Influence of Input Data Modification of Neural Networks Applied to the Fetal Outcome Classification

MICHAL JEZEWSKI [1], ROBERT CZABANSKI [1], DAWID ROJ [2], JANUSZ WROBEL [2]
[1] Institute of Electronics, Silesian University of Technology,
Akademicka 16, 44-100 Gliwice, POLAND
[2] Department of Biomedical Signal Processing, Institute of Medical Technology and Equipment,
Roosevelta 118, 41-800 Zabrze, POLAND
Michal.Jezewski@polsl.pl

Abstract: - Cardiotocographic (CTG) fetal monitoring based on automated analysis of the fetal heart rate signal is widely used for fetal assessment. The high efficiency in diagnosis of cases with no fetal risk makes it a valuable screening method. However, the conclusion generation system is still needed to improve the fetal outcome prediction. Classification of the CTG records by means of neural networks is presented in this paper. Multi-layer perceptron neural networks were learned through 17 parameters obtained from computerized analysis of 749 traces from 103 patients, where 210 records related to abnormal fetal outcome. Classification efficiency was retrospectively verified by the real fetal outcome defined by newborn delivery data. Influence of numerical and categorical representation of the input variables, different data sets during learning, and gestational age as an additional information, were investigated in various experiments. The cases were fifty times randomly assigned to learning, validating and testing data sets. The best sensitivity and specificity were achieved for numerical input variables and with real proportion between normal and abnormal cases during learning.

Key-Words: - fetal heart rate monitoring, neural networks, pattern classification, signal analysis

1 Introduction

Cardiotocographic (CTG) monitoring is a routine procedure for assessment of the fetal state during pregnancy and labour. It relies upon noninvasive recording of Fetal Heart Rate (FHR), maternal Uterine Contractions (UC) and fetal movement activity. The CTG signals undergo analysis aimed at extraction and quantitative description of the features essential for classification of the traces as corresponding to normal or abnormal fetal state. At present, quantitative analysis of CTG records is performed with a help of computerized fetal monitoring system [1]. However, fetal assessment is still done by clinicians who finally classify the trace features as relating to normal or abnormal fetal outcome.

The benefit of fetal monitoring is that the reassuring CTG features are usually confirmed by normal fetal outcome. While the abnormal signal patterns can relate both to abnormal and normal fetal state, and false assessment of the abnormal fetal state very often causes unnecessary operative interventions. Since the cardiotocography is the primary method for fetal state assessment, looking for automated methods for efficient conclusion generation is extremely needed.

In the learning process of the neural networks, aimed at cardiotocographic traces classification, a knowledge of clinical experts is applied, which is based on evaluation of selected parameters from the newborn description [2]. The input set for automated classifier was most often formed by time and/or frequency domain parameters from computerized analysis of the FHR signal [3],[4],[5]. Usually the automated approach provided better results than human experts. However, particular emphasis was placed on the classification method itself, whereas the different aspects of input data applied during learning process should have been considered as well.

Thus, in this work we focused on the CTG signal features extraction and testing different approaches to input data preprocessing. We investigated how two possible representations of the input variables – numerical versus categorical, the different structures of learning data sets as well as additional information on a fetal gestational age affect the quality of classification. In the proposed work, the learning process was accomplished with the true fetal outcome evaluated on the newborn data by experts just after delivery. Several experiments were performed using multi-layer perceptron neural networks as the most representative. They were aimed to show the directions in which the performance measures – the prognostic indices – can be expected to change. Therefore, the obtained results do not represent their maximum values.

2 Methods

2.1 Data collection

The data were obtained from the archive of computerized fetal surveillance system MONAKO [1]. They consisted of a set of parameters of quantitative description of CTG signals in time domain together with the associated medical history referring to patients and their newborns. The FHR was acquired via pulsed Doppler ultrasound transducer placed on the maternal abdomen. Uterine contractions were recorded via strain gauge transducer. The research material consists of 2431 one-hour traces collected from 293 patients. All patients gave their informed consent prior to participation in the study, which was approved by the ethical committee.

In the first step of data cleaning, the traces with the signal loss higher than 20% were excluded. The variability indices calculated for one-minute intervals were assumed as reliable only if at least 60% of samples were valid. We also excluded all the records, where the percentage of missing one-minute variability index values exceeded 20%. We rejected records from patients with incomplete delivery and newborn data forms. Finally, we obtained 749 records from 103 patients, where 210 (28%) records related to abnormal fetal outcome. The number of traces recorded from particular patient varied from one to ten, and the distribution of traces among patients according to gestational age is shown in Fig. 1.

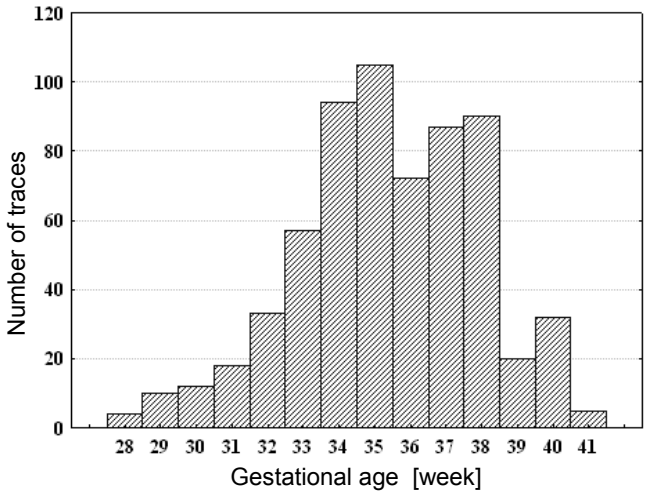

Fig. 1 Number of all antenatal traces recorded in particular weeks of pregnancy.

2.2 Features extraction

Fetal heart activity is described by changes of cardiac intervals T_i determined between two consecutive heart beats or using the instantaneous fetal heart rate values FHR_i, which is an extrapolation of the interval T_i into one-minute period:

$$FHR_i = \frac{60000}{T_i\,[miliseconds]} \quad [beats\ per\ \ minute] \qquad (1)$$

Significant FHR variability in time (Fig. 2) is caused by complex heartbeat regulation system. There are a number of different variability patterns which can be grouped into:

a) changes of the basal fetal heart rate called baseline that comprise very slow and usually long-lasting decrease or increase of the heart rate;
b) changes of the fetal heart rate in certain direction, e.g. transitory increases above the baseline defined as accelerations, as well as transient slowing of the FHR in relation to the baseline called decelerations;
c) short-lasting changes of the FHR also called instantaneous variability. There are two types of this variability: short-term variability with changes of consecutive T_i intervals duration (called beat-to-beat variability), and long-term variability with periodical changes of beat-to-beat variability concerning both direction and magnitude (called oscillations of FHR).

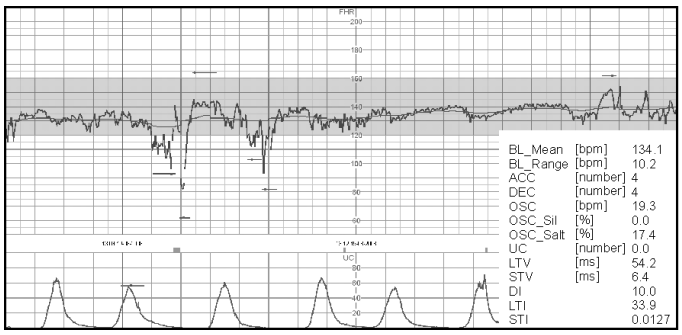

Fig. 2 Segment of cardiotocographic trace. The horizontal bars above/below the waveforms identify the recognized trace acceleration/deceleration patterns. Additionally, window presents 13 CTG parameters: ACC, DEC and UC represent number of patterns recognized in segment, and the OSC values – percentage of time in the segment duration.

Two parameters directly describing the FHR baseline have been included into an input data set for neural network: the mean baseline value (*BL-Mean*) and the fluctuation range of baseline values calculated as a difference between maximum and minimum values (*BL-Range*). Numbers of A/D patterns detected per hour were represented by *ACC* and *DEC* variables. All the variability indices (short term: *STV* and *DI*, long term: *LTV* and *LTI*, oscillation amplitude – OSC) are calculated for separated one-minute signal windows with samples averaged over 2.5 s. Exceptions are the indices marked with *BB*, which are calculated using the signal in a form of certain number of events T_i within one-minute fragment. With exception of percentage of silent (*OSC-Sil*) and saltatory (*OSC-Salt*) oscillations, the final value of a given index relates to the whole signal record and it

is calculated as a mean of all one-minute values. The number of contractions (*UC*), and the number of fetal movements (per hour) perceived by mother during the entire monitoring session were included in the neural network inputs set as *MOV* variable.

2.3 Fetal outcome

When using neural networks for classification of CTG signals, the learning process is accomplished with the known results of fetal outcome evaluated by experts just after delivery with a help of three main attributes of the newborn. The Apgar score is a simple method for evaluation of newborn's physical condition just after the childbirth using five factors: appearance, pulse, grimace, activity and respiration. The Apgar score ranges from 0 to 10, and the value below 7 is regarded as abnormal. Percentile of birth weight is determined basing on neonatal birth weight in relation to its reference percentiles stratified by infant sex, and the gestational age (from 28th to 44th completed week of gestation) determined from USG. Birth weight below the 10th reference percentile is regarded as abnormal. At the time of birth the umbilical cord blood sampling for gas values analysis (especially for pH measurement) is considered as very important for fetal oxygenation status. The value of pH below 7.20 means an abnormal fetal state.

The neural network output (classification result) represents the predicted fetal outcome, but in practice, it means the fetal state at the time of CTG monitoring. In our application the two-state output represents normal or abnormal fetal outcome. Common approach in clinical practice is to assume the fetal outcome as abnormal, if at least one attribute is outside the physiological range. In our research material the neonatal birth weight was the most decisive attribute as it classified 131 fetal outcomes as abnormal, whereas Apgar score 45 and pH only 17.

2.4 Neural network modelling

The set of 17 parameters of quantitative description of the CTG signals as the input data were normalized according to their minimum and maximum values. Fifteen parameters describe the FHR features in time domain: baseline (two indices), number of recognized A/D patterns (two), the short-term variability (six), as well as the long-term variability parameters (five indices). Additionally, the number of identified uterine contractions and fetal movements were involved.

We used MLP neural networks with the sigmoid activation function. The number of neurons in a hidden layer were changed in two ranges: from 2 to 10 with one-neuron step and from 5 to 250 with five-neuron step. Additionally, different learning time in epochs of 200, 500, 1000, 5000, 10000 as well as the learning rate with values of 0.001, 0.01, 0.1, 0.15, 0.5 were applied. The output was a single neuron and the threshold between the two classes analyzed was automatically determined, while minimizing the classification error. The steepest descent gradient algorithm was used. To avoid the situation when the network with a given structure provides very good results only by chance, we applied a set of trials with randomly arranged contents of data sets. In every experiment, the cases were 50 times randomly assigned to three data sets: learning, validating and testing. As the result we obtained 50 neural networks with a given constant structure, but with different performance parameters resulted from the learning process. The normal and abnormal cases were partitioned in the 50%, 25% and 25% respectively. Thus, the ratio of cases with normal fetal outcome to abnormal one in each set was constant in all trials. The exception was the Equal-Approach learning which will be described later. Results of particular experiments are presented as mean values (with standard deviations) calculated for all trials.

2.5 Experiments

Pregnant woman can be monitored many times, especially in case of high-risk pregnancy, which causes that a given fetal outcome is related to several CTG traces. For the classification procedure it is possible to select one trace for each patient, for example the one registered as close as possible to the delivery. However, in [6] we stated that remaining the research material unchanged gives better results, and thus this approach was used in the current study.

Seventeen parameters representing fetal records, were fed to neural network inputs in numerical and categorical form. Information on gestational age was used additionally either as the number of completed week of pregnancy or antenatal group number. In highly-developed countries the fetal outcome is normal in most of the cases. Our relatively high number of abnormal cases – 210 (28%) in relation to the normal ones – 539 (72%), is caused by the fact that the research material was obtained from clinical centre which represents in Poland the highest level of perinatal care. Therefore, during the learning stage the input data were grouped in two ways: Real Approach - where the original sizes of classes were maintained, and Equal Approach – with classes adjusted to the same size by removing some randomly selected records referring to the normal outcome. Finally, the reduced learning data set comprised the same number (105) of normal and abnormal cases. Classification efficiency was determined in relation to fetal outcome for a whole research material and separately for the particular antenatal groups.

Eleven parameters of quantitative description of CTG signals could be converted from their original numerical values into categorical ones basing on the established ranges of physiology [7]. After conversion the value of 0 means that the corresponding numerical value is within the normal range, whereas 1 means the abnormal value. The range is a function of gestational age at which the CTG signals were recorded.

Table 1. Descriptive statistics of eleven input parameters with double representation – Numerical and Categorical

Input parameter		Number/ percentage [%] of traces in abnormal range	Mean ± SD
ACC	[number]	205/27.4	8.49 ± 5.77
LTI	[-]	167/22.3	24.08 ± 7.79
STI	[-]	138/18.4	1.14 ± 0.33
DI	[-]	188/25.1	9.34 ± 2.77
UC	[number]	65/8.7	2.91 ± 4.97
LTV	[ms]	117/15.6	42.45 ±11.04
MOV	[number]	423/56.5	33.38 ±47.54
DEC	[number]	30/4.0	1.56 ± 2.87
OSC-Sil	[%]	206/27.5	6.01 ± 8.57
STV	[ms]	137/18.3	6.03 ± 1.86
OSC-Salt	[%]	15/2.0	11.44 ± 0.81

Descriptive statistics of the input parameters and their abnormal and normal values within the material collected are listed in Table 1. For both data set representations the neural networks were proposed. The distribution of the values of De Haan long-term variability index *LTI* is presented in details in Fig. 3.

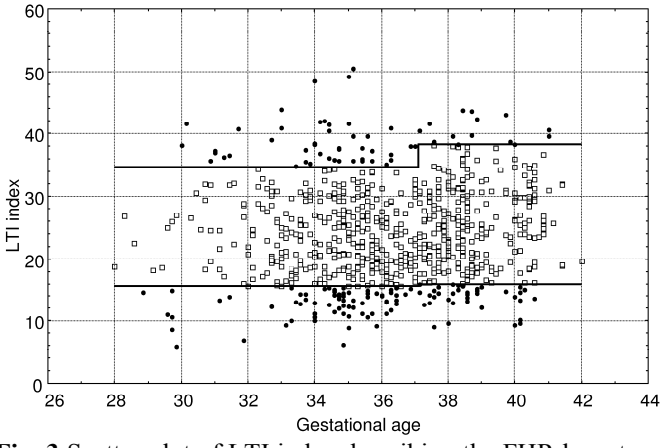

Fig. 3 Scatter plot of LTI index describing the FHR long-term variability calculated for CTG traces recorded between 26th and 44th week of pregnancy. Normal values represented by white squares.

Real influence of the input parameters on the classification quality was investigated by the importance index calculated for each particular input variable within all 50 learning trials. The importance index was defined as the ratio of the error of classification by the neural network learned without the given input variable, to the error of the network trained with all input variables. The classification error represents the number of false classified cases (fetal outcomes) related to all cases.

With the progress of pregnancy, the features characterizing the CTG signals change. The current gestational age (for a given date of CTG recording) is determined in relation to the gestational age obtained through the ultrasound examination that had been performed before the 20th week of gestation. That enabled two successive experiments to be performed in order to investigate how a different gestational age at which the signals were recorded, influences the classification process. In the first experiment this information was directly applied as an input in two ways: as the number of completed week of pregnancy, and as the number of one of previously established four groups of antenatal traces. For a given gestational age expressed with accuracy of day its distances as an absolute differences to centers of the four groups were determined: G1 – 34.5 , G2 – 35.5, G3 – 36.5, G4 – 38.5. A given CTG record was assigned to the group with minimum distance. If the absolute difference was equal for two groups, this one of the higher centre was chosen. After conversion we obtained 228 traces with gestational age assigned to the G1, 105 to G2, 123 to G3 and 183 to G4 group. In the second experiment, the research material was partitioned by the distribution of antenatal CTG traces recorded between 28th and 41st week of pregnancy. In order to ensure comparable number of records, four overlapping sets S1 ÷ S4 were proposed (Table 3). Additionally, 286 traces recorded at labour were extracted from the material and assigned to set SL. Classification was performed using neural networks for each group separately. For every experiment by the trial and error method the best network structure was selected and basic parameters of learning algorithm were set.

Classifying the CTG signals as corresponding to abnormal or normal fetal state is a kind of diagnostic test giving positive or negative result respectively. Relating it to a true result – the fetal outcome, enables performance measurement using sensitivity (SE), specificity (SP), positive (PPV) and negative (NPV) predictive values. Additionally, all these indices were combined to form the overall index OI:

$$OI = \sqrt{\frac{(2 \cdot SE + NPV)}{3} \cdot \frac{(SP + PPV)}{2}} \quad [\%] \qquad (2)$$

From the clinical point of view, it is crucial to minimize the number of false negative cases, because they have more serious consequences than the false positive ones. Thus, the sensitivity weight is doubled in OI formula.

3 Results

An increase of number of epochs above 500 did not affect significantly the prediction quality. As for the learning rate the best results were obtained for 0.01 and 0.15. An increase of number of hidden neurons led to decrease of the generalization ability, and the above tendencies were noted for all experiments. Testing the influence of data set representation, we noticed the decreasing of all mean values of prognostic indices for categorical inputs (together with increasing of their SD values). In turn, changing the type of data learning from Real-Approach to Equal-Approach caused the increase of *SP* and *NPV* indices, whereas the *SE* and *PPV* decreased. This tendency resulted in decreasing of *OI*, but this change was not statistically significant. The results of this part of study caused that in the next experiments the networks were designed with the numerical input data representation and the Real-Approach mode of learning.

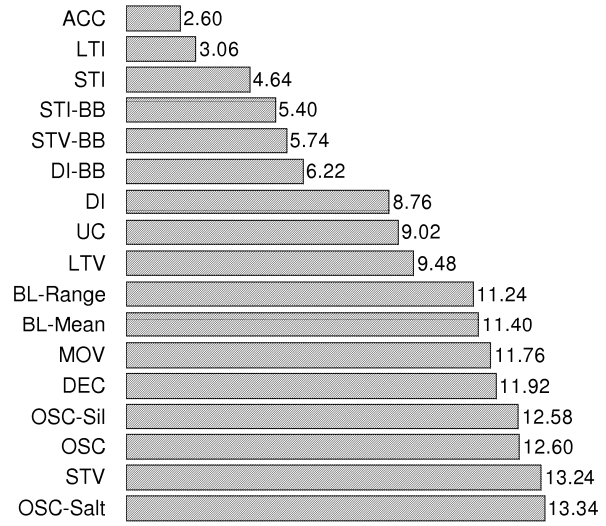

Fig. 4 The mean value of the numerical inputs positions taken in all fifty trials with a use of MPL neural network.

Additionally, the importance index was calculated for all inputs, and then the inputs were ranked according to the index value. The mean value of ranks (1 to 17) for a given input in all trials, which represents its real influence on classification quality, is shown in Fig. 4. It is easy to see, that the most significant parameter is the number of accelerations as well as the indices describing the FHR variability. This confirms that they are regarded as crucial signs of fetal wellbeing.

Analysing the influence of the gestational age, as an additional neural networks input (Table 2), it was surprising that only for neural network fed with the information on gestational age as a number of the appropriate antenatal group, the quality of classification was higher. The summary statistics concerning the classification quality obtained in the next experiment, when neural networks were separately learned with four sets of gestational traces as well as the set of labour traces, is presented in Table 3. In general, in sets S1, S2 and S3 we obtained higher values for most prognostic indices in comparison to the previous experiment, when gestational age was applied as additional input. Taking into account statistical significance of the differences among OI values, the best results were obtained for the neural network designed for the set of signals recorded in the earliest period of pregnancy, i.e. between 33rd and 36th week. With the increase of gestational age the decrease of prognostic indices was observed. Significantly better results were obtained for labour set in relation to the fourth antenatal one.

Table 2. Classification results obtained for two NNs designed with two representations of gestational age applied as additional input parameter.

Gestational age representation	Completed week of pregnancy	Number of the antenatal group
Prognostic index	Mean ± SD [%]	Mean ± SD [%]
SE	62.6 ± 8.0	65.7 ± 8.5
SP	64.4 ± 5.5	68.5 ± 6.2
PPV	42.2 ± 7.5	45.6 ± 7.9
NPV	80.6 ± 4.6	83.3 ± 4.6
OI	60.5 ± 11.1	63.9 ± 10.2

The best classification quality was achieved for the MLP neural network with six hidden neurons fed with numerical parameters and with additional input being the gestational age information as a number of the gestational group, and with original proportion between normal and abnormal fetal outcomes. The obtained sensitivity of 65.7% and specificity of 68.5% can be in some way related to those obtained in [8] – sensitivity of 73% and specificity of 94%. It must be pointed out, that our results are more rigorous because they concern mean values obtained after 50 trials with randomly mixed learning and testing subsets. Additionally, in [8] as much as 30 input variables were used, and the abnormal fetal outcome was defined by more strict criteria applied to the newborn description. Improvement of the classification quality was noted when the whole data set was divided into the sets according to the gestational age and neural networks were designed separately for each set. The highest sensitivity of 71.3% and specificity of 72.5% was obtained for the set of signals recorded in the earliest period of pregnancy, i.e. between 33rd and 36th week. However, the classification quality indices decreased with the increase of gestational age.

Table 3. Characteristics of the sets with gestational and intrapartum traces and summary statistics of associated prognostic indices.

Traces sets	SL	S1	S2	S3	S4
Gestational age [weeks]	Labour	<33÷36>	<34÷37>	<35÷38>	<36÷41>
Material characteristics	286 / 43%	285 / 32% [a]	284 / 31%	279 / 30%	278 / 29%
Prognostic index (Mean ± SD [%])					
SE	66.5 ± 8.4	71.3 ± 8.8	66.1 ± 10.3	67.8 ± 10.2	61.8 ± 10.1
SP	68.0 ± 9.7	72.5 ± 7.8	70.6 ± 7.4	67.0 ± 6.2	66.8 ± 8.2
PPV	62.0 ± 11.0	55.9 ± 10.3	51.4 ± 10.0	46.6 ± 7.7	43.1 ± 10.4
NPV	72.2 ± 7.5	83.8 ± 5.9	81.7 ± 6.1	83.1 ± 5.5	81.6 ± 4.7
OI	66.6 ± 13.2 [b]	**69.6 ± 13.4** [c]	66.0 ± 13.0	64.3 ± 12.6	61.3 ± 12.9

[a] number of records / percentage with abnormal fetal outcome.
[b] statistically significant difference SL vs. S4 ($p < 0.01$).
[c] statistically significant difference S1 vs. S2, S3, S4, SL ($p < 0.02$).

4 Conclusions

Various structures of learning subsets were tested to consider that during pregnancy the fetus is usually monitored several times which leads to assigning a number of CTG traces to one fetal outcome. Representation of the input variables in categorical form caused the decrease of all performance indices (with an increase of their standard deviations), so the numerical representation should be preserved. Among the numerical input parameters the most significant were the number of accelerations and the indices describing the instantaneous variability of FHR. Improvement of the classification quality was noted when whole data set was divided into sets according to the gestational age, and neural networks were designed separately for each set. However, the classification quality indices decreased with the increase of gestational age. We found these results encouraging and our plan for future research is to collect larger database in order to select more representative groups with gestational as well as intrapartum traces.

Acknowledgements
This work was supported in part by the Ministry of Sciences and Higher Education resources in 2007-2010 under Research Project N518 014 32/0980.

References:

[1] Jezewski J, Wrobel J, Horoba K, et al., Centralised fetal monitoring system with hardware-based data flow control, *Proceedings of III International Conference MEDSIP,* Glasgow, 2006, pp. 51-54.

[2] Beksac MS, Ozdemir K, Erkmen A, Karakas U, Assessment of antepartum fetal heart rate tracings using neural networks, *A critical appraisal of fetal surveillance*. Edited by van Geijn HP, Copray FJA. Elsevier Science BV, 1994, pp. 354-362.

[3] Georgoulas G, Stylios ChD, Groumpos P, Predicting the risk of metabolic acidosis for newborns based on fetal heart rate signal classification using support vector machines, *IEEE Trans Biomed Eng*, Vol. 53, 2006, pp. 875-884.

[4] Arduini D, Giannini F, Magnes G, et al., Fuzzy logic in the management of new perinatal variables, *Proceedings of the 5th World Congress of Perinatal Medicine,* 23-27 August 2001, Barcelona, Edited by Carrera JM, Monduzzi E, Cabero L, pp. 1211-1216.

[5] Salamalekis E, Thomopoulos P, Giannaris D, et al., Computerised intrapartum diagnosis of fetal hypoxia based on fetal heart rate monitoring and fetal pulse oximetry recordings utilising wavelet analysis and neural networks, *Int J Obstet Gynaecol,* Vol. 109, 2002, pp. 1137-1142.

[6] Jezewski M, Henzel N, Wrobel J, et al., Application of neural networks for prediction of fetal outcome, *J Medical Informatics and Technologies,* Vol. 10, 2006, pp. 127-132.

[7] Sikora J, Digital analysis of cardiotocographic traces for clinical fetal outcome prediction, *Klin Perinat Ginekol*, Vol. 10 (Suppl 21), 2001, pp. 57-88.

[8] Arduini D, Giannini F, Magenes G, Intrapartum surveillance: computer cardiotocography, *Proceedings of the 5th World Congress of Perinatal Medicine*, 2001, pp. 1217-1223.