THE JOURNAL OF COGNITIVE SYSTEMS
Journal Homepage: https://dergipark.org.tr/en/pub/jcs
Vol.7, No.2, 2022

Prediction of Placenta Accreta Spectrum by Machine Learning Methods and Determination of Candidate Biomarkers

[1]Seyma Yasar, [1]Saim Yologlu

[1]Inonu University Department of Biostatistics and Medical Informatics, Faculty of Medicine, Malatya, Turkey. (e-mail: {seyma.yasar, saim.yologlu}(1)@inonu.edu.tr).

ARTICLE INFO

Received: Sep.,09.2022
Revised: Oct.,19.2022
Accepted: Oct.,29.2022

Keywords:
Placenta Accreta Spectrum
Biomarker
Machine learning
XGBoost
Proteomics

Corresponding author: Şeyma Yasar
✉ seyma.yasar@inonu.edu.tr
☎ +90 507 639 71 21

ISSN: 2548-0650

DOI: https://doi.org/10.52876/jcs.1180200

ABSTRACT

Placenta accreta spectrum (PAS) disorders; Abnormal adhesion of placental villi to the myometrium associated with endometrial trauma or dysplasia. Placenta previa and previous cesarean section operations are two major risk factors for PAS disorders. It is usually diagnosed by ultrasound examinations performed during pregnancy follow-up. After this diagnosis is made, a very careful and strict pregnancy follow-up should be done. If the diagnosis is made during pregnancy, the delivery should be done by cesarean section and the bleeding that the mother will experience should be stopped with an appropriate method. However, no protein candidate to be used in clinical diagnosis has been found so far. The aim of this study is to identify candidate biomarkers that can be used in the diagnosis and follow-up of PAS with machine learning methods.

In this study, proteomic data obtained from 26 women with and without PAS were used. After using the Lasso method as the variable selection method, machine learning models (XGBoost, Adaboost) were created with 5-fold cross-validation. Accuracy, Balanced accuracy, Sensitivity, Specificity, Positive Predictive Value, Negative Predictive Value, F1-Score, MCC and G-mean metrics were used in the performance evaluation of the models created.

When the performance metrics of the two models are compared, the best result belongs to the XGBoost machine learning model. Therefore, the Accuracy, Balanced accuracy, Sensitivity, Specificity, Positive Predictive Value, Negative Predictive Value, F1-Score, MCC, and G-mean performance criteria for the XGBoost model are 0.962, 0.950, 1.00, 0.90, 0.94, 1.00, 0.97, 0.92, and 0.97, respectively.

As a result, considering the experimental results, it can be said that the created machine learning model is quite successful in classifying PAS. In addition, it can be said that KDR and AMH proteins are candidate biomarkers that can be used in the diagnosis and follow-up of PAS according to the significance of the variables related to the model.

1. INTRODUCTION

PLACENTA accreta is defined as abnormal trophoblast invasion of part or all of the placenta into the uterine wall myometrium. The diagnoses of placenta accreta, placenta percreta and placenta accreta, formerly known as placenta attachment anomaly, are currently called placenta accreta spectrum (PAS). It is known to cause obstetric hemorrhages, which often require blood transfusions after delivery and even threaten life. The most common known risk factor is the history of previous cesarean section and the number of previous cesarean sections. The risk increases as the number of previous cesarean sections increases. In a systemic review, when patients with a history of 1 cesarean section were compared with patients with 4 and 5, the incidence of PAS increased from 0.3 to 6.74 [1]. Other risk factors are; maternal age, multiparity, history of previous uterine surgery, history of curettage, history of Asherman syndrome and presence of placenta previa [2]. On the other hand, in clinical practice, up to 50% of pregnancies with PAS are not diagnosed before delivery, resulting in increased morbidity [3,4]. Therefore, a new and improved paradigm is urgently needed for early and accurate diagnosis. PAS antenatal diagnosis is very important. Early diagnosis of patients with PAS will reduce poor outcomes, as it will be appropriate to follow up in hospitals with a comprehensive multidisciplinary approach for problems that may occur in the follow-up of patients with PAS and bleeding that may occur during delivery.

Machine learning, as a sub-branch of artificial intelligence, is a method based on the use of computer-aided mathematical models to reveal the relationships between data and make them meaningful. Thanks to the mathematical methods used, analyzes can be made on a large number of instantly updated data stacks. It is possible to draw meaningful conclusions from the tested data using machine learning methods and to interpret these results on untested data. It allows the learning activity to be done by computers in the process of data processing. It gives computers the ability to predict future events with the results obtained from the analysis of past data [5]. Recently, machine learning methods have been used frequently in the diagnosis and treatment of diseases in the field of health.

The aim of this study is to determine the proteins that can be used in the diagnosis and treatment of PAS and that can be biomarker candidates by applying tree-based methods such as XGBoost and Adaboost, which are machine learning methods, on open access PAS data.

2. MATERIAL and METHODS

2.1. Dataset

The open-source dataset used in this study consists of 1305 proteins obtained after proteomic analysis of plasma samples taken before birth from 16 patients with PAS at 35.1 gestational week and 10 control subjects at 35.5 gestational weeks with similar gestational ages [6]. Descriptive statistics for the subjects that make up the data set are given in Table 1.

TABLE I
DESCRIPTIVE FEATURES OF PLACENTA ACCRETA SPECTRUM CASES AND CONTROL CASES

Variable	Invasive placenta (n=16)	Control (n=10)
Maternal age, year[a]	34.1 (32.4-37.2)	30.8 (30.0-36.7)
Body mass index at delivery, kg/m2[a]	33.7 (26.5-43.0)	28.5 (27.6-31.1)
Previous cesarean deliveries, n (%)		
0	1 (6.3)	4 (40.0)
1	7 (43.8)	5 (50.0)
2	2 (12.5)	0 (0.0)
3	2 (12.5)	1 (10.0)
4	4 (25.0)	0 (0.0)
Previa in current pregnancy, n (%)		
Yes	13 (81.3)	4 (40.0)
No	2 (12.5)	5 (50.0)
Unknown	1 (6.3)	1 (10.0)
Gestational age at delivery, wka[a]	35.1 (34.6-35.4)	35.5 (35.2-35.7)

[a] **Data are presented as median (interquartile range)**

2.2. XGBoost

XGBoost algorithm; It was proposed by Tianqi Chen in 2016. It is an improved and higher performance version of the Gradient Boosting Machine algorithm based on decision trees. The XGBoost algorithm works 10 times faster than other algorithms and it does this thanks to its scalability. Its scalability is due to some algorithmic optimizations. XGBoost works much faster and with higher performance thanks to its parallel operation and hardware optimizations [7].

2.3. Adaboost

AdaBoost algorithm was developed by Robert Schapire and Yoav Freund in 1995. AdaBoost is a classifier method that can obtain a strong estimator as a result of combining the singular and weak estimators obtained in each iteration. Coefficients are assigned to the estimators generated at each iteration. In these steps, the weights of the weak estimators are increased. With the completion of the iterations, the estimators to be combined are obtained [8].

2.4. Performance Evaluation of the Models

The conformity of the variables in the data set to the normal distribution was evaluated with the Kolmogorov-Smirnov test. Quantitative variables are expressed as median (Inter Quantile Range), while qualitative variables are expressed as numbers (percentage). Lasso regression method, which is frequently used in large data sets, was used as a variable selection method because it is effective and fast [9]. Then, 5-fold cross validation method was used. Accuracy, Balanced accuracy, Sensitivity, Specificity, Positive predictive value, Negative predictive value, Matthews correlation coefficient (MCC), G-mean and F1-Score metrics in the performance evaluation of XGBoost and Adaboost machine learning models created to identify candidate biomarkers that can be used in the diagnosis and follow-up of PAS used.

3. RESULTS

As a result of the variable selection method applied to 1305 proteins in the open source data set used in the study, 125 proteins were included in the study. Classification matrices for XGBoost and Adaboost models created with these 125 proteins obtained to classify PAS are given in Table 2 and Table 3, respectively.

TABLE II
CLASSIFICATION MATRIX OF THE TESTING STAGE FOR THE XGBOOST MODEL

		Real		
		Control	**PAS**	**Total**
Predicted	**Control**	9	0	9
	PAS	1	16	17
	Total	10	16	26

TABLE III
CLASSIFICATION MATRIX OF THE TESTING STAGE FOR THE ADABOOST MODEL

		Real		
		Control	**PAS**	**Total**
Predicted	**Control**	9	2	11
	PAS	1	14	15
	Total	10	16	26

The performance metrics calculated through the obtained classification matrices are given in Table 4.

TABLE IV
PERFORMANCE METRICS OF XGBOOST AND ADABOOST MODELS

METRICS	MACHINE LEARNING MODELS	
	XGBOOST	ADABOOST
	VALUE (95% CI)	VALUE (95% CI)
ACCURACY	0.96 (0.88-1)	0.89 (0.76-1.00)
BALANCED ACCURACY	0.95 (0.86-1)	0.88 (0.77-1.00)

SENSITIVITY	1.00 (0.79-1)	0.88 (0.62-0.98)
SPECIFICITY	0.90 (0.55-0.99)	0.90 (0.55-0.99)
POSITIVE PREDICTIVE VALUE	0.94 (0.71-0.99)	0.93 (0.68-0.99)
NEGATIVE PREDICTIVE VALUE	1.00 (0.66-1.00)	0.82 (0.48-0.98)
F1-SCORE	0.97 (0.90-1)	0.90 (0.79-1.00)
MCC	0.92 (0.82-1)	0.76 (0.60-0.93)
G-MEAN	0.97 (0.91-1)	0.87 (0.75-1.00)

Considering the performance metrics in Table 4, the values for the XGBoost model are higher. Therefore, the importance values of PAS-related proteins determined by this model are shown in Table 5 and Figure 1.

TABLE V
THE IMPORTANCE VALUES OF PAS-RELATED PROTEINS DETERMINED BY XGBOOST MODEL

EXPLANATORY VARIABLES	IMPORTANCE VALUE	EXPLANATORY VARIABLES	IMPORTANCE VALUE
KDR	100	PPBP	21.356
AMH	65.246	TFF2	20.813
CNDP1	44.879	MAP2K4	20.079
CXCL8	40.416	FGA FGB FGG	19.0
TNFRSF11B	38.246	PPBP	18.477
IL18R1	31.612	FTCD	18.382
THBS1	31.362	SIRT2	18.345
CA1	24.294	HP	18.287
TFPI	22.534	TPT1	18.113
VEGFA	21.526	FGFR4	17.914

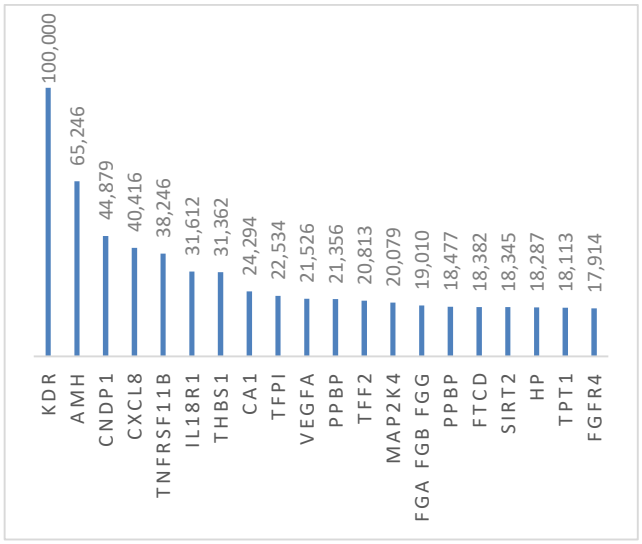

Fig. 2. The importance values for possible biomarkers

3. DISCUSSION

PAS is defined as one of the most serious disorders of pregnancy, as it is associated with a significant risk of massive obstetric hemorrhage and thus a high risk of admission to the maternal intensive care unit, reoperation, and prolonged hospitalization. On the other hand, early detection of the placental accreta spectrum (PAS), which is defined as the abnormal invasion of trophoblasts into the myometrial layer at different depths of invasion, is very important for the most appropriate surgical management and to prevent bleeding during delivery [10].

In the case of PAS, the best acceptable approach is cesarean hysterectomy [11]. However, this is not accepted by most patients. PAS can be diagnosed with high sensitivity and specificity by ultrasonography in the second and third trimesters [12, 13]. On the other hand, MRI is reported to be beneficial in cases of posterior PAS. However, MRI is expensive and requires expertise not commonly found in the diagnosis of accreta. In two studies that directly compared the two imaging modalities, MR was not found to be superior to ultrasonography [14, 15]. Therefore, there is a need for easier and cheaper diagnostic methods to be used in clinical diagnosis.

Studies in which PAS and machine learning models are integrated are very limited in the literature. In a study involving 727 women with PAS, a machine learning model was created combining baseline and perioperative variables, and the highest AUC values at which the model was evaluated were found to be 0.90 [16]. In another study, the accuracy values of different machine learning algorithms (Random Forest, K-Nearest Neighbor, Naive Bayes, Multilayer Perceptron) created using region of interest (ROI) were obtained as 95.6, 98.1,80.5 and 88.6, respectively [17].

In this study, machine learning methods (XGBoost, Adaboost) created using proteomic data from 26 subjects (PAS=16, Control=10) were used to classify PAS. According to the experimental results obtained, the accuracy values obtained from the XGBoost and Adaboost models are 0.962 and 0.885, respectively.

As a result, the proteomic dataset used in the study and the XGBoost model have a very high classification performance. Therefore, considering the variable significance obtained from the model result, it can be said that KDR, AMH proteins are candidate biomarkers that can be used in the diagnosis of PAS. In addition, as a result of the confirmation analyzes to be made, it is predicted that these two proteins can be used in the clinical diagnosis of PAS.

REFERENCES

[1] Marshall, N.E., Fu R., Guise, J.M.(2011). Impact of multiple cesarean deliveries on maternal morbidity: a systematic review. American journal of obstetrics and gynecology. 205(3):262. e1-. e8.
[2] Miller, R., Timor-Tritsch, I.E., Gyamfi-Bannerman, C., Medicine SfM-F. (2020) Society for Maternal-Fetal Medicine (SMFM) consult series# 49: cesarean scar pregnancy. American Journal of Obstetrics and Gynecology.222(5):B2-B14.
[3] Bailit, J.L., Grobman, W., Rice, M.M., Reddy, U.M., Wapner, R.J., Varner, M.W., et al. (2015) Morbidly adherent placenta treatments and outcomes. Obstetrics and gynecology.125(3):683.
[4] Fitzpatrick, K., Sellers, S., Spark, P., Kurinczuk, J., Brocklehurst, P., Knight, M. (2015) The Management and Outcomes of Placenta Accreta, Increta, and Percreta in the United Kingdom: A Population-based Descriptive Study. Obstetric Anesthesia Digest. 35(1):24-5.
[5] Greener, J.G., Kandathil, S.M., Moffat, L., Jones, D.T. (2022) A guide to machine learning for biologists. Nature Reviews Molecular Cell Biology. 23(1):40-55.
[6] Shainker, S.A., Silver, R.M., Modest, A.M., Hacker, M.R., Hecht, J.L., Salahuddin, S., et al. (2020) Placenta accreta spectrum: biomarker discovery using plasma proteomics. American Journal of Obstetrics and Gynecology. 223(3):433. e1-. e14.
[7] Chen, T., He, T., Benesty, M., Khotilovich, V., Tang, Y., Cho, H., et al. (2015) Xgboost: extreme gradient boosting. R package version 04-2. 1(4):1-4.
[8] Schapire, R.E. (2013) Explaining adaboost. Empirical inference: Springer. p. 37-52.
[9] Ranstam, J., Cook, J. (2018) LASSO regression. Journal of British Surgery. 105(10):1348.
[10] Silver, R.M., Branch, D.W. (2018) Placenta accreta spectrum. New England Journal of Medicine. 378(16):1529-36.

 ISSN: 2548-0650 http://dergipark.gov.tr/jcs

[11] Jauniaux, E., Alfirevic, Z., Bhide, A., Belfort, M., Burton, G., Collins, S., et al. (2018) Placenta Praevia and Placenta Accreta: Diagnosis and Management: Green-top Guideline No. 27a. Bjog. 126(1):e1-e48.
[12] Berkley, E.M., Abuhamad, A.Z. (2013) Prenatal diagnosis of placenta accreta: is sonography all we need? Journal of ultrasound in medicine. 32(8):1345-50.
[13] D'antonio, F., Iacovella, C., Palacios-Jaraquemada, J., Bruno, C., Manzoli, L., Bhide, A. (2014) Prenatal identification of invasive placentation using magnetic resonance imaging: systematic review and meta-analysis. Ultrasound in Obstetrics & Gynecology. 44(1):8-16.
[14] Riteau, A-S., Tassin, M., Chambon, G., Le, Vaillant. C., de Laveaucoupet, J., Quere, M-P., et al. (2014) Accuracy of ultrasonography and magnetic resonance imaging in the diagnosis of placenta accreta. PLoS One. 9(4):e94866.
[15] Balcacer, P., Pahade, J., Spektor, M., Staib, L., Copel, J.A., McCarthy, S. (2016) Magnetic resonance imaging and sonography in the diagnosis of placental invasion. Journal of Ultrasound in Medicine. 35(7):1445-56.
[16] Shazly, S.A., Hortu, I., Shih, J-C., Melekoglu, R., Fan, S., Ahmed, FuA., et al. (2021) Prediction of clinical outcomes in women with placenta accreta spectrum using machine learning models: an international multicenter study. The Journal of Maternal-Fetal & Neonatal Medicine. 1-10.
[17] Romeo, V., Ricciardi, C., Cuocolo, R., Stanzione, A., Verde, F., Sarno, L,, et al. (2019) Machine learning analysis of MRI-derived texture features to predict placenta accreta spectrum in patients with placenta previa. Magnetic resonance imaging. 64:71-6.

BIOGRAPHIES

Şeyma YAŞAR obtained her BSc. degree in mathematics from Gaziosman Paşa University in 2009. She received MSc. degree in biostatistics and medical informatics from the Inonu University in 2018. She currently continues Ph.D. degrees in biostatistics and medical informatics from the Inonu University. In 2014, she joined the Department of Biostatistics and Medical Informatics at Inonu University as a researcher assistant. Her research interests are cognitive systems, data mining, machine learning, deep learning, proteomics, bioinformatics.

Saim YOLOĞLU was born in Malatya, Turkey. He received his BS in Statistics from Hacettepe University in 1980. He completed his master's degree from Hacettepe University Health Sciences Institute Biostatistics program in 1983 and his doctorate in 1988 in the same program. In the same year, he started to work as a lecturer in the Department of Biostatistics and Medical Informatics, Faculty of Medicine, İnönü University. Currently in the same department, Prof. Dr. works as.