده است. لذا، یافته‌های این مطالعه در راستای پژوهش حاضر نیست. برخی از مطالعات متمرکز بر پیش‌بینی زردی نوزادی با استفاده از سایر الگوریتم‌ها به غیر از SVM بوده که صحت یا دقتی بین ۸۹ تا ۱۰۰ درصد گزارش‌شده است. [۱،۱۷،۲۹] بنابراین، پیشنهاد می‌شود در پژوهش‌های آتی از سایر الگوریتم‌ها برای پیش‌بینی زری نوزادان استفاده شود تا بتوان نتایج آن را باهم مقایسه نمود. افزون براین، انجام یک کارآزمایی بالینی به‌منظور ارزیابی اثربخشی مدل‌های یادگیری توصیه می‌گردد. همچنین، می‌توان با استفاده از ترکیب ویژگی‌ها و استفاده از متغیرهای نشانگر زیستی (Biomarkers) مدل پیش‌بینی را بهینه‌تر کرد. کم بودن تعداد نمونه پژوهش یکی از محدودیت‌های پژوهش حاضر بود. لذا، توصیه می‌شود جهت افزایش دقت پیش‌بینی زردی نوزادان از مجموعه داده‌های وسیع‌تر استفاده شود.

ملاحظات اخلاقی

رعایت دستورالعمل‌های اخلاقی: این مقاله حاصل بخشی از پایان‌نامه با عنوان ایجاد سیستم پیش‌بینی زردی نوزادان بر اساس الگوریتم ماشین بردار پشتیبان، در مقطع کارشناسی ارشد، مصوب دانشگاه علوم پزشکی ایران در سال ۱۳۹۵ با کد اخلاق به شماره IR.IUMS.REC.1395.9311304003 اخذشده از کمیته ملی اخلاق در پژوهش‌های زیست پزشکی است.

حمایت مالی: این پژوهش با حمایت مالی دانشگاه علوم پزشکی ایران انجام‌شده است.

تضاد منافع: نویسندگان اظهار داشتند که تضاد منافعی وجود ندارد.

تشکر و قدردانی: از کلیه مشارکت‌کنندگان پژوهش که ما را یاری نمودند تشکر و قدردانی می‌گردد.

References

1. Ferreira D, Oliveira A, Freitas A. Applying data mining techniques to improve diagnosis in neonatal jaundice. BMC Med Inform Decis Mak. 2012;12:1-6.
2. Nelson WE, Kliegman R. Nelson textbook of pediatrics. 19th ed. Philadelphia: Elsevier Saunders; 2011.
3. Onyearugha CN, Chapp-Jumbo A, George IO. Neonatal jaundice: Evaluating the knowledge and practice of expectant mothers in aba, Nigeria. Journal of Health Science Research. 2016;1(2):42-7.
4. Mansor MN, Yaacob S, Muthusamy H, Nisha Basah S, Ahmad Jamil SHFS, Mohd Khidir ML, et al. PCA-based feature extraction and K-NN algorithm for early jaundice detection. International Journal of Soft Computing and Software Engineering. 2011;1(1):25-9.
5. Maisels MJ, Bhutani VK, Bogen D, Newman TB, Stark AR, Watchko JF. Hyperbilirubinemia in the newborn infant≥ 35 weeks' gestation: An update with clarifications. Pediatrics. 2009;124(4):1193-8.

6. Bhutani VK. Phototherapy to prevent severe neonatal hyperbilirubinemia in the newborn infant 35 or more weeks of gestation. Pediatrics. 2011;128(4):1-7.
7. Wong RJ, Bhutani VK. Pathogenesis and etiology of unconjugated hyperbilirubinemia in the newborn. 2014.
8. Schwartz HP, Haberman BE, Ruddy RM. Hyperbilirubinemia: Current guidelines and emerging therapies. Pediatr Emerg Care. 2011;27(9):884-9.
9. Boskabadi H, Zakeri Hamidi M, Goudarzi M. Investigating the effect of maternal risk factors in incidence of neonatal jaundice. Iranian journal of obstetrics, gynecology, and infertility. 2013;15(34):1-6. [In Persian]
10. Najib Kh, Saki F, Hemmati F, Inaloo S. Incidence, risk factors and causes of severe neonatal hyperbilirubinemia in the South of Iran (Fars province). Iran Red Crescent Med J. 2013;15(3):260-3. [In Persian]
11. Mansor MN, Yaacob S, Hariharan M, Basah SN, Ahmad Jamil SHFS, Mohd Khidir ML, et al. Jaundice in newborn monitoring using color detection method. Procedia Eng. 2012;29:1631-5.
12. Shukla M, Agarwal M. Knowledge of mothers regarding neonatal jaundice attending immunisation clinic at a tertiary care hospital of Lucknow. International Journal of Applied Research. 2016;2(6):297-9.
13. Arulmozhi A, Ezhilarasi M. Maximal information compression index (MICI) and PSO based detection of jaundice. Journal of Multiple-Valued Logic and Soft Computing. 2015;24(5-6):583-97.
14. Ramachandran A. Neonatal hyperbilirubinaemia. Paediatr Child Health. 2016;26(4):162-8.
15. Lauer BJ, Spector ND. Hyperbilirubinemia in the newborn. Pediatr Rev. 2011;32(8):341-9.
16. Calado CS, Pereira AG, Santos VN, Castro MJ, Maio JF. What brings newborns to the emergency department? A 1-year study. Pediatr Emerg Care. 2009;25(4):244-8.
17. Safdari R, Kadivar M, Tabari P, Shawky Own H. Comparison of data classification algorithms to determine the type of neonatal jaundice. Payavard Salamat. 2018;11(5):541-8. [In Persian]
18. Taylor RA, Pare JR, Venkatesh AK, Mowafi H, Melnick ER, Fleischman W, et al. Prediction of in-hospital mortality in emergency department patients with sepsis: A local big data driven, machine learning approach. Acad Emerg Med. 2016;23(3):269-78.
19. Sayadi M, Varadarajan V, Sadoughi F, Chopannejad S, Langarizadeh M. A machine learning model for detection of coronary artery disease using noninvasive clinical parameters. Life. 2022;12(11):1-10.
20. Mohammadzadeh N, mosayebi Z, Beigy H, Shojaeinia M. Prediction of sepsis due to acinetobacter infection in neonates admitted to nicu. Payavard Salamat. 2021;14(6):497-505. [In Persian]
21. Mueller M, Almeida JS, Stanislaus R, Wagner CL. Can machine learning methods predict extubation outcome in premature infants as well as clinicians? J Neonatal Biol. 2013;2:1-18.
22. Rezaee Kh, Haddadnia J, Rasegh Ghezelbash M. A novel algorithm for accurate diagnosis of hepatitis B and its severity. International Journal of Hospital Research. 2014;3(1):1-10.
23. Pal M, Foody GM. Feature selection for classification of hyperspectral data by SVM. IEEE Trans Geosci Rem Sens. 2010;48(5):2297-307.
24. Dormohammadi S, Alizadeh S, Asghari M, Shami M. Proposing a prediction model for diagnosing causes of infertility by data mining algorithms. Journal of Health Administration. 2014;17(57):46-57. [In Persian]
25. Farajollahi B. Presenting of prediction model for successful allogenic hematopoietic stem cell transplantation in adults with acute myeloid leukemia [master's thesis]. Theran: Iran University of Medical Sciences; 2022. [In Persian]
26. Hashemian AH, Manochehri S, Afshari D, Manochehri Z, Salari N, Shahsavari S. Prognosis of multiple sclerosis disease using data mining approaches random forest and support vector machine based on genetic algorithm. Tehran University Medical Journal. 2019;77(1):33-40. [In Persian]
27. Setareh S, Zahiri M, Zare Bandamiri M, Raeesi A, Abbasi R. Using data mining for survival prediction in patients with colon cancer. Iranian Journal of Epidemiology. 2018;14(1):19-29. [In Persian]
28. Cruz JA, Wishart DS. Applications of machine learning in cancer prediction and prognosis. Cancer Inform. 2006;2:59-77.
29. Castro-Ramos J, Toxqui-Quitl C, Villa Manriquez F, Orozco-Guillen E, Padilla-Vivanco A, Sanchez-Escobar J. Detecting jaundice by using digital image processing. Proceedings of SPIE. 2014;8949:1-8.

30. Adebayo Idowu P, Egejuru NC, Ademola Balogun J, Ajibola Sarumi O. Comparative analysis of prognostic model for risk classification of neonatal jaundice using machine learning algorithms. Computer Reviews Journal. 2019;3:122-46.

31. Umer A. Early prediction of neonatal jaundice using machine learning [master's thesis]. Ethiopia: Bahir Dar University; 2021.

32. Sussma S, Srivignesh S, Kishore VS, Marimuthu M. Jaundice prediction using machine learning approach. International Journal of Advance Research, Ideas and Innovations in Technology. 2022;7(6):493-5.

33. 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. Journal of Mazandaran University of Medical Sciences. 2016;25(133):171-82. [In Persian]

پژوهشگاه علوم انسانی و مطالعات فرهنگی
پرتال جامع علوم انسانی