

**MULTI-CLASSIFICATION OF FETAL HEALTH STATUS USING EXTREME LEARNING MACHINE**

AŞIRI MAKİNE ÖĞRENMESİ KULLANARAK FETAL SAĞLIK DURUMUNUN ÇOKLU SINIFLANDIRMASI

Ömer KASIM

Department of Electrical and Electronics Engineering, Simav Technology Faculty, Kutahya Dumlupınar University, Kutahya 43500, Turkey

ORCID ID: [0000-0003-4021-5412](https://orcid.org/0000-0003-4021-5412)**ABSTRACT**

Cardiotocography (CTG) is used for monitoring the fetal heart rate signals during pregnancy. Evaluation of these signals by specialists provides information about fetal status. When a clinical decision support system is introduced with a system that can automatically classify these signals, it is more sensitive for experts to examine CTG data. In this study, CTG data were analysed with the Extreme Learning Machine (ELM) algorithm and these data were classified as normal, suspicious and pathological as well as benign and malicious. The proposed method is validated with the University of California International CTG data set. The performance of the proposed method is evaluated with accuracy,  $f_1$  score, Cohen kappa, precision, and recall metrics. As a result of the experiments, binary classification accuracy was obtained as 99.29%. There was only 1 false positive. When multi-class classification was performed, the accuracy was obtained as 98.12%. The amount of false positives was found as 2. The processing time of the training and testing of the ELM algorithm were quite minimized in terms of data processing compared to the support vector machine and multi-layer perceptron. This result proved that a high classification accuracy was obtained by analysing the CTG data both binary and multiple classification.

**Keywords:** Cardiotocography, Fetal Health Status, Extreme Machine Learning, Computer aided diagnosis.

**1. INTRODUCTION**

Monitoring of the fetal distress by specialists is important for the health of the baby during pregnancy. Cardiotocography (CTG) signals are used for this aim Ashwal et al. (2019). Fetal Heart Rate (FHR) measurements are obtained from CTG by processing the fetal heartbeat signals of babies at 28 weeks of gestation Wu et al. (2020). In these measurements, the basic heart rate fluctuation is evaluated. The fetal status is classified into 3 separate categories as normal, suspicious and pathological with Acceleration (ACC) and Deceleration (DCL) metrics Kannan et al. (2021). Imprecise monitoring of fetal pain leads to unnecessary treatments, and improper diagnosis can eliminate necessary treatments. On the other hand, it has been reported that more than 50% of deaths result from inability to recognize abnormal FHR patterns and lack of appropriate diagnosis Campos et al. (2005). Therefore, successful analysis of CTG is the most essential tool for further analysis and treatment.

In the studies conducted in the literature, the feature vector was obtained with the measurements obtained from CTG Subha et al. (2013). The presence of the pathology case was evaluated by classifying this feature vector with artificial intelligence algorithms Kannan et al. (2021). Remarkable successes were achieved with machine learning methods and shallow artificial neural networks in classification studies made according to 3 different expert decisions in the University of California International (UCI) data set Karabulut et al. (2014), Shah et al. (2015). Algorithms such as principal component analysis (PCA) and Correlation-based feature selection (CFS) were used to optimize the features in the data set Arif (2015), Georgoulas et al. (2007). When the features extracted from the data are optimized, the processing time can be shortened without decreasing the classification success. In shallow machine learning, sensitivity of 97.94 and specificity performance of 99.73 were obtained by multi-layer perceptron (MLP) method Cömert and Kocamaz (2017). The Cohen kappa value, which measures the classification success in the unbalanced data set, was obtained as 0.861 Zhang and Zhao (2017). Synthetic minority high-speed sampling technique (SMOTE) approach was used to balance the data in the data set with an unbalanced label distribution. In the classification made in the balanced data set, only 90.79 sensitivity and 91.35 specificity were obtained. Fergus et al. (2017).

Arif et al. tried to classify the fetal negativities in the UCI data set. In the experiments conducted with the decision-tree classifier, the classification error rate was obtained as 0.2107 in 408 validation samples Arif et al. (2020). Avuçlu

et al. suggested a system that met the need for early intervention by detecting the adverse situation in the fetus. In experiments with the UCI data set, the authors achieved an accuracy of 95.68% in multiple classifications with the Naive Bayes classifier Avuçlu and Abdullah (2020). In the study conducted by Nandipati et al., the results of the classification was performed by using different features together. Then, the authors discussed the accuracy of the different machine learning techniques. They obtained an average accuracy of 95.07% and the highest accuracy of 96.96% with the Random Forest classifier Nandipati et al. (2020). Jagannathan study with the UCI data set. The authors calculated the decision tree classification accuracy to 97.41% while the sensitivity was 97.93% Jagannathan (2018).

Although certain accuracy and other metrics have been achieved in studies with machine learning and shallow artificial neural networks, multi-class and binary classification is needed with highly stable algorithms such as the extreme learning machine (ELM) method. Since the ELM method enables direct decision making with a single neuron, it is suitable for achieving high accuracy multi-classification success. In this study, both binary and multiple classification accuracies were experimentally demonstrated by applying the ELM method to the UCI data set. In addition, the effective processing time obtained with ELM compared to support vector machine (SVM) and MLP methods contributed to the performance of the method.

## 2. MATERIALS AND METHODS

### 2.1. Proposed Method

In the proposed method, UCI data set which contains 21 features derived from CTG signals was used. These 21 features consist of histogram analysis, long-term variability, prolonged deceleration, acceleration numbers, beats per minute, fetal movement number and uterine contractions data. These data were organized and applied to the ELM algorithm. Problematic fetus data were detected with the ELM algorithm. Later, a multi-class classification was made, including the suspicious situations of the same data. In the first experimental setup, the normal data of 2126 data in the data set were combined with pathological and suspicious labels and classified as benign and malicious. In the second experimental setup, a classification was made into 3 classes as normal, suspicious and pathological cases. The results were validated with the accuracy, precision, recall, Cohen kappa and  $f_1$  score metrics.

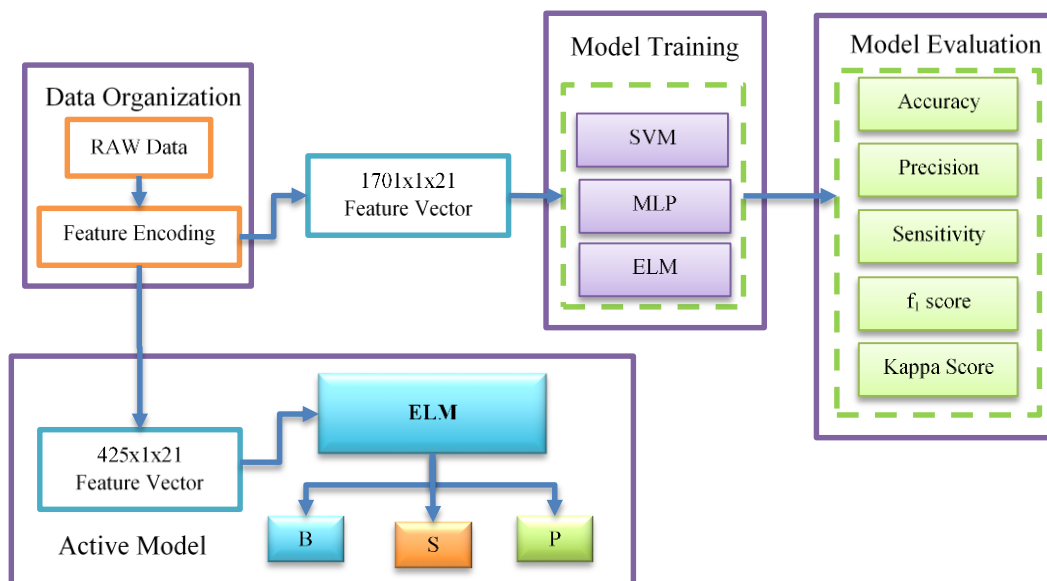


Figure 1. Flow Diagram of Proposed Method

## 2.2. Cardiocography Data Set

Cardiography is a medical device used to diagnose fetal heart rate status. This device is used by specialists for simultaneous recording and monitoring of FHR and UC patterns during pregnancy Kapaya et al. (2020). It was obtained from the open source UCI repository for the experimental analysis of the method proposed in the study UCI Machine Learning Repository (2021). The SiSporto tool was used to obtain the attributes in this data set. The data set consists of data labelled the combined characteristics of FHR and uterine contraction by specialist obstetricians and a consensus classification assigned to each. There are 2126 features with 21 features in the data set. In the features provided, 21 features can be classified according to the FHR model class or fetal status class code. In this study, fetal condition class code was used as the target feature, and each sample was classified into one of three groups as normal, suspicious or pathological. In the data set, 1655 cases were defined as normal, 295 cases as suspicious and 176 cases as pathological Campos et al. (2000).

## 2.3. Extreme Learning Machine

The ELM is a single layer feed forward network model Huang et al. (2006). The training of this model is performed with randomly selected hidden node parameters. There is no iterative optimization in this model. In training, output weights are calculated by taking the inverse of the Moore-Penrose matrix with Equation 1 Çil et al. (2020). The ELM with hidden nodes ( $L$ ) is modelled in Equation 1.

$$\sum_{m=1}^N \beta_m h_m(x_i) = y_i, i = 1, 2, \dots, L \quad (1)$$

In Equation 1,  $L$  denotes the hidden node. The sum of the products of the ELM's output neurons  $m$  connecting the  $m^{th}$  latent neurons and  $h_m(x_i)$ , which allows mapping for the output of the hidden node, is obtained. The structural formula of the  $h_m$  is expressed in Equation 2.

$$h_m(x_i) = \frac{1}{1 + e^{-(w_m^t x_i + b_m)}} \quad (2)$$

The expression in Equation 2 is a sigmoid activation function. In the equation  $w_m = [w_{m1}, \dots, w_{mD}]$  represents the input neurons and  $b_m$  represents the bias term.

Unlike traditional learning algorithms, ELM has an architecture that minimizes the training error and the smallest output weights norm simultaneously. Randomly initiated hidden node parameters are generated from the least squares solution in Equation 4 with  $w_j$  and  $b_j$ .

$$\beta = H' \cdot y \quad (3)$$

When the general form of the activation function is expressed in a general form, the calculation function of ELM is as represented in Equation 4. When the result obtained from the hidden layer with the hidden node matrix is passed through a signum function, output class as  $y$  is obtained.

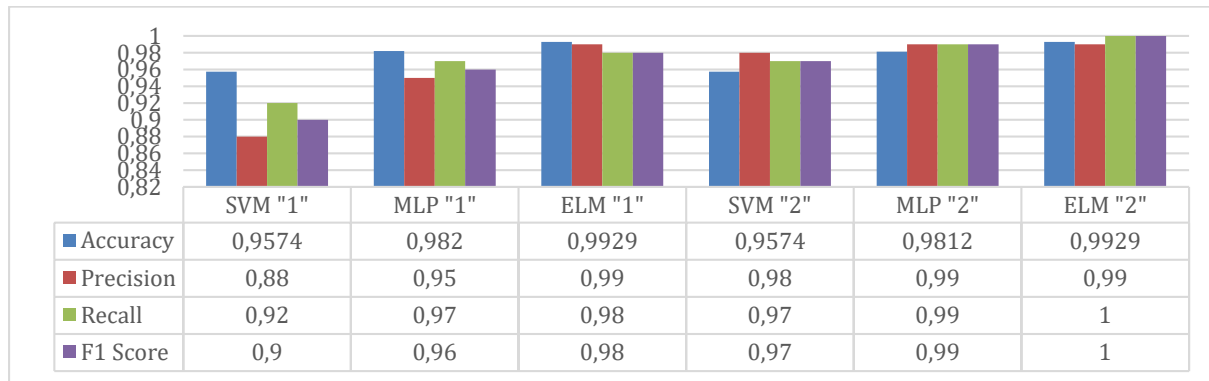
$$y' = \text{sign}(h(x) \cdot \beta) \quad (4)$$

## 3. EXPERIMENTAL RESULTS

Experiments of the proposed method were performed using the Python framework. The proposed algorithm performance metrics were achieved with 8 GB RAM and I72730 - 3GHz processor. In classification, the data lines in the UCI data set were arranged as feature vectors. These vectors were separated as 80% training and 20% validation for hold-out validation.

**Table 1.** Confusion Matrix of the Binary Classification of SVM, MLP and ELM Classifiers

	SVM		MLP		ELM	
	Malignant	Benign	Malignant	Benign	Malignant	Benign
Malignant	83	11	88	5	92	1
Benign	7	324	3	329	2	330



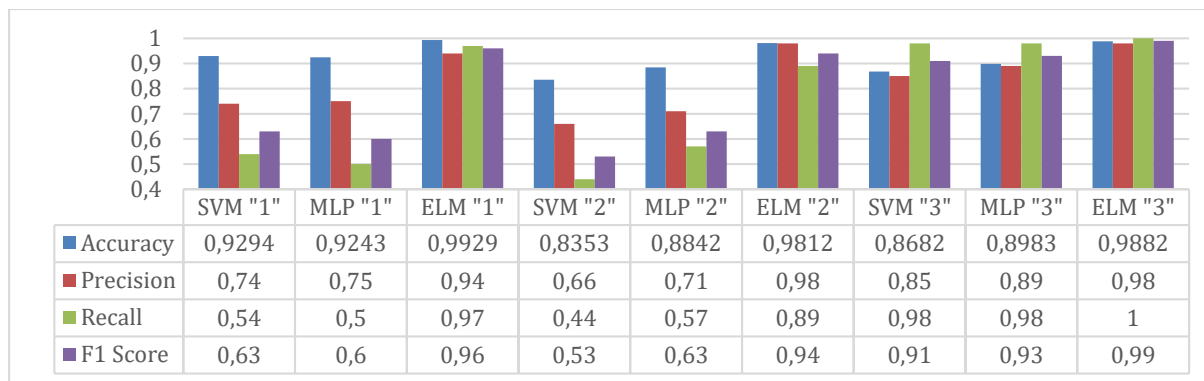
**Figure 2.** Binary Classification accuracy, precision, recall, f<sub>1</sub> score and Cohen kappa metrics for Benign (“2”) and Malicious (“1”) Labels

In the binary classification, the data labels in the UCI data set were arranged as benign and malicious. Experiments were performed with 1655 data labelled as benign and 471 data labelled as Malignant. Suspicious and pathological evaluations were merged in the malignant category. The confusion matrix in Table 1 was obtained as a result of the hold out validation made with the training of the ELM algorithm. Only one of the data labelled as benign was evaluated as malignant. In the data labelled as malignant, only 2 false positives were observed. The accuracy, precision, recall, f<sub>1</sub> score and Cohen kappa values obtained from these results were calculated as 87.4% for SVM, 94.5%, for MLP and 97.9% for ELM, respectively. The performance metrics of the binary classification is shown in Figure 2.

In multi-class classification, 3 labels were used as normal, suspicious and pathological. The holdout validation result of the ELM algorithm trained with these data is shown in Table 2. In the multi-class classification experiment, 1655 of 1226 data were normal, 295 were suspicious and 176 were pathological. When these data were classified, 2 of 39 data that were labeled as suspicious in validation were classified as normal and 3 as pathological. 1 of 59 pathologically labeled data were obtained as normal and 3 as suspicious. The accuracy, precision, recall, f<sub>1</sub> score and Cohen kappa values calculated from these results were calculated as 81.65% for SVM, 64.1% for MLP and 94.9% for ELM, respectively. The performance metrics of the multi-class classification is shown in Figure 3.

**Table 2.** Confusion Matrix of the Multi-Class Classification of SVM, MLP and ELM Classifiers

	SVM			MLP			ELM		
	Pathologi c	Suspiciou s	Norma l	Pathologi c	Suspiciou s	Norma l	Pathologi c	Suspiciou s	Norma l
Pathologi c	25	8	1	26	6	2	32	2	-
Suspiciou s	14	39	6	13	42	4	1	58	-
Normal	7	42	283	11	26	295	-	5	327



**Figure3.** Multi-class classification accuracy, precision, recall, f<sub>1</sub> score and Cohen kappa metrics for Pathologic (“1”), Suspicious (“2”) and Normal (“3”)Labels

#### 4. DISCUSSION

The data in the UCI dataset was used directly without data cleaning, data editing and data normalization. The methods proposed were validated with the hold-out validation technique and the accuracy, precision, recall, f<sub>1</sub> score and Cohen-kappa metrics were calculated. The hold-out validation provides validity in the literature, especially in the test process. The accuracy of 99.16% and only 1 FP rate obtained in binary classification show that the ELM algorithm can make the classification effectively. The accuracy of 98.12% in multiple classification shows that the classification of pathological cases and suspicious cases can be performed almost as successfully as specialists. The Cohen kappa value used in this study provides a valid measurement especially in classification problems with unbalanced data set. The kappa value of 94.9% for multi-class classification and 97.9% for binary classification obtained in this study is the proof that the experiments performed with the unbalanced data set give successful results.

In similar studies, the performance metrics obtained with the studies performed with the UCI data set were compared. Similar studies and experimental results obtained in these studies are presented in Table 3.

The studies conducted in the literature to examine the UCI data set and CTGs were compared with the method proposed in this study. When the obtained results were presented briefly, more successful results were obtained in this study in terms of evaluation metrics. Cömert et al. used an extreme learning machine, radial basis function, random forest, support vector machine and artificial neural network to classify fetal heart rate. As a result of the experiments, they recommended MLP as the most successful method. The sensitivity of 97.94 and specificity of 99.73 were obtained with the MLP algorithm [11]. Menai et al. selected 4 features out of 21 features in the UCI data set that affect the success of the classification with the Relieff algorithm. When the selected features were applied to the Naive Bayes method, they obtained an accuracy of 93.97% Menai et al. (2013). Fergus et al. divided the cases into normal and pathological using the FHR. Recursive Feature Elimination (RFE) were used for feature optimization. SMOTHE method was used to stabilize the unbalanced data set. A sensitivity of 90.79% and a specificity performance of 91.35%

**Table 3.** Additional Performance comparison with several methods

Study	Description	Accuracy Rate (%)	Performance Metrics (%)
Cömert et al.	MLP		97,94 sensitivity 99,73 specificity
Menai et al.	Relief with Naïve Bayes	93.97	
Fergus et al.	RFE and Smote		%90,79 sensitivity %91,35 specificity
Chamidah et al.	K-means and SVM (Cross Validation)	90,64	
Arif	RF Classifier	93,6	
Zhang et al.	PCA-Adaboost-SVM	98.6	
Karabulut et al.	Decision tree with adaboost	95.01	0.861 Kappa
Şahin et al.	k-NN and RF	98.4 (k-NN) 99.18 (RF)	
Georgoulas et al.	PSO-SVM	83.8	
Shah et al.	Binary classification with CFS	94.7	
Bhatnagar et al.	Naïve Bayes	82.32	
Proposed Method	ELM (Binary Classification)	99,29	
Proposed Method	ELM (Multi-Class Classification)	99,12	

were achieved in the classification made with a balanced data set Fergus et al. (2017). Chamidah et al. analysed the cardiography data in the UCI data set and obtained 7 features using the k-means algorithm. When these features were classified binary with SVM algorithm, cross-validation success was obtained as 90.64% Chamidah and Wasito (2015). Arif examined the fetal situation with 2126 data belonging to 21 feature vectors in the UCI data series. The author performed normal, suspect and pathological classification in the experiments with the RF algorithm. As a result, an accuracy of 93.6% was obtained. Zhang et al. performed feature selection by applying Principal Component Analysis (PCA) method to the CTG data set taken from the UCI data set. When the selected features were classified with the SVM classifier integrated with adaptive Boosting (AdaBoost), a success of 98.6% was achieved. The success achieved without PCA application was presented by the authors as 93%. The authors also experimentally demonstrated that with fewer features, the processing time of the algorithm was reduced Huang et al. (2006). Karabulut et al. used a decision tree-based AdaBoost to determine fetal distress, and an accuracy of 95.01% and a kappa statistic of 0.861 were calculated Karabulut et al. (2014). Şahin et al. used various machine learning algorithms to analyze CTG data with binary classification. The authors achieved the highest accuracy with k-NN (98.4%) and RF (99.18%) classifiers, respectively Sahin and Subaşı (2015). Georgoulas et al. performed classification with the k-NN algorithm using features optimized by the PSO algorithm. The authors found that the classification accuracy was more successful in identifying pathological data with SVM. The authors obtained 83.8% accuracy Georgoulas et al. (2007). Shah et al. focused on detecting normal and pathological status with binary classification in the UCI data set. The selected 7 features with high distinctiveness in feature vectors were obtained by subset evaluation CFS method. In the binary classification performed with these 7 features, a success of 94.7% was achieved Shah et al. (2015). Bhatnagar et al. used UCI data set and various classifiers (J48, JRIP, Naïve Bayes, Random forest, MLP and regression) in their studies to classify pathological cases. The authors achieved the highest success with the Naïve Bayes method with 82.32% success Bhatnagar and Maheshwari (2016).

**Table 4.** Training, and Testing Time comparison for Different Classifiers

Methods	Training Time (seconds)	Testing Time (milliseconds)
SVM Binary Classification	23,635	0,0138
MLP Binary Classification (100 iteration)	0,2102	0,0104
ELM Binary Classification	0,0520	0,0042
SVM Multi-Class Classification	42,868	0,0197
MLP Multi-Class Classification (100 iteration)	0,3683	0,0117
ELM Multi-Class Classification	0,0534	0,0053

In this study, instead of obtaining specific measurements from ECG signals, an effective classification was made based on measurements Zhong et al. (2019). The effectiveness of this classification has been proven by experiments. The ELM algorithm is an algorithm that can produce efficient classification results. Machine learning and shallow neural networks are independent of the convergence problem in the learning curve as the number of data increases. In addition, ELM does not require high processing volume like deep learning algorithms. The experiment was compared with SVM machine learning algorithm, MLP shallow neural network and ELM algorithm's processing time, accuracy and FP rate to demonstrate the effective performance of ELM. Comparative results are shown in Table 5. As can be seen in Table 4, the ELM algorithm has less training and validation times than other methods in terms of processing time.

#### 4. CONCLUSION

CTG data are used by specialists to determine the risks of fetal abnormalities. Specialists reveal normal, suspicious and pathological results with the measurement values obtained from CTG. The overlap between the ELM algorithm used in this study and the measurement metrics in the UCI data set and the decisions performed by the specialists were revealed. The proposed method validated with accuracy, precision, recall,  $f_1$  score and Cohen kappa values gave better results than SVM and MLP. When compared with the various methods suggested in the literature, it was demonstrated that quite high success metrics were obtained. In addition, it has been proven to be better than SVM and MLP in terms of processing performance since the ELM algorithm contains a transformation matrix with a single neuron. These results obtained in the study were validated in both the binary classification and the multi-class classification, which highly overlapped with the decisions performed by the specialists. This result proves that the proposed method can support specialists.

#### REFERENCES

- Arif, M. 2015. Classification of cardiocograms using random forest classifier and selection of important features from cardiocogram signal. *Biomaterials and Biomechanics in Bioengineering*, 2(3):173-183.
- Arif, M. Z., Ahmed, R., Sadia, U. H., Tultul, M. S. I., and Chakma, R. 2020. Decision Tree Method Using for Fetal State Classification from Cardiography Data. *Journal of Advanced Engineering and Computation*, 4(1): 64-73.
- Ashwal, E., Shinar, S., Aviram, A., Orbach, S., Yogev, Y., and Hirsch, L. 2019. A novel modality for intrapartum fetal heart rate monitoring. *The Journal of Maternal-Fetal & Neonatal Medicine*, 32(6): 889-895.
- Avuçlu, E., and Abdullah, E. 2020. Classification of Cardiocography Records with Naïve Bayes. *International Scientific and Vocational Studies Journal*, 3(2): 105-110.
- Ayres-de-Campos, D., Bernardes, J., Garrido, A., Marques-de-Sa, J., and Pereira-Leite, L. 2000. SisPorto 2.0: a program for automated analysis of cardiocograms. *Journal of Maternal-Fetal Medicine*, 9(5): 311-318.
- Ayres-de-Campos, D., Costa-Santos, C., and Bernardes, J., Multicentre 2005. Prediction of neonatal state by computer analysis of fetal heart rate tracings: the antepartum arm of the SisPorto® multicentre validation study. *European Journal of Obstetrics & Gynecology and Reproductive Biology*, 118(1): 52-60.

- Bhatnagar, D., and Maheshwari, P. 2016. Classification of cardiocography data with WEKA. *International Journal of Computer Science and Network-IJCSN*, 5(2):412-418.
- Chamidah, N., and Wasito, I. 2015. Fetal state classification from cardiocography based on feature extraction using hybrid K-Means and support vector machine. In 2015 international conference on advanced computer science and information systems (ICACSIS) : 37-41.
- Cömert, Z., and Kocamaz, A. F. 2017. Comparison of machine learning techniques for fetal heart rate classification. *Acta Phys. Pol. A*, 132(3): 451-454.
- Çil, B., Ayyıldız, H., and Tuncer, T. (2020). Discrimination of  $\beta$ -thalassemia and iron deficiency anemia through extreme learning machine and regularized extreme learning machine based decision support system. *Medical hypotheses*, 138: 109611.
- Fergus, P., Hussain, A., Al-Jumeily, D., Huang, D. S., and Bouguila, N. 2017. Classification of caesarean section and normal vaginal deliveries using foetal heart rate signals and advanced machine learning algorithms. *Biomedical engineering online*, 16(1): 1-26.
- Georgoulas, G., Stylios, C., Chudacek, V., Macas, M., Bernardes, J., & Lhotska, L. 2007. Classification of fetal heart rate signals based on features selected using the binary particle swarm algorithm. In *World Congress on Medical Physics and Biomedical Engineering 2006* :1156-1159. Springer, Berlin, Heidelberg.
- Huang, G. B., Zhu, Q. Y., Siew, C. K. 2006. Extreme learning machine: theory and applications. *Neurocomputing*, 70(1-3): 489-501.
- Jagannathan D. 2018. Cardiocography - a comparative study between support vector machine and decision tree algorithms. *International Journal of Trend in Research and Development*, 4: 148-151.
- Kannan, E., Ravikumar, S., Anitha, A., Kumar, S. A., and Vijayasathy, M. 2021. Analyzing uncertainty in cardiocogram data for the prediction of fetal risks based on machine learning techniques using rough set. *Journal of Ambient Intelligence and Humanized Computing*: 1-13.
- Kapaya, H., Jacques, R., Almond, T., Rosser, M. H., and Anumba, D. (2020). Is short-term-variation of fetal-heart-rate a better predictor of fetal acidaemia in labour? A feasibility study. *Plos one*, 15(8): e0236982.
- Karabulut, E. M., and Ibriki, T. 2014. Analysis of cardiocogram data for fetal distress determination by decision tree based adaptive boosting approach. *Journal of Computer and Communications*, 2(9): 32-37.
- Menai, M. E. B., Mohder, F. J., and Al-mutairi, F. 2013. Influence of feature selection on naïve Bayes classifier for recognizing patterns in cardiocograms. *Journal of Medical and Bioengineering*, 2(1):66-70.
- Nandipati, S. C. R., & XinYing, C. 2020. Classification and Feature Selection Approaches for Cardiocography by Machine Learning Techniques. *Journal of Telecommunication, Electronic and Computer Engineering (JTEC)*, 12(1): 7-14.
- Sahin H., and Subasi A. 2015. Classification of the cardiocogram data for anticipation of fetal risks using machine learning techniques, *Applied Soft Computing*, 33: 231-238.
- Shah, S. A. A., Aziz, W., Arif, M., and Nadeem, M. S. A. 2015. Decision trees based classification of cardiocograms using bagging approach. In 2015 13th international conference on frontiers of information technology (FIT):12-17.
- Subha, V., Murugan, D., Rani, J., Rajalakshmi, K., and Tirunelveli, T. 2013. Comparative analysis of classification techniques using Cardiocography dataset. *International Journal of Research in Information Technology*, 1(12): 274-280.
- UCI Machine Learning Repository: Cardiocography Data set, Available: <http://archive.ics.uci.edu/ml/datasets/Cardiocography>, (2021, March).
- Wu, X., Liu, Y., Kearfott, K., and Sun, X. 2020. Evaluation of public dose from FHR tritium release with consideration of meteorological uncertainties. *Science of the Total Environment*, 709: 136085.



Zhang, Y., and Zhao, Z. 2017. Fetal state assessment based on cardiotocography parameters using PCA and AdaBoost. In 2017 10th international congress on image and signal processing, BioMedical engineering and informatics (CISP-BMEI): 1-6.

Zhong, W., Liao, L., Guo, X., and Wang, G. 2019. Fetal electrocardiography extraction with residual convolutional encoder–decoder networks. Australasian physical & engineering sciences in medicine, 42(4): 1081-1089.