Biomaterials and Biomechanics in Bioengineering, *Vol. 2, No. 3 (2015) 173-183*
(Formerly, Biomaterials and Biomedical Engineering)
DOI: http://dx.doi.org/10.12989/bme.2015.2.3.173

Classification of cardiotocograms using random forest classifier and selection of important features from cardiotocogram signal

Muhammad Arif*

Department of Computer Science, College of Computer and Information systems, Umm-Alqura University, KSA

(Received March 19, 2015, Revised November 3, 2015, Accepted November 6, 2015)

Abstract. In obstetrics, cardiotocography is a procedure to record the fetal heartbeat and the uterine contractions usually during the last trimester of pregnancy. It helps to monitor patterns associated with the fetal activity and to detect the pathologies. In this paper, random forest classifier is used to classify normal, suspicious and pathological patterns based on the features extracted from the cardiotocograms. The results showed that random forest classifier can detect these classes successfully with overall classification accuracy of 93.6%. Moreover, important features are identified to reduce the feature space. It is found that using seven important features, similar classification accuracy can be achieved by random forest classifier (93.3%).

Keywords: cardiotocography; fetal heart rate; random forest classifier; uterine contractions; biomedical data classification

1. Introduction

In Cardiotocography (CTG), fetal heart rate (FHR) and uterine contraction data are recorded simultaneously during pregnancy. Obstetricians use Cardiotocography as an important tool to evaluate the well-being of the fetus before delivery. This is typically done in the third trimester of the pregnancy. A standardized nomenclature has been adopted to read the cardiotocographs (Macones *et al*. 2008). It includes description of uterine activity, baseline fetal heart rate (110 to 160 beats per minute), baseline FHR variability (5 to 25 beats per minute above and below the stable FHR baseline), periods of reduced and increased FHR variability and presence of any acceleration or deceleration (Ugwumadu 2013). By monitoring FHR, it is possible to identify the fetal hypoxia (shortage of oxygen typically in the range of 1 to 5%). If fetal hypoxia is prolonged then chances of disability of the newborn baby becomes high and sometime it may leads to the death. Hence it is very important to identify abnormal FHR patterns and take appropriate actions to avoid prenatal morbidity and mortality (Chen *et al*. 2011, Lees *et al*. 2013). Cardiotocography can be used to investigate the status of fetus health, normoxia (oxygen tensions between 10-21%) (Carbonne *et al*. 1997) and normal or abnormal fetus acid base status (Spencer 1993). Hence many indicators, occurring days or hours before fetus death, if detected promptly can lead to proper

*Corresponding author, Professor, E-mail: mahamid@uqu.edu.sa, syedmarif2003@yahoo.com

obstetric intervention which could help in delivering a healthy baby. CTG monitoring is done manually which may leads to human error. A computerized CTG may produce automatic interpretation of CTG reducing the prenatal mortality rate (PMR) (Grivell *et al*. 2010, Brown *et al*. 2014).

Automated continuous FHR monitoring is an important tool for intrapartum surveillance. Variability of FHR is characterized by measuring the variance or standard deviation of FHR about the baseline values(Ayres-de-Campos *et al*. 2008). Recently, many researchers have focused on other properties related to variability of the fetal rate signal. It includes spectral analysis (Logier *et al*. 2008, Kwon *et al*. 2012), nonlinear analysis (Gonçalves *et al*. 2006), matching pursuit technique to detect fetal hypoxia (Salamalekis *et al*. 2006), entropy measures as an indicator of complexity loss of FHR in acidemia (Costa *et al*. 2014).

Different computational intelligence methods are used to classify the CTG data. Czabanski *et al*. (Czabanski *et al*. 2012) used two steps mechanism comprising of weighted fuzzy scoring systems and LSVM algorithm and applied this mechanism to FHR to predict the acidemia risk. Georgieva *et al*. (Georgieva *et al*. 2013) and Jezewski *et al*. (Jezewski *et al*. 2007) used artificial neural network to monitor the fetal wellbeing. Other approaches include neuro-fuzzy system (Czabanski *et al*. 2008), naïve bayes classifier (Menai *et al*. 2013) and support vector machines (Ocak 2013). Beatrijs *et al*. (van der Hout-van der Jagt *et al*. 2012) proposed mathematical modeling approach to simulate early deceleration in CTG. Ensemble classification algorithms consist of weak learners when combined together produce higher and robust classification accuracies (Dietterich 2000). Bagging (Dietterich 2000, Galar *et al*. 2012) and boosting (Svetnik *et al*. 2005, Galar *et al*. 2012) are two famous techniques used in the ensemble classifiers to combine the classification outputs of the weak learners. Esra *et al*. (Karabulut and Ibrikci 2014) used adaptive boosting ensemble of decision trees to analyze cardiotocogram for identifying pathologic fetus. Random forest (Breiman 2001) is an ensemble classifier that is built on multiple trees from randomly sampled subspaces of the input features and combine the output of the trees using bagging. Random forest classifier is used in numerous real life applications like protein sequencing (Kandaswamy *et al*. 2011), classification of Alzheimer's disease (Gray *et al*. 2013), cancer detection (Ozcift 2012), physical activity classification (Arif *et al*. 2014) and so on.

In this paper, random forest classifier is used to classify the cardiotocograms into normal, suspicious and pathological classes in the cardiotocogram database provided as public dataset (Ayres-de-Campos, Bernardes et al. 2000). Feature importance index is used to identify important features of the database. It is shown that good classification accuracy can be achieved by using only seven important features out of total twenty one features.

2. Materials and methods

Cardiotocography (CTG) is a method of recording the fetal heartbeat and the uterine contractions during pregnancy usually in the last trimester. The data set consists of 2126 cardiotocograms collected in the Maternity and Gynecological Clinic (University Hospital of Porto in Portugal) (Ayres-de-Campos, Bernardes *et al*. 2000).

Cardiotocograms (CTG) are classified by three expert obstetricians and their majority opinion has defined the class of the cardiotocogram. Cardiotocograms are also processed by SisPorto (Ayres-de-Campos *et al*. 2000) and 21 features are extracted automatically. List of features is tabulated in Table 1. The detailed description of the features is described in (Ayres-de- Campos *et*

Table 1 Explanation of features

Feature	Explanation	Feature	Explanation
LB	FHR Baseline value	**mLTV**	Mean value of long term variability
AC	Accelerations in FHR	**DL, DS, DP**	Light, Severe, Prolonged Decelerations
FM	Fetal movement	**Width**	Width of the Histogram
UC	Uterine contractions	**Min, Max**	Low frequency and High frequency of the Histogram
ASTV	Percentage of time with abnormal short term variability	**Nmax, Nzeros**	number of histogram Peaks and Zeros
mSTV	Mean value of short term variability	**Mode, Mean, Median, Variance**	Mode, Mean, Median and Variance of the Histogram
ALTV	Percentage of time with abnormal long term variability	**Tendency**	histogram tendency: -1=left asymmetric; 0=symmetric; 1=right asymmetric

Table 2 Class distribution of CTGs

Class (Fetal State)	Number of FHR Recordings
N (Normal)	1655
S (Suspect)	295
P (Pathologic)	176
Total	2126

al. 2000). Baseline value (LB) is estimated by an algorithm using fetal heart rate (FHR) and short term variability (STV) mentioned in (Ayres-de-Campos *et al*. 2000). Acceleration is defined as increase in the FHR above the baseline for 15 to 120 seconds and deceleration is defined as the decrease in the FHR below the baseline for 15 to 120 seconds. Uterine contraction signal after filtering is evaluated for the contraction episodes. Variability analysis is also done by the Sisporto software and features are extracted as listed in Table 1. Histogram of FHR is calculated and features as explained in the Table 1 are extracted. Distribution of classes of CTG database is listed in Table 2.

Random forest classifier (Breiman 2001) is used for the classification of three classes. Let X is the data set comprising of N data points and $\boldsymbol{d}$ features containing M classes, C_i, i=1,2,…,M. An independent subset of data set Φ_k selected randomly from the dataset X, such that $\Phi_k \subseteq X$ (bootstrap sample), containing **s** features set $\boldsymbol{s} \subseteq \boldsymbol{d}$. It is used to train a tree $h(\boldsymbol{x}, \Phi_k)$ as weak classifier for the training set where $\boldsymbol{x}$ is the input. In the random forest classifier, Combination of various such trees (as weak classifiers) is used to predict the class of a particular feature vector by majority voting. Tibshirani (Tibshirani 1996) and Wolpert *et al*. (Wolpert and Macready 1999) proposed out-of-bag estimates to estimate the generalization error and it is found to be as good as the test data of the same size as of the training set (Breiman 2001). Out-of-bag error is calculated by putting aside a fraction of Φ_k as out-of-bag. It is done for all k trees. At the end, class of the vector in the out-of-bag is predicted as the class who got most of the votes in the prediction by the trees. This class is compared with the true class and used in the estimate of out-of-bag error

estimate.

Classification results are presented by using precision, recall and F-measure. Precision or positive predictive value (PPV) is defined as the proportion of instances which belongs to a class (TP: True Positive) out of the total instances including TP and FP (False Positive) classified by the classifier as belong to this particular class

$$Precision = \frac{TP}{TP+FP} \tag{1}$$

Recall or Sensitivity is defined as proportion of instances classified in one class out of the total instances belonging to that class. Total number of instances of a class includes TP and FN (False Negative)

$$Recall = \frac{TP}{TP+FN} \tag{2}$$

F-measure is the combination of precision and recall and defined as

$$F - measure = \frac{2 \times Precision \times Recall}{Precision + Recall} \tag{3}$$

3. Results and discussions

As discussed in the above section, there are 2126 number of instances which are labeled as one the three classes, Normal (N), Suspicious (S) and Pathological (P) as mentioned in Table 2. Random forest classifier is applied to classify these three classes. First step is to optimize the structure of the random forest classifier. So it is important to decide the optimal leaf size and the number of trees generated in the random forest classifier. Number of features included in the subset Φ_k randomly is selected to be one third of the input features (total features are 21). Out-of-bag classification error is calculated for different leaf sizes (2, 5, 10 and 15) and plotted in Fig. 1. For every leaf size, classifier is trained 10 times and the mean value of the out-of-bag classification error is plotted in Fig. 1 for different tree sizes. It can be seen from the figure that out-of-bag classification error decreased as the number of grown trees are increased and becomes almost constant after the grown trees size was greater than fifty. It can also be observed that out-of-bag classification error for leaf size equals to two remains minimum for all sizes of the grown trees.

Hence for the further classification experiments, leaf size is fixed to be 2 and the number of trees is fixed to 50.

Training and testing data sets are generated by dividing the whole data set into 70-30 split randomly without replacement. Random forest classifier is trained on the training set and class labels of the testing set is predicted by the trained classifier. The whole process is repeated 10 times and mean and standard deviation of Precision, Recall and F-measure is reported for training and testing data in Table 3 and Table 4. For the training data, random forest classifier showed very good performance with high values of precision, recall and F-measure. The last row of both tables show weighted average of the values. Weights are assigned as the fraction of number of data points of each class. For the testing data sets, Precision and recall of the Normal class (comprising of approximately 78% of the testing set) are more than 0.94 with F-measure of 0.96. Suspect class (S) showed low values of precision and recall as compared to other two classes. It is obvious because experts also put these cardiotocogram in the suspect class. So it is easier for this class to be confused by the classifier with either normal (N) class or pathological (P) class.

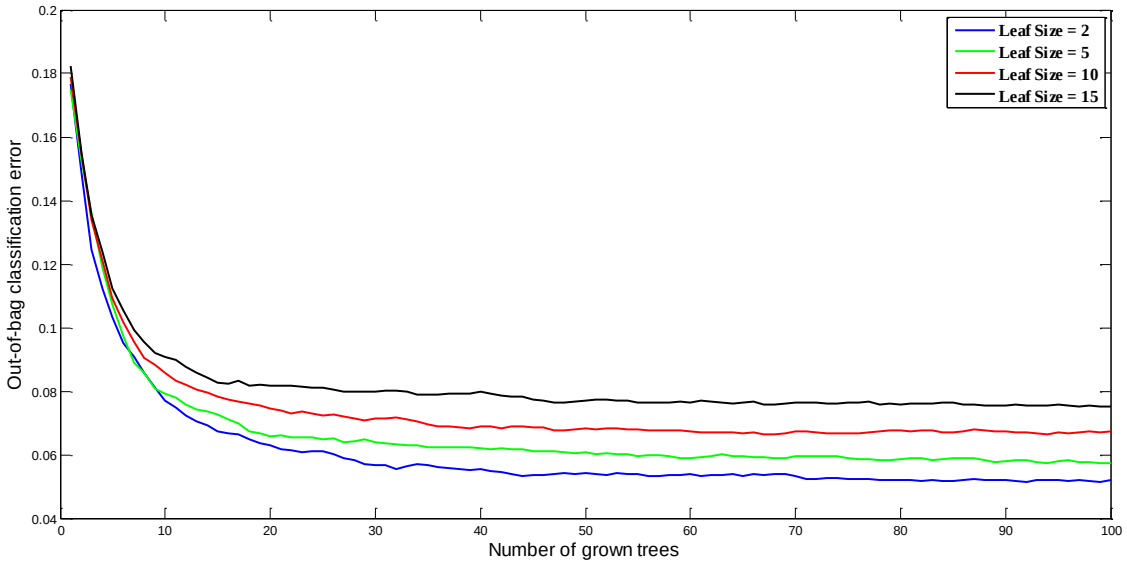

Fig. 1 Classification error for different leaf sizes

Table 3 Classification result for training; Values are reported as mean (standard deviation)

	Precision	**Recall**	**F-measure**
N (Normal)	0.999 (0.001)	0.995 (0.002)	0.997 (0.001)
S (Suspect)	0.973 (0.009)	0.989 (0.005)	0.981 (0.006)
P (Pathologic)	0.984 (0.011)	0.995 (0.006)	0.989 (0.006)
Weighted Average	**0.994**	**0.994**	**0.994**

Table 4 Classification result for testing; Values are reported as mean (standard deviation)

	Precision	**Recall**	**F-measure**
N (Normal)	0.981 (0.006)	0.947 (0.01)	0.964 (0.004)
S (Suspect)	0.729 (0.035)	0.879 (0.025)	0.796 (0.017)
P (Pathologic)	0.895 (0.043)	0.931 (0.043)	0.912 (0.029)
Weighted Average	**0.939**	**0.936**	**0.936**

Table 5 Confusion matrix for one of testing set

	N (Normal)	**S (Suspect)**	**P (Pathologic)**
N (Normal)	479	25	4
S (Suspect)	8	70	0
P (Pathologic)	1	0	51

Confusion matrix for one of the testing data set is given in Table 5. Most of the Normal class is identified as Normal class whereas 25 cases of suspect (S) class is confused with normal (N) class. Few cases of pathological (P) class (only 4) are confused with the normal class. Overall classification accuracy is 94% for the testing data set shown in Table 5.

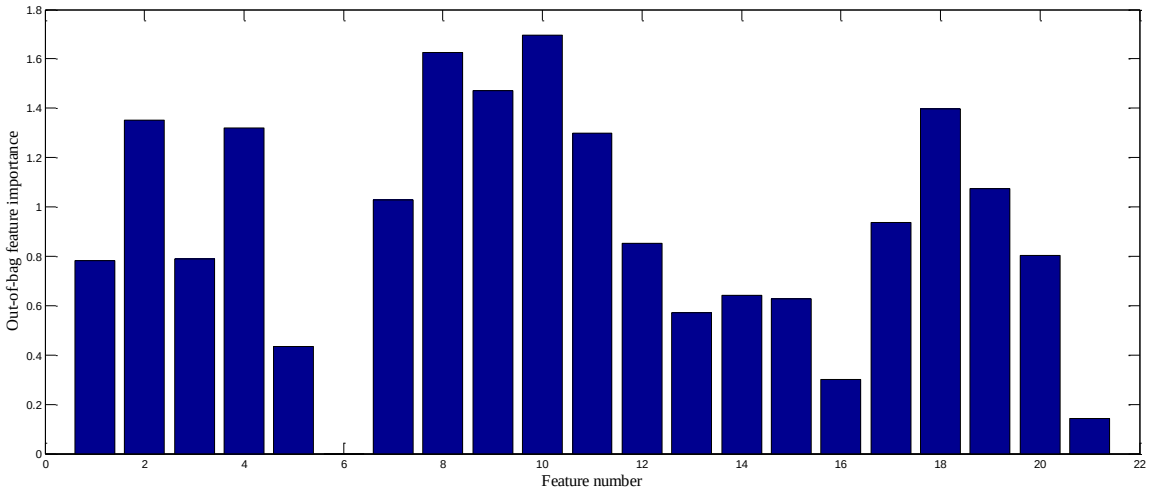

Fig. 2 Feature importance

Table 6 Performance of selection of reduced feature set using different threshold values

			F-measure of all three classes		
Threshold	**Feature Size**	**Overall Accuracy**	**N (Normal)**	**S (Suspect)**	**P (Pathologic)**
0.1	20	0.936	0.964	0.812	0.876
0.3	18	0.926	0.957	0.784	0.869
0.5	17	0.936	0.964	0.807	0.891
0.7	14	0.936	0.963	0.802	0.902
0.9	10	0.931	0.960	0.804	0.874
1.1	7	0.933	0.961	0.798	0.889
1.3	6	0.918	0.952	0.759	0.860
1.5	2	0.865	0.919	0.698	0.493

In the dataset, there are 21 features and all features may not be equally important in contributing the classification. So it is importance to study the importance of features in the classification of all three classes. In this paper, we have studied permutation based feature importance criteria. In this criterion, values of one feature are randomly permuted in a classification tree and increase in the mean square error (MSE) is recorded for out-of-bag samples. Hence this increase in MSE reflects how sensitive the values of the particular feature are with respect to the classes. This increase in MSE is averaged over all the trees in the random forest and divided by the standard deviation of MSE to get the feature importance. Large value of feature importance corresponds to the more important feature in the classification. Fig. 2 shows the out-of-bag importance of all the features. It can be observed that feature number 6(DS) and 21(Tendency) are the least important features whereas feature 10(ALTV) has the highest value of feature importance. A reduced set of the features can be selected by defining an appropriate threshold value below which all features are rejected. To study the effect of threshold on the size of reduced feature set and the classification accuracy, threshold values are iterated from 0.1 to 1.5 in the step

of 0.2 and classification accuracy is analyzed for the reduced feature sets in terms of F-measure. Whole data set is divided randomly into training and testing data sets (70-30) and all experiments are done ten times and mean values are reported. Table 6 summarizes the results. As the threshold value is increasing, number of selected features is decreasing because only features having the feature importance values above the threshold is selected. It can be observed that at the threshold value of 1.1, number of features is reduced to 7 but the overall classification accuracy is not significantly degraded. For threshold value of 1.5, only two most important features are selected but the overall classification accuracy is dropped to 86.5%.

F-measure for all three classes is plotted in Fig. 3 for different threshold values. In Fig. 3, Threshold values along with the number of features selected are shown on *x*-axis. It can be observed from the figure that f-measure remains almost constant for all three classes up to threshold value of 1.1 (number of features for this threshold value is 7). Hence threshold value of 1.1 is optimal and reduced feature set contains only these seven important features. These features are AC, UC, ASTV, MSTV, ALTV, MLTV and Mean of Histogram (described in Table 7). Reduced feature set elaborates that abnormal short term and long term variability of FHR is an important analysis. Similarly uterine contractions, Acceleration of FHR and Mean value of the histogram of FHR are also equally important. For this reduced set, overall classification accuracy is 93.3% and F-measures are 0.961, 0.798 and 0.889 for N, S and P classes respectively. Detailed result of reduced feature set is tabulated in Table 8. Precision, recall and f-measure for all three classes are found to be comparable with the full feature set (Table 4). Clinical importance of these features are highlighted in the guidelines (Macones *et al*. 2008) and other publish literature (Ugwumadu 2013, Chen *et al*. 2014).

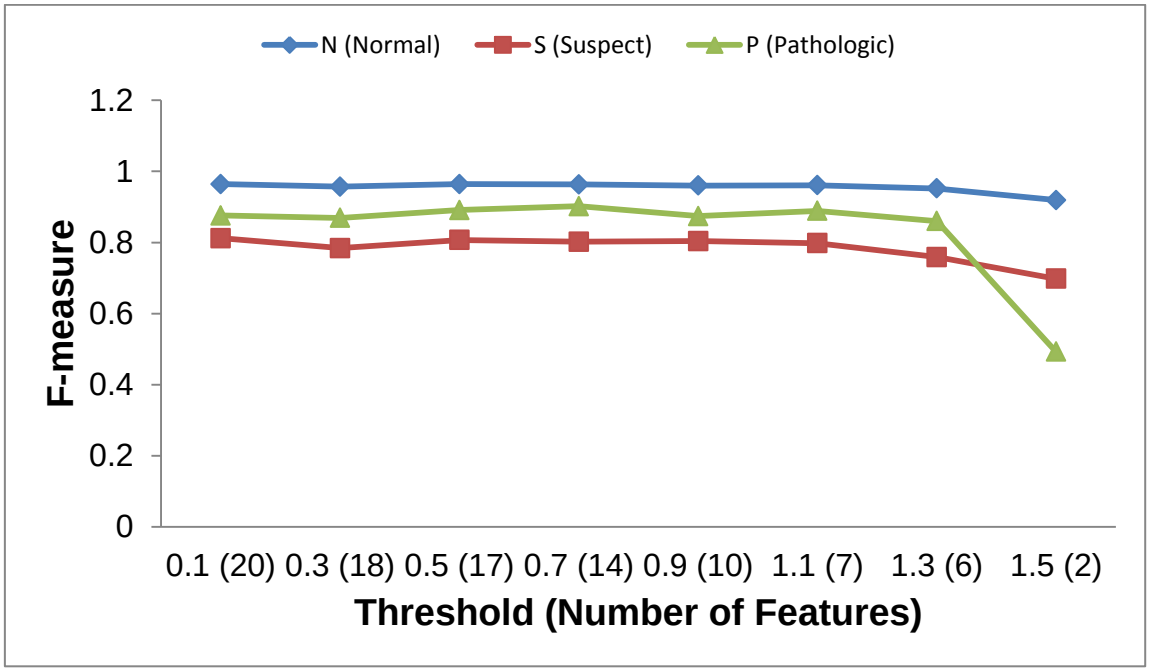

Fig. 3 F-measure of all three classes for different number of features

Table 7 Reduced features set

Feature	Explanation	Feature	Explanation
ASTV	Percentage of time with abnormal short term variability	**ALTV**	Percentage of time with abnormal long term variability
UC	Uterine contractions	**mLTV**	Mean value of long term variability
AC	Accelerations	**Mean**	Mean of the Histogram
mSTV	Mean value of short term variability		

Table 8 Classification result for testing data (Reduced feature set containg 7 features)

Class	Precision	Recall	F-measure
N (Normal)	0.982 (0.008)	0.941 (0.009)	0.961 (0.006)
S (Suspect)	0.717 (0.047)	0.899 (0.036)	0.798 (0.028)
P (Pathologic)	0.88 (0.041)	0.898 (0.039)	0.889 (0.032)
Weighted Average	**0.936**	**0.932**	**0.932**

Table 9 Comparison with reported results

Ref	Method	Results
(Sundar *et al*. 2013)	Neural network based classifier	Weighted average of Precision, Recall and F-measure of three classes is reported as, Precision (0.91), Recall (0.90) and F-Measure (0.90)
(Karabulut and Ibrikci 2014)	Decision Tree (C4.5)	Overall accuracy 95.014 (10 folds cross-validation)
(Menai, Mohder *et al*. 2013)	Relief F-15	Accuracy (0.939), Recall (0.915), Specificity (0.958)
(JEZEWSKI, NSKI *et al*.)	LSVM classifier	Sensitivity 83%, Specificity 92% (on all features)
(Chen *et al*. 2012)	FG-Kmeans	Precision (0.76) Recall (0.81) F-measure (0.77)
(Zhou and Sun 2014)	Active learning of Guassian Processes	Overall Accuracy 89% (small training dataset of 140 examples only)
(Cruz *et al*. 2014)	META-DES Ensemble Classifier	Overall accuracy 84.6%
This paper	Random Forest (Full Features)	**Training:** Precision, Recall and F-measure are **0.99** **Testing:** Precision, Recall and F-measure are **0.936** **Overall Accuracy: 93.6%**
This paper	Random Forest (Reduced Features, Seven Only)	**Testing:** Precision (**0.936**), Recall (**0.932**) and F-measure (**0.932)** **Overall Accuracy: 93.3%**

Classification results reported in the paper is compared with the published results in the literature in Table 9. In the paper, whole dataset is divided into 70% (training set) and 30% (testing set) randomly and classification results are reported as the average value of 10 independent runs. It can be seen that overall classification accuracy is better than most of the published results.

Karabulut *et al.* (Karabulut and Ibrikci 2014) reported overall accuracy of 95% for 10 folds cross validation using decision tree (C4.5) classifier. In our paper, we have used larger testing dataset for better analysis as the dataset is not balanced (N class is dominated). The most important contribution of our paper is the identification of important features (only seven) which can be used for the classification with the comparable classification accuracy of full features.

4. Conclusions

In this paper, we have evaluated the performance of random forest classifier using three different performance measures, namely Precision, Recall and F-measure to identify the pathological and suspicious states of the fetus from the normal state. Different parameters of the random forest classifier including leaf size and number of trees are optimized before applying the classifier. CTG dataset is divided into training and testing datasets randomly (70% for training and 30% for testing). Since the classifier is stochastic so ten folds cross validation is used with 70%-30% split of the CTG dataset. It is observed that overall classification accuracy of 93.6% can be achieved by the classifier on the test data set when whole feature space is used. Precision, Recall and F-measure are found to be 0.936 for the full feature space. Feature Importance index is used to identify the important features and a subspace of seven features is selected. Classification results on this feature subspace are found to be comparable with the full feature space.

Acknowledgements

The authors would like to thank the Deanship of Scientific Research at Umm Al-Qura University (project # 43408032) for the financial support.

References

Arif, M., Bilal, M., Kattan, A. and Ahamed, S.I. (2014), "Better physical activity classification using Smartphone acceleration sensor", *J. Med. Syst.*, **38**(9), 1-10.

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 cardiotocograms", *J. Maternal-Fetal Neonatal Med.*, **9**(5), 311-318.

Ayres-de-Campos, D., Sousa, P., Costa, A. and Bernardes, J. (2008), "Omniview-SisPorto® 3.5-a central fetal monitoring station with online alerts based on computerized cardiotocogram+ ST event analysis", *J. Perinatal Med.*, **36**(3), 260-264.

Breiman, L. (2001), "Random forests", *Machine Learn.*, **45**(1), 5-32.

Brown, R., Wijekoon, J.H., Fernando, A., Johnstone, E.D. and Heazell, A.E. (2014), "Continuous objective recording of fetal heart rate and fetal movements could reliably identify fetal compromise, which could reduce stillbirth rates by facilitating timely management", *Med. Hypotheses*, **83**(3), 410-417.

Carbonne, B., Langer, B., Goffinet, F., Audibert, F., Tardif, D., Le Goueff, F. and French Study Group on Fetal Pulse Oximetry (1997), "Multicenter study on the clinical value of fetal pulse oximetry", *Am. J. Obstetric. Gynecol.*, **177**(3), 593-598.

Chen, C.Y., Yu, C., Chang, C.C. and Lin, C.W. (2014), *Comparison of a novel computerized analysis program and visual interpretation of cardiotocography*, e112296.

Chen, H.Y., Chauhan, S.P., Ananth, C.V., Vintzileos, A.M. and Abuhamad, A.Z. (2011), "Electronic fetal heart rate monitoring and its relationship to neonatal and infant mortality in the United States", *Am. J. Obstetric. Gynecol*., **204**(6), 491-e1.

Chen, X., Ye, Y., Xu, X. and Huang, J.Z. (2012), "A feature group weighting method for subspace clustering of high-dimensional data", *Patt0 Recognition*, **45**(1), 434-446.

Costa, M.D., Schnettler, W.T., Amorim-Costa, C., Bernardes, J., Costa, A., Goldberger, A.L. and Ayres-de-Campos, D. (2014), "Complexity-loss in fetal heart rate dynamics during labor as a potential biomarker of acidemia", *Ear. Human Develop*., **90**(1), 67-71.

Cruz, R.M., Sabourin, R., Cavalcanti, G.D. and Ren, T.I. (2015), "META-DES: A dynamic ensemble selection framework using meta-learning", *Patt. Recognition*, **48**(5), 1925-1935.

Czabanski, R., Jezewski, J., Matonia, A. and Jezewski, M. (2012), "Computerized analysis of fetal heart rate signals as the predictor of neonatal acidemia", *Expert Syst. Appl*., **39**(15), 11846-11860.

Czabanski, R., Jezewski, M., Wrobel, J., Horoba, K. and Jezewski, J. (2008, January), "A neuro-fuzzy approach to the classification of fetal cardiotocograms", *In 14th Nordic-Baltic Conference on Biomedical Engineering and Medical Physics*, Springer Berlin Heidelberg.

Dietterich, T.G. (2000), "Ensemble methods in machine learning", *In Multiple classifier systems*, Springer Berlin Heidelberg.

Dietterich, T.G. (2000), "An experimental comparison of three methods for constructing ensembles of decision trees: Bagging, boosting, and randomization", *Machine Learn..*, **40**(2), 139-157.

Galar, M., Fernandez, A., Barrenechea, E., Bustince, H. and Herrera, F. (2012), "A review on ensembles for the class imbalance problem: bagging-, boosting-, and hybrid-based approaches. Systems, Man, and Cybernetics, Part C: Applications and Reviews", *IEEE Trans.*, **42**(4), 463-484.

Georgieva, A., Payne, S.J., Moulden, M. and Redman, C.W. (2013), "Artificial neural networks applied to fetal monitoring in labour", *Neural Comput. Appl.*, **22**(1), 85-93.

Gonçalves, H., Rocha, A.P., Ayres-de-Campos, D. and Bernardes, J. (2006), "Linear and nonlinear fetal heart rate analysis of normal and acidemic fetuses in the minutes preceding delivery", *Med. Biol. Eng. Comput*., **44**(10), 847-855.

Gray, K.R., Aljabar, P., Heckemann, R.A., Hammers, A., Rueckert, D. and Alzheimer's Disease Neuroimaging Initiative (2013), "Random forest-based similarity measures for multi-modal classification of Alzheimer's disease", Neuroimage, **65**, 167-175.

Grivell, R.M., Alfirevic, Z., Gyte, G.M. and Devane, D. (2010), "Antenatal cardiotocography for fetal assessment", *Cochrane Database Syst. Rev*., **1**.

Jezewski, M., Wrobel, J., Labaj, P., Leski, J., Henzel, N., Horoba, K. and Jezewski, J. (2007), "Some practical remarks on neural networks approach to fetal cardiotocograms classification", *In Engineering in Medicine and Biology Society, 2007. EMBS 2007. 29th Annual International Conference of the IEEE*, IEEE.

Jezewski, M. and Eski, J.Ł. (2014), "The influence of cardiotocogram signal feature selection method on fetal state assessment efficacy".

Kandaswamy, K.K., Chou, K.C., Martinetz, T., Möller, S., Suganthan, P.N., Sridharan, S. and Pugalenthi, G. (2011), "AFP-Pred: A random forest approach for predicting antifreeze proteins from sequence-derived properties", *J. Theo. Biol*., **270**(1), 56-62.

Karabulut, E.M. and Ibrikci, T. (2014), "Analysis of cardiotocogram data for fetal distress determination by decision tree based adaptive boosting approach", *J. Comput. Commun*., **2**(09): 32.

Kwon, J.Y., Park, I.Y., Shin, J.C., Song, J., Tafreshi, R. and Lim, J. (2012), "Specific change in spectral power of fetal heart rate variability related to fetal acidemia during labor: comparison between preterm and term fetuses", *Ear. Human Develop*., **88**(4), 203-207.

Lees, C., Marlow, N., Arabin, B., Bilardo, C.M., Brezinka, C., Derks, J.B. and Wolf, H. (2013), "Perinatal morbidity and mortality in early-onset fetal growth restriction: cohort outcomes of the trial of randomized umbilical and fetal flow in Europe (TRUFFLE)", *Ultrasound Obstetric. Gynecol.*, **42**(4), 400-408.

Logier, R., Jonckheere, J.D., Jeanne, M. and Matis, R. (2008), "Fetal distress diagnosis using heart rate variability analysis: design of a high frequency variability index", *In Engineering in Medicine and*

Biology Society, 2008. EMBS 2008. 30th Annual International Conference of the IEEE, IEEE.

Macones, G.A., Hankins, G.D., Spong, C.Y., Hauth, J. and Moore, T. (2008), "The 2008 National Institute of Child Health and Human Development workshop report on electronic fetal monitoring: update on definitions, interpretation, and research guidelines", *J. Obstetric. Gynecol. Neonatal Nurs.*, **37**(5), 510-515.

Menai, M.E.B., Mohder, F.J. and Al-mutairi, F. (2013), "Influence of feature selection on naïve Bayes classifier for recognizing patterns in cardiotocograms", *J. Med. Bioeng.*, **2**(1).

Ocak, H. (2013), "A medical decision support system based on support vector machines and the genetic algorithm for the evaluation of fetal well-being", *J. Med. Syst.*, **37**(2), 1-9.

Ozcift, A. (2012), "Enhanced cancer recognition system based on random forests feature elimination algorithm", *J. Med. Syst*., **36**(4), 2577-2585.

Salamalekis, E., Hintipas, E., Salloum, I., Vasios, G., Loghis, C., Vitoratos, N. and Creatsas, G. (2006), "Computerized analysis of fetal heart rate variability using the matching pursuit technique as an indicator of fetal hypoxia during labor", *J. Maternal-Fetal Neonatal Med*., **19**(3), 165-169.

Spencer, J.A. (1993), "Clinical overview of cardiotocography", *BJOG: Int. J. Obstetrics Gynaecol.*, **100**(s9), 4-7.

Sundar, C., Chitradevi, M. and Geetharamani, G. (2013), "An overview of research challenges for classification of cardiotocogram data", *J. Comput. Sci*., **9**(2), 198.

Svetnik, V., Wang, T., Tong, C., Liaw, A., Sheridan, R.P. and Song, Q. (2005), "Boosting: An ensemble learning tool for compound classification and QSAR modeling", *J. Chem. Inform. Model*., **45**(3), 786-799.

Tibshirani, R. (1996), *Bias, variance and prediction error for classification rules*, University of Toronto, Department of Statistics.

Ugwumadu, A. (2013), "Understanding cardiotocographic patterns associated with intrapartum fetal hypoxia and neurologic injury", *Best Practice Res. Clin. Obstetric. Gynaecol*., **27**(4), 509-536.

van der Hout-van, M.B., Oei, S.G. and Bovendeerd, P.H. (2012), "A mathematical model for simulation of early decelerations in the cardiotocogram during labor", *Med. Eng. Phys*., **34**(5), 579-589.

Wolpert, D.H. and Macready, W.G. (1999), "An efficient method to estimate bagging's generalization error", *Machine Learn*., **35**(1), 41-55.

Zhou, J. and Sun, S. (2014), "Active learning of Gaussian processes with manifold-preserving graph reduction", *Neural Comput. Appl*., **25**(7-8), 1615-1625.

DH