Applied Soft Computing 33 (2015) 231–238

Contents lists available at ScienceDirect

Applied Soft Computing

journal homepage: www.elsevier.com/locate/asoc

Classification of the cardiotocogram data for anticipation of fetal risks using machine learning techniques

Hakan Sahin*, Abdulhamit Subasi

International Burch University, Faculty of Engineering and Information Technologies, Francuske Revolucije b.b., Ilidza, Sarajevo 71000, Bosnia and Herzegovina

ARTICLE INFO

Article history:
Received 8 December 2014
Received in revised form 31 March 2015
Accepted 16 April 2015
Available online 25 April 2015

Keywords:
Cardiotocogram
Support vector machines
Artificial neural network
Radial basis functions
Decision trees
k-Nearest neighbor and Random Forest

ABSTRACT

The aim of the research is evaluating the classification performances of eight different machine-learning methods on the antepartum cardiotocography (CTG) data. The classification is necessary to predict newborn health, especially for the critical cases. Cardiotocography is used for assisting the obstetricians' to obtain detailed information during the pregnancy as a technique of measuring fetal well-being, essentially in pregnant women having potential complications. The obstetricians describe CTG shortly as a continuous electronic record of the baby's heart rate took from the mother's abdomen. The acquired information is necessary to visualize unhealthiness of the embryo and gives an opportunity for early intervention prior to happening a permanent impairment to the embryo. The aim of the machine learning methods is by using attributes of data obtained from the uterine contraction (UC) and fetal heart rate (FHR) signals to classify as pathological or normal. The dataset contains 1831 instances with 21 attributes, examined by applying the methods. In the paper, the highest accuracy displayed as 99.2%.

1. Introduction

During pregnancy, it is not easy to acquire direct information about the fetus. Therefore, obstetricians should use indirect information related to a fetal state. Most important one is measuring the fetal heart rate (FHR) [1].

The term 'electronic fetal monitoring' is the other alternative name of CTG, however, it is conceivable as less specific appellation since CTG contains observation about the contractions of the mother and different kind of fetal monitoring too. One type of fetal assessments in pregnancy can be the antenatal CTG and it checks the fetal heart rate regarding as a biological indicator of baby health [2]. CTG is usually applied during the last trimester it means after the 28th week of pregnancy. The idea behind the usage of CTG is a monitoring test for the diagnosis of babies having either acute or chronic fetal hypoxia and getting information about the baby under the risk of developing hypoxia [2].

The accepted parameters for the fetus are stated as follows (they are all in beats per minute). Baseline fetal heart rate between 110 and 160. Baseline variance is larger than 5 [3].

Nowadays, cardiotocography is the most preferable oblique, investigative method, in practical usage, to observe fetal well-being. Cardiotocogram (CTG) comprises of two different signals, its uninterrupted recording of instant fetal heart rate (FHR) and uterine activity (UC). The information, which is obtained from CTG, is used for early identification of a pathological state (i.e. inherited heart deficiency, fetal suffering or hypoxia, etc.) and may assist the obstetrician to anticipate future complications and interfere before there is a permanent harm to the fetus. During the delivery the baby who is exposed to hypoxia may cause of death or temporary disablement. Due to improper diagnosis of the FHR pattern recordings and improper treatments applied to the fetus can cause more than half of these deaths [4].

While its practicality, there has been some discrepancy as to the utility and the success of CTG monitoring, particularly in low-risk pregnancies. If there is an incorrectly analyzed fetal pain, then, it may be referred to needless treatments or if there is an improper analysis of fetal welfare then it may be rejected required treatments [5].

The classification uses a performance evaluation measure, but it is not enough to decide for the vital case especially in medical diagnosis. Therefore, it is also suggested another type of performance evaluation tools such as the ROC (receiver operation characteristics) [6] and F_1-measure [7].

* Corresponding author. Tel.: +387 33 944 513; fax: +387 33 782 131.
E-mail addresses: hsahin@ibu.edu.ba (H. Sahin), asubasi@ibu.edu.ba (A. Subasi).

http://dx.doi.org/10.1016/j.asoc.2015.04.038
1568-4946/

Table 1
Summary of all CTG features used in classification [11].

Symbol	Attribute information
LB	FHR baseline (beats per minute)
AC	# of accelerations per second
FM	# of fetal movements per second
UC	# of uterine contractions per second
DL	# of light decelerations per second
DS	# of severe decelerations per second
DP	# of prolonged decelerations per second
ASTV	Percentage of time with abnormal short-term variability
MSTV	Mean value of short-term variability
ALTV	Percentage of time with abnormal long-term variability
MLTV	Mean value of long-term variability
Width	Width of FHR histogram
Min	Minimum of FHR histogram
Max	Maximum of FHR histogram
Nmax	# of histogram peaks
Nzeros	# of histogram zeros
Mode	Histogram mode
Mean	Histogram mean
Median	Histogram median
Variance	Histogram variance
Tendency	Histogram tendency
NSP	Fetal state class (code (N = normal; P = pathological))

In this study, it is used eight different machine-learning methods to examine the CTG data categorized as healthy or unhealthy according to the decisions of three obstetricians. This study presents a comparison of the performances of the machine learning methods using the open source software WEKA [8] in terms of accuracy, specificity, sensitivity, *F*-measure and ROC curve.

2. Materials and methods

2.1. Dataset

The analysis shows the performance of the methods over the datasets from UCI [9] including CTG data with some indicative features. Three expert obstetricians decided to condition of the CTG data, whether normal or pathological checking the status of the embryo. The UCI cardiotocography data [9] was obtained by the automatic SISPORTO 2.0 [10] software. It is isolated from the suspicious entries and normal and pathologic class added to the NP feature. The Table 1 gives an explanation for each property of the respective features in the data. The CTG data has 21 features, 8 of them are continuous and 13 are discrete. The classification of the data with respect to condition of the fetus either normal or pathological. After giving a short explanation for each of the algorithms which are ANN, SVM, SL, RBF, C4.5, CART and RF decision trees. WEKA classification algorithms available at [11] were used for the algorithms. Each algorithm performance was tested using 10-fold cross validation during the examination for CTG dataset that tested around 1831 entries accurately. The CTG data entered in the WEKA classification module and trained several times by varying the respective parameters of the each algorithm to maximize the classification performance. The performances of the algorithms recorded and then tabulated.

2.2. Logistic regressions

Logistic regression analyzes the effect of several dynamics in a two-grouped outcome via approximating the probability of the experimental event [12]. The outcome of logistic regression is a binary event, like spam versus not spam, or malignant versus benign. The mathematical background of logistic regression is the concept logit that is described as "the natural logarithm of the odds ratio". In the literature, you can come up against the researchers who were calling the logistic regression as the logit model or logistic model formulated as follows,

$$log\frac{p(x)}{1-p(x)} = \beta_0 + x\cdot\beta \tag{1}$$

Why? Because, to achieve a goal, such as predicting the dichotomous outcome, instead of using a complicated equation, it opts linear one like its cousin linear regression. Solving for p, gives us a new formula depends on the variable x

$$p(x; w, n) = \frac{e^{\beta_0+x\cdot\beta}}{1+e^{\beta_0+x\cdot\beta}} = \frac{1}{1+e^{-(\beta_0+x\cdot\beta)}} \tag{2}$$

That equation includes all necessary coefficients, which are accepted as the inputs of the model. In the literature, it is encountered with two different types of usage of logistic model: Of course, to categorize the data according to its classes by calculating the ratio of probabilities success and failure in the form of the odds ratio. Secondly, to find out the relationships between the variables and how effects each other [13].

2.3. k-Nearest neighbor

The *k*-nearest-neighbors are an example of a classification method with the independence of the parameters. The method can be classified as by the point of implication as simple, but efficient in many datasets [14].

The *k*-NN algorithm is defined by three terms (S, k, T) where S represents a resemblance measure which links to each pair of data in an appropriate N-dimensional space at (real or integer) number, k represents the number of nearest data that are trained to carry out the classification and T represents the vector of M training data applied by the classifier to actually carry out the classification [15].

k-NN uses distance metrics usually Euclidean distance to perform the similarity measure formulated by

$$d_E = \sum_{i=1}^{N}\sqrt{x_i^2 - y_i^2} \tag{3}$$

If t is a data sample, which classification implemented and its k nearest neighbors turn up, then this makes a *neighborhood of t*. While using the method to classify the data sample in the neighborhood about t, the distance is whether considered or not. Of course, to use *k*-NN choosing the right value for k is critical; due to the rate of success of classification directly appertain to this value. The *k*-NN method can pass for biased by k. It is possible to choose the k-value in so many different ways. Of course, the smooth and efficient one is applying the algorithm several times changing the value of k to get the highest accuracy. To decrease the dependency level of choice of k for *k*-NN method, Wang [16] recommended checking multiple sets of nearest neighbors instead of one set of nearest neighbors.

k-NN can be applied by using the following procedure,

1.	Keep all the outputs of the P nearest neighbors to examine a part r in the vector set $m = \left\{m^1, \ldots, m^p\right\}$ by applying the steps P times:
a.	Take the next part r^i in the data, where i represents the current iteration in the given set $\{1, \ldots, Q\}$
b.	If r is not assigned or $r < d(r, m^i) : r \leftarrow d(r, m^i)$, $t \leftarrow s^i$
c.	Repeat until the last element of the data (i.e. $i = Q$)
d.	Keep r as the vector c and t as the vector m
2.	Compute the average output for the vector m by the formula as follows: $\bar{m} = \frac{1}{P}\sum_{i=1}^{P} m_i$
3.	Assign $\bar{m}$ as the output value of the examination data part r.

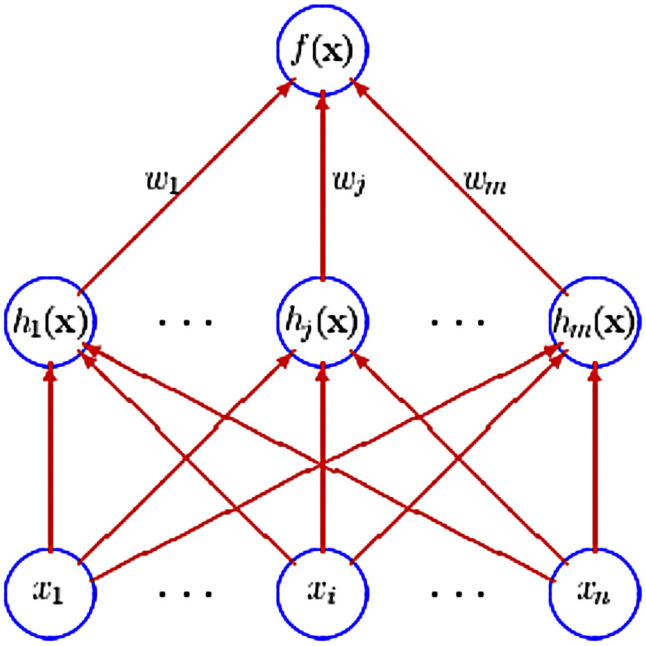

Fig. 1. The typical radial basis function network. Each of n elements of the input vector $\boldsymbol{x}$ sends to m basis function whose outputs are linear combination of weights $\{w_j\}_{j=1}^{m}$ and network output $f(\boldsymbol{x})$.

2.4. Radial basis function network

The **RBFN** is derived from the methods for executing full interpolation of a subpart of data points in a high dimensional feature space [17]. A usual radial function is the Gaussian that is having an arbitrary value c, is as follows

$$h(x) = e^{\left(-\frac{(x-c)^2}{r^2}\right)} \tag{4}$$

The parameters of the function are c for the center and r for the radius.

The **RBFN** shows similarity in network design with the typical regularization network [18]. Since the basis functions are radially symmetrical due to the stabilizer having radial symmetry, it calls the **RBFN**. From the point of approximation theory, the regularization network needs three properties [19] expressed as follows,

- It can converge any multivariate continuous function on a compact set to a random precision, indicated an infinitely many units.
- The regularization network shows the best-approximation attribute since the unidentified coefficients are linear.
- The result calculated by the regularization network is optimum. Optimality shows that the regularization network reduces a function that measures how much the solution diverges from its actual value as denoted by the training data [20].

Fig. 1 illustrates a typical **RBF** network [21]

2.5. Artificial neural networks (ANNs)

ANNs are computing structures inspired from the biological neural networks. ANN constitutes of the concomitant operating units. They are capable of learning by adjusting the weights of the interconnections through the input data [22].

Fig. 2 demonstrates the typical feed forward neural network.

ANN consists of neurons, each of which has one or more weighted inputs and one or more output that are weighted when connecting to other neurons. Neuron collects the weighted inputs and transports the net input through an initiation function in order to create a result [23].

$$h_i = \sigma\left(\sum_{j=1}^{N} V_{ij}x_j + T_i^{\text{hidden}}\right) \tag{5}$$

where $\sigma()$ represents activation function, N symbolizes the number of input neurons V_{ij} the weights, x_j inputs for the input neuron, and T_i^{hidden} the threshold value for the hidden neurons [24].

The limitations of perceptron can overcome by feed forward multilayer networks with non-linear node functions. However, the simple perceptron learning algorithm cannot be converted just when it is transported from a single layer of perceptrons in multiple layers of perceptrons. These feed forward networks are sometimes given the name "multilayer perceptrons" (MLPs). The expression "back propagation network" is sometimes referred as the definition of feed forward neural networks trained by the back propagation learning method [25].

2.6. Support vector machine

Support vector machines (SVM) were developed by [26] for using binomial classification. It gained popularity amongst the other machine learning algorithms. Because of, implementing it in wide areas of biomedical science (i.e. classification of protein compounds as high as 90%) to computer science (i.e. classification of handwritings, images, text and hypertext) without any need to look for different machine learning algorithm. It needs a hyperplane to split multidimensional data into two classes. The following formula is used two linearly separable data if the ordered $(\mathbf{w}, b)$ exists

$$\boldsymbol{w}^T x_i + b \geq 1, \quad \text{for all } x_i \in C_+ \text{ or } \quad \boldsymbol{w}^T x_i + b \leq -1, \quad \text{for all } x_i \in C_- \tag{6}$$

The data gathered from the real life usually are nonlinear that means it is not easy to find a line to separate data into two groups. SVM accomplishes this nonlinearity problem by coming up with the concept of a "kernel-induced feature space", which transfers the data into higher dimensional space to convert it as separable data. It comes to mind how SVM overcomes that time-consuming complexity of transformation and over fitting data problem. Since during the computation, it is used just dot product in a higher dimensional space, which provides SVM avoiding such computational and over fitting problem. Besides the classification, the other implementation purpose of SVM is regression [27].

2.7. Classification and regression trees (CART)

Classification and regression trees (CART) considered as a decision tree method applied for classification purposes using the diachronic data. CART as a machine learning method is required to be determined number of classes. CART was proposed in 80s by [28]. In order to build decision trees, CART needs learning example. Decision trees are symbolized by a group of queries, which is splitting the learning sample into small and insignificant parts. In order to achieve the best split, the question that separates the data into two homogeneous parts, CART method examines for all possible variables and values [29].

CART uses Gini index to select the attribute which has maximum information. If a data A with n classes has Gini index that is defined as,

$$\text{Gini}(A) = 1 - \sum_{k=1}^{n} p_j^2 \tag{7}$$

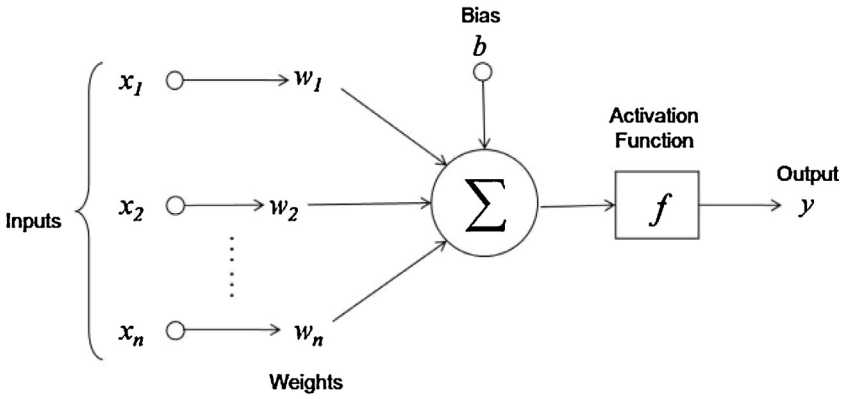

Fig. 2. The typical feedforward neural network with bias.

2.8. C4.5 decision tree classifier

Most experimental learning systems need to know the type of classes exists in a given dataset. In addition, for each system, there is a vector of attribute values, and a mapping function to correspond from attribute values to classes. The attributes applied to define examples can be assembled into continuous attributes, whose figure are numeric, and discrete attributes with ungraded nominal values. C4.5 [30] can be given an example of that system which learns decision-tree classifiers [31].

C4.5 utilize "Information Gain" to get a new measurement that is called as "Gain Ratio". They are defined as in the following formula.

$$\text{Entropy}(P) = -\sum_{i=1}^{N} p_i \log(p_i) \tag{8}$$

$$\text{GainRatio}(p, T) \frac{\text{Gain}(p, T)}{\text{SplitInfo}(p, t)} \tag{9}$$

where splitInfo is defined as

$$\text{SplitInfo}(p, \text{test}) = -\sum_{j=1}^{N} p'\left(\frac{j}{p}\right) \cdot \log\left(p'\left(\frac{j}{p}\right)\right) \tag{10}$$

p is the probability distribution of the given data and log uses base as 2 due to measure information as 'bit'.

2.9. Random Forest (RF)

Random Forest (RF) gives a particular synthesis of classification accuracy and being a model having an exposition between the traditional artificial intelligence algorithms. The arbitrary instantiation and way of creating group applied in Random Forest give it a chance to succeed reliable prediction besides better generalizations. The property of generalization is derived from the bagging sketch, which improves the generalization by diminishing variance, even though dependent processes like boosting perform this by diminishing bias [32].

Three features of Random Forest obtain the primary focus [33]:

- It gives the opportunity of reliably classification in a different field of technique.
- It provides evaluating the significance of each attribute having model training.
- The trained model is used to measure pairwise proximity among samples.

3. Result and discussion

Due to use of the WEKA workbench, it allows to change and find the appropriate parameters. Firstly, the default values of algorithms' parameters are tested and according to results of confusion matrices, it is tested to find the best accuracy that is achieved by the respective classifier.

ANN in the workbench has seven parameters, but we have tested only three parameters as follows number of hidden layers was assigned 11, learning rate was 0.21 and momentum was 0.2.

SVM has four parameters, but it is used only tested two parameters that were kernel as RBF, cost value was 9000.

Logistic regression has four parameters just two were tested they were M as 1000 and heuristic stop as 151 respectively.

RBF has three parameters just two were tested, which were clustering seed as 121 and number of clusters as 15.

C4.5 (J48) has four parameters we were tested three of them such as confidence factor as .025, seed as 100 and number of folds as 6.

CART has four parameters we have tested three of them in order number of folding for pruning as 0.2, seed as 100 and minimum number of object as 2

RF has four parameters we tested three of them such as N as 13, F as 4 and T as 29.

And the last one k-NN has three parameters and just two were tested k as 11 and linear search algorithm as Euclidean distance.

3.1. Performance evaluation measurements

Predicting performance of a machine learning method based on inadequate data is difficult. Therefore, Cross-validation becomes the favorite when the researcher got a small amount of data [34]. When machine-learning algorithms are used, decisions must be made on how to divide data for training and testing. With the aim of calculating the performance of machine learning methods, the entire CTG data split training and testing sets, and 10-fold cross-validation, which is a famous method for evaluation, is applied afterwards. The classification of data examined by a training set as forming a model, the verification of the model performed by the test set.

The number of true negatives (TN), false negatives (FN), true positives (TP), and false positives (FP) are used to compute the efficiency of the classifier. The sensitivity and specificity are statistical measurements of checkout tests. Sensitivity states in the rate of positive test result,

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}} \times 100\%$$

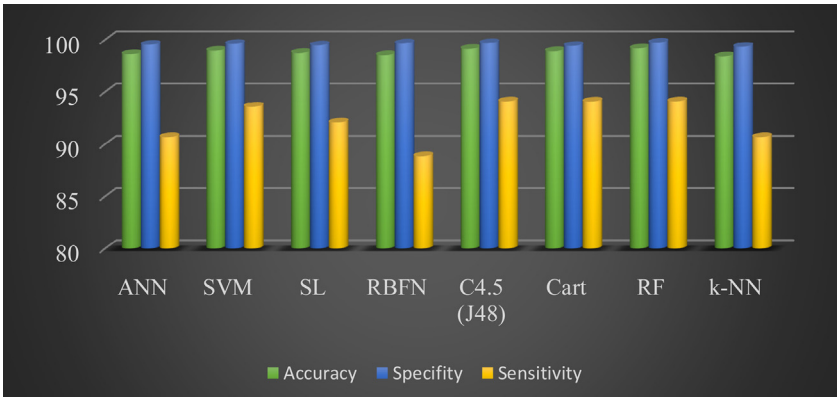

Fig. 3. The performances of the respective machine learning methods.

Specificity indicates to the ratio of a negative test result, which has the following formula

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}} \times 100\%$$

Accuracy shows overall measure, which is:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{TN} + \text{FN}} \times 100\%$$

ROC represents the classifier performance without considering class distribution or error costs. A receiver operating characteristics is made by drawing all specific values versus correspondent sensitivity values [35,36]. The approximation excellence relies upon the amount of thresholds tested. Regardless of its good sides, the ROC plot does not provide a standard for the classification of cases. On the other hand, there are methods that can be applied to get decision formulas by using the ROC drawing [36,37].

Deciding the suitable threshold from a ROC curve is contingent upon on getting figures for the notional costs of false-negative and false positive errors.

There is another statistical measurement call *F*-measure; to evaluate characterization of the performance, has the following formula:

$$F\text{-measure} = \frac{2\text{TP}}{2\text{TP} + \text{FP} + \text{FN}}$$

In Table 2, the applied measurements are given with their mathematical expressions.

3.2. Experimental results

The performance calculation of the each method; all CTG data broken up two parts, one for training and the rest as test sets. That process calls *k*-fold cross validation, which asserted by Salzberg [38]. *k*-Fold cross validation provides avoiding to pick a particular parts that are for training and testing. In the paper, the number of *k* adjusted to 10; for this reason, the CTG dataset was split 10 parts. For the training process, it is necessary to use nine data elements while the testing process needs the residuary part to perform [39]. To finish up the process, the procedure was repeated 10 times to allow each fraction of data being as a testing data.

Fig. 3 shows the comparison of the algorithm testing performance. As it displayed in Fig. 3, the Random Forest gets the maximum performance of classification with respect the others. As a result of that decision trees (Random Forest is a greedy method that selects the best solution at hand when choosing classification structures that are to be applied for tests at each tree node. The solution is using pruning the tree.

In the paper, all classifiers were performed on the data without excluding any features. During the training phase, the classifier parameters chosen with respect to a *k*-fold cross-validation (CV) procedure [40]. For each algorithm, respected ROC and *F*-measure also computed at the end of the process.

To assure for the performance of each algorithm, the value of *k* assigned to 10 to implement 10-fold cross validation. Table 3 displays respected calculation of statistical measurements of ROC and *F*-measure.

As shown in Fig. 4, total accuracy, ROC and *F*-measures accomplished with the RF classifier on the test set were equal to 99.18%, 0.999 and 0.992, respectively.

These results were comparatively better than the other classifiers got. The overall accuracies of the classifiers laid in the in the interval (98.42–99.13) where 99.13%, 98.96%, 98.91%, 98.74%, 98.63%, 98.53% and 98.42% for the C4.5, SVM, CART, SL, ANN, RBFN and *k*-NN classifiers, respectively.

The area under curves were equal to 0.996, 0.993, 0.985, 0.976, 0.966, 0.966 and 0.954 for the SL, ANN, *k*-NN, C4.5, SVM, RBFN, and CART classifiers respectively. The *F*-measures were getting the following result 0.991 for the C4.5 classifier, 0.990 for the SVM classifier and 0.989, 0987, 0986, 0985, 0.984 for the classifiers CART, SL. ANN, RBF as the last classifier *k*-NN, respectively.

This paper reveals to corroborate the investigation of the other field of application, in other words, the supremacy of Random Forest about conventional methods when handling with CTG data.

Exact specification of CTG data has been critical for not just a diagnosis, but also an amendment evaluation. The Random

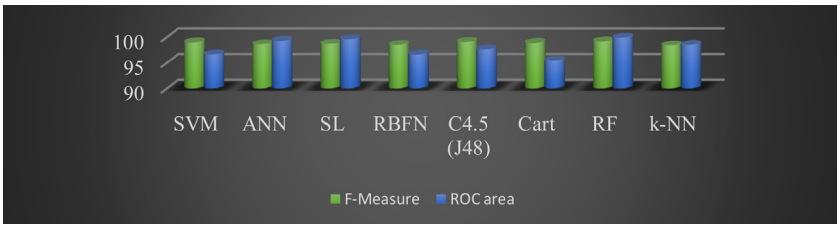

Fig. 4. Respective *F*-measure and ROC area of the algorithms.

Table 2
Measures for binary classification [33].

Measure	Formula	Evaluation focus
Accuracy	$\frac{TP+TN}{TP+TN+FP+FN}$	Overall efficiency of a classifier
Sensitivity(recall)	$\frac{TP}{TP+FN}$	The efficiency of a classifier to categorize positively labeled data
Specificity	$\frac{TN}{TN+FP}$	Performance of a classifier when categorizes negative labels
Precision	$\frac{TP}{TP+FP}$	The data with the positive labels correctly classified by the classifier
F-score	$\frac{2*P*R}{P+R}$	Harmonic mean between precision and recall
AUC	$\frac{1}{2}\left(\frac{TP}{TP+FN}+\frac{TN}{TN+FP}\right)$	The classifier's power to prevent misclassification

Table 3
Comparison chart of the classifiers given in percentage (%).

	ANN	SVM	SL	RBF	C4.5	CART	RF	*k*-NN
Accuracy	98.63	98.96	98.74	98.53	99.13	98.91	**99.18**	98.42
Specificity	99.53	99.61	99.47	99.66	99.68	99.42	**99.74**	99.33
Sensitivity	90.69	93.59	92.11	88.86	94.10	94.09	**94.10**	90.69

The bold entries show the highest value in the table respective row and column.

Forest algorithm classifies CTG data with an accuracy of 99.18%. The sensitivity and the specificity values of Random Forest are 94.18% and 99.74%, respectively. Observation of the table forms an opinion about being the best under the consideration of ROC area (AUC = 0.999) and *F*-measure (0.992) of Random Forest among the other classifiers by a long way.

The analysis of the values of specific as for the biomedical data has more importance than sensitivity. So, it can be said that the classifier shows its power of classification in the pathological state with obtaining high specificity. According to Table 3, it is easily said all the classifiers show high specificity. It means that, all algorithms show a nearly same performance as a classifier, since the error range of the classifiers is just between 0.26% and 0.67%.

However, the analysis of the sensitivities of the classifiers shows a different face of the data. According to Table 2, the sensitivities are less than specified for all classifiers. It means that the algorithms had the difficulties to classify the positively labeled data. The error rates vary between 5.90% and 11.14%. Again, RF got the minimum error rate, and *k*-NN got the maximum error rate.

According to sensitivity, specificity and accuracy criteria RF preserved leadership in its classification capacity. The Table 3 displays another two-evaluation measurement. One is *F*-measure that is used to show how similar two classifying results are. Since the higher value gives a better result, therefore, again the winner is RF then in order it follows as C4.5, SVM, CART, SL, ANN, RBF and *k*-NN.

The area under the ROC curve call as one of the important statistical measures, which is needed during the classification of the data. If the curve area is exactly one, then, the ROC curve establishes an intersection of two perpendicular lines. One of them is perpendicular to the *x*-axis with one unit length and similarly the other is perpendicular to the *y*-axis with one unit length . This analysis shows 100% precise inasmuch as both the specificity and sensitivity set the value of one, so there is no misclassification. However, if a classifier cannot distinguish normal and pathological corresponds then a ROC curve is constructed as a line segment plotted from the origin to the point [1, 1]. The area of the ROC is calculated under this line segment as half of full. The computed ROC curve areas are lying between the interval 0.5 and 1.0. It is enough to check the contrast of the ROC curves for each classifier to be compared [41].

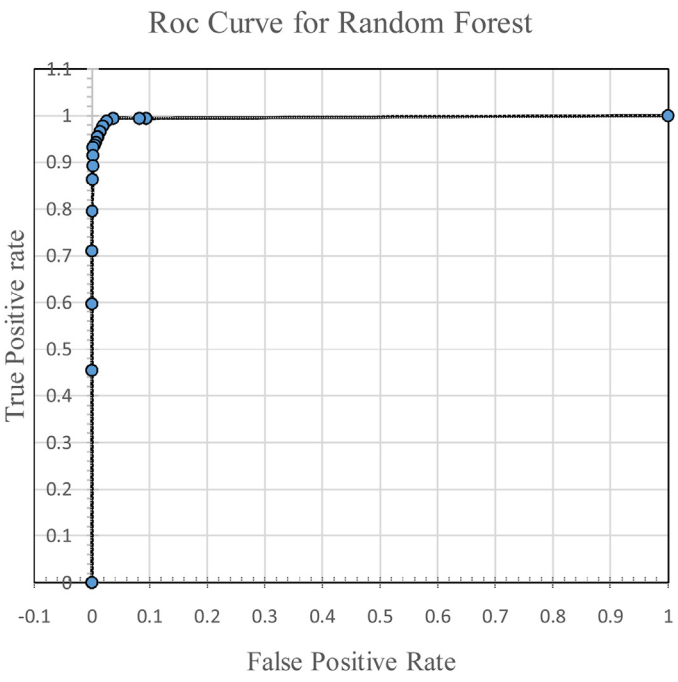

Fig. 5. The respective ROC curve for Random Forest algorithm.

The highest ROC curve value belongs to RF as 0.999 nearly 1. It means that the Random Forest can classify the data with high precision. The ROC curve of random forest is illustrated in Fig. 5. After the Random Forest, the highest value was got by simple logistics surprisingly with the value of 0.996. The respective values are in order 0.993, 0.985, 0.976, 0.966, 0.954, which were got by ANN, *k*-NN, C4.5, SVM, RBF and CART.

3.3. Discussion

Several researches have done about the classification of the Cardiotocogram data. Huang [42] analyzed the CTG data by three machine learning methods to create the classification models to

Table 4
F-Measure and ROC (receiver operating characteristic) area.

	ANN	SVM	SL	RBF	C4.5	CART	RF	*k*-NN
F-measure	0.986	0.990	0.987	0.985	0.991	0.989	**0.992**	0.984
ROC area	0.993	0.966	0.996	0.966	0.976	0.954	**0.999**	0.985

The bold entries show the highest value in the table respective row and column.

predict fetal distress. Krupa et al. [43] suggested using statistical features extracted from empirical mode decomposition (EMD). The extracted features from the decomposed components classified as normal and at risk. The reported accuracy of the test data is 86%.

In another study, it is proposed a two-step investigation of fetal heart rate data that lets for efficient prediction of the acidemia risk. After the classification of FHR signals by fuzzy, multilayer perceptron and Lagrangian support vector machines (LSVM) were performed to classify the nonlinear recording. The performance depends on the number of accurate classifications [44] and quality index [45] specified as the geometric mean of sensitivity and specificity.

According to [46] CC and QI get through to the highest value for the Weighted Fuzzy Scoring System combined with the LSVM algorithm. Another work [47] suggest classification method using SVM but before the applying the method they eliminated the noisy data from the FHR recording and also reduced the dimension by PCA. It is reported that the overall classification accuracy is 75.61% and AUC as 0.78. Again the same author [48] used artifact removal from the FHR data followed by feature extraction, then applied HMM (hidden Markov models) to classify the data and reported overall accuracy 83%.

Sundar et al. [49] tried to develop a new model uses CTG data to classify by ANN. The statistical parameters such as *F*-score, Recall were computed to evaluate the performance and reported overall performance are 84%, 78% and 80%, respectively. Furthermore, they proposed [50] calculation of the performance of the clustering of CTG data by *k*-means using the precision, recall and *F*-score measures. The general accuracies stated as 80%, 35% and 45%, respectively [50]. Ocak and Ertunc [51] proposed adaptive neuro-fuzzy inference systems (ANFIS) and reported 97.5% and 95.8% accuracies for ANFIS and ANN based classifiers, respectively. Ocak [52] proposed SVM and genetic algorithm (GA) based classification technique and 99.4% accuracy reported which is the highest as reported in this dataset.

Recent research has done [53] using hybrid model. The model suggests using LS-SVM and PSO (particle swarm optimization) to train the CTG data after getting trained data again applying the same procedure then in the last step applying Binary Decision Tree to get classified data. The reported classification rate is 91.62%.

In our experiment, we used UCI CTG dataset which has actually 2126 entries, but 295 were belonging to suspicious class. Since the aim of classification is to help to clinicians during the diagnosis, 295 of the data does not help and provide useful information to apply any treatment. Because of the given idea. It is excluded from the dataset.

The comparison of the study with others is difficult due to not using the same CTG data. Mainly we can emphasize that according to all classifiers it seems that Random Forest gave the highest classification ratio in all type of measurements.

4. Conclusion

CTG data is useful for obstetricians to diagnose fetal abnormalities and used to decide for medical intervention before persistent damage on baby. However, interpretations of the CTG data after visual analysis done by obstetrician could not be objective.

This experiment aimed to evaluate the performance of eight different machine learning algorithms that are ANN, SVM, *k*-NN, RF, CART, Logistic Regression, C4.5 and RBFN over the UCI CTG dataset. To measure the classification performance of these classifiers it is applied 10-fold cross validation. The results of this study show that Random Forest can be accepted as a good classifier of normal and pathological classes of the CTG data.

References

[1] M. Cesarelli, M. Romano, P. Bifulco, Comparison of short term variability indexes in cardiotocographic fetal monitoring, Comput. Biol. Med. 39 (2009) 106–118.
[2] R.M. Grivell, Z. Alfirevic, G. Gyte, D. Devane, Antenatal cardiotocography for fetal assessment, Cochrane Database Syst. Rev. (2010) CD007863.
[3] C. Gribbin, J. Thornton, Critical evaluation of fetal assessment methods, High-Risk Pregnancy Manag. Options 3 (2006) 229–239.
[4] D. Ayres-de-Campos, C. Costa-Santos, J. Bernardes, S.M.V.S. Group, Prediction of neonatal state by computer analysis of fetal heart rate tracings: the antepartum arm of the SisPorto® multicentre validation study, Eur. J. Obstet. Gynecol. Reprod. Biol. 118 (2005) 52–60.
[5] H.P. van Geijn, H.W. Jongsma, J. de Haan, T.K. Eskes, Analysis of heart rate and beat-to-beat variability: Interval difference index, Am. J. Obstet. Gynecol. 138 (1980) 246–252.
[6] F.J. Provost, T. Fawcett, Analysis and visualization of classifier performance: Comparison under imprecise class and cost distributions, in: Proceedings of the Third International Conference on Knowledge Discovery and Data Mining, 1997, pp. 43–48.
[7] M. Sokolova, N. Japkowicz, S. Szpakowicz, Beyond accuracy, *F*-score and ROC: a family of discriminant measures for performance evaluation, in: AI 2006: Advances in Artificial Intelligence, Springer, 2006, pp. 1015–1021.
[8] M.L.G.a.t.U.o. Waikato, The WEKA Classification Algorithms, 2010.
[9] K. Bache, M. Lichman, Cardiotocography data set, in: UCI Machine Learning Repository, 2010.
[10] D. Ayres-de Campos, J. Bernardes, A. Garrido, J. Marques-de-Sa, L. Pereira-Leite, SisPorto 2.0: a program for automated analysis of cardiotocograms, J. Matern.-Fetal Med. 9 (2000) 311–318.
[11] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, I.H. Witten, The WEKA data mining software: an update, ACM SIGKDD Explor. Newsl. 11 (2009) 10–18.
[12] J. Anderson, Logistic regression, in: Handbook of Statistics, North-Holland, New York, 1982, pp. 169–191.
[13] R.P. Burns, R. Burns, Business Research Methods and Statistics Using SPSS, Sage, 2008.
[14] D. Hand, H. Mannila, P. Smyth, Principles of Data Mining, The MIT Press, 2001.
[15] Z.M.V. Kovacs, R. Guerrieri, A generalization technique for nearest-neighbor classifiers, in: Neural Networks, 1991, IEEE International Joint Conference, vol. 2743, 1991, pp. 2740–2745.
[16] H. Wang, Nearest Neighbours without k: A Classification Formalism based on Probability, Faculty of Informatics, University of Ulster, N. Ireland, 2002.
[17] M.J.D. Powell, Radial basis functions for multivariable interpolation: a review, in: Algorithms for Approximation, Clarendon Press, Oxford, 1987, pp. 143–167.
[18] T. Poggio, F. Girosi, Networks for approximation and learning, Proc. IEEE 78 (1990) 1481–1497.
[19] F. Girosi, T. Poggio, Networks and the best approximation property, Biol. Cybern. 63 (1990) 169–176.
[20] K.L. Du, M.N.S. Swamy, Radial basis function networks, in: Neural Networks in a Softcomputing Framework, Springer, London, 2006, pp. 251–294.
[21] M.J. Orr, Introduction to Radial Basis Function Networks, Technical Report, Center for Cognitive Science, University of Edinburgh, 1996.
[22] S. Haykin, Neural Networks: A Comprehensive Foundation, Prentice Hall, 1999.
[23] M.T. Jones, Artificial Intelligence: A Systems Approach, Infinity Science Press LLC, 2008.
[24] S.-C. Wang, Artificial neural network, in: Interdisciplinary Computing in Java Programming, Springer, US, 2003, pp. 81–100.
[25] K. Mehrotra, C.K. Mohan, S. Ranka, Elements of Artificial Neural Networks, MIT Press, 2000.
[26] C. Cortes, V. Vapnik, Support-vector networks, Mach. Learn. 20 (1995) 273–297.
[27] D. Boswell, Introduction to Support Vector Machines, 2002.
[28] L. Breiman, J.H. Friedman, R.A. Olshen, C.J. Stone, Classification and Regression Trees, Chapman & Hall, New York, NY, 1984.
[29] R. Timofeev, Classification and Regression Trees Theory and Applications, Berlin, Germany, 2004.
[30] J.R. Quinlan, C4.5: Programs for Machine Learning, Morgan Kaufmann, San Mateo, 1993.
[31] J.R. Quinlan, Improved use of continuous attributes in C4.5, J. Artif. Intell. Res. (1996) 77–90.
[32] P. Yang, Y. Hwa Yang, B. Zhou, Y. Zomaya, A review of ensemble methods in bioinformatics, Curr. Bioinform. 5 (2010) 296–308.
[33] L. Breiman, Random forests, Mach. Learn. 45 (2001) 5–32.
[34] I.H. Witten, E. Frank, Data Mining: Practical Machine Learning Tools and Techniques, Morgan Kaufmann Publishers (Elsevier), San Francisco, CA, 2005.
[35] A. Fielding, Cluster and Classification Techniques for the Biosciences, Cambridge University Press, 2007.
[36] S.L. Salzberg, On comparing classifiers: pitfalls to avoid and a recommended approach, Data Min. Knowl. Discov. (2007) 317–328.
[37] A.H. Fielding, Cluster and Classification Techniques for the Biosciences, Cambridge University Press, The Edinburgh Building, Cambridge, UK, 2007.
[38] S.L. Salzberg, On comparing classifiers: pitfalls to avoid and a recommended approach, Data Min. Knowl. Discov. 1 (1997) 317–328.
[39] R. Johnson, D. Wichern, Applied Multivariate Statistical Analysis, 5th ed., Prentice Hall, 2001.
[40] B.R. Bakshi, Multiscale PCA with application to multivariate statistical process monitoring, AIChE J. 44 (1998) 1596–1610.

[41] M.H. Zweig, G. Campbell, Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine, Clin. Chem. 39 (1993) 561–577.
[42] M.-L. Huang, Fetal distress prediction using discriminant analysis, decision tree, and artificial neural network, J. Biomed. Sci. Eng. 05 (2012) 526–533.
[43] N. Krupa, M. Ali, E. Zahedi, S. Ahmed, F.M. Hassan, Antepartum fetal heart rate feature extraction and classification using empirical mode decomposition and support vector machine, Biomed. Eng. Online 10 (2011) 6.
[44] C.A. Micchelli, Interpolation of scattered data: distance matrices and conditionally positive definite functions, Constr. Approxim. (1986) 11–22.
[45] Y. Qi, Random forest for bioinformatics, in: Ensemble Machine Learning: Methods and Applications, 2012, p. 307.
[46] R. Czabanski, J. Jezewski, A. Matonia, M. Jezewski, Computerized analysis of fetal heart rate signals as the predictor of neonatal acidemia, Expert Syst. Appl. 39 (2012) 11846–11860.
[47] G. Georgoulas, D. Stylios, P. Groumpos, Predicting the risk of metabolic acidosis for newborns based on fetal heart rate signal classification using support vector machines, IEEE Trans. Biomed. Eng. 53 (2006) 875–884.
[48] G.G. Georgoulas, C.D. Stylios, G. Nokas, P.P. Groumpos, Classification of fetal heart rate during labour using hidden Markov models, in: Neural Networks, 2004, Proceedings of IEEE International Joint Conference, vol. 2473, 2004, pp. 2471–2475.
[49] C. Sundar, M. Chitradevi, G. Geetharamani, Classification of CTG data using neural network based machine learning, Int. J. Comput. Appl. 47 (2012) 19–25.
[50] C. Sundar, An analysis on the performance of K-means clustering algorithm for cardiotocogram data clustering, Int. J. Comput. Sci. Appl. 2 (2012) 11–20.
[51] H. Ocak, H. Ertunc, Prediction of fetal state from the cardiotocogram recordings using adaptive neuro-fuzzy inference systems, Neural Comput. Appl. 23 (2013) 1583–1589.
[52] H. Ocak, 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 (2013) 9913.
[53] E. Yilmaz, C. Kilikcier, Determination of fetal state from cardiotocogram using LS-SVM with particle swarm optimization and binary decision tree, Comput. Math. Methods Med. 2013 (2013) 487179.