Advances in Computer Science and Information Technology (ACSIT)
p-ISSN: 2393-9907; e-ISSN: 2393-9915; Volume 6, Issue 2; April-June, 2019, pp. 86-88

http://www.krishisanskriti.org/Publication.html

Classification of Fetal Heart Rate and Thyroid Disease using Ensemble Learning Approach

Musavir Hassan[*], Muheet Ahmad Butt[*2] and Majid Zaman Baba[*3]

[*]*PG Department of Computer Science, University of Kashmir, Srinagar*
E-mail: musavirhassan.1319@gmail.com, ermuheet@gmail.com

Abstract—*Ensemble learning approach combines number of weak learners that are generated by using different algorithms, different parameters/hyper parameters (hidden layers), different representations and different training sets to improve the classification accuracy. In this paper the ensemble learning approach is used as a classification task to accurately predict fetal heart risks and thyroid disease. We have selected two standard datasets from UCI Repository and have performed extensive experimentation using random forest Classifier. In first dataset records are classified based on fetal state into Normal, Suspect and Pathologic State of fetus. Fetal heart rate and uterine contractions are recorded using cardiotocography usually during pregnancy. By observing the cardiotocography doctors can detect pathologies and understand the status of fetus. In another dataset a patient data is classified into normal or hyperthyroidism or hypothyroidism. Thyroid is an incessant and intricate infection happened because of unseemly TSH (Thyroid Simulating Hormone) levels or might be brought on by the issues in thyroid organ itself. Performance of the model is evaluated using accuracy, F-measure and Kappa statistics analysis. Experimental results reveal that the proposed model is efficient enough to classify the data.*

Keywords*: cardiotocography, random forest, accuracy, F-measure, Cohen's Kappa statistics*

1. INTRODUCTION

Thyroid is one of the major health challenges all over the world. The prevalence of thyroid is increasing at a fast pace, deteriorating human, economic and social fabric. It is the chronic and one of the dramatically increasing endocrine disorders in the world. It has been projected that about 200 million people suffer from thyroid disorders worldwide and amongst those 42 million are in India [1]. A recent survey conducted by Indian Thyroid Society depicts awareness for the disease ranked ninth as compared to other common ailments such as asthma, cholesterol problem, depression, diabetes, insomnia and heart problem. A study conducted by SRL diagnostics (2012-2014) in which 14,24,008 samples were studied, 22.68 per cent, of the total samples were found with abnormal TSH levels. The younger population within the age group of 3 - 45 years was at higher risk of thyroid dysfunction (30.33% of the samples) than the older population within the age group of 46-60 years (25.81% of the samples). The analysis showed the highest prevalence of the disease among men in the eastern zone of the country. East Zone had highest percentage of abnormality with 25.2% while northern and western zone had 23.9% and 21.1% respectively. Among the four zones, southern India showed the lowest percentage of abnormality with 19.4%. Prevention and prediction of thyroid is increasingly gaining interest in healthcare community. Although several clinical decision support systems have been proposed that incorporate several data mining techniques for thyroid prediction and course of progression. These conventional systems are typically based either just on a single classifier or a plain combination thereof. Recently extensive endeavors are being made for improving the accuracy of such systems using ensemble classifiers.

Cardiotocography is a medical test used to analyze the uterine contractions and fetal heart rate (FHR) in the third trimester of pregnancy. A standard nomenclature to read cardiotocographs [2] includes the uterine activity description, fetal heart rate, presence of deceleration or acceleration [3] and reduced or increased FHR variability. [4] developed a cardiotocographic ML algorithm, a random forest (RF) model, which accurately predicts neonatal pathologic disposition to inform obstetric intervention. In [5] Cardiotocography (CTG) is used as a technique of measuring fetal well-being. Different machine learning techniques were used for anticipating of fetal risks.

2. MATERIAL AND METHOD

In section, we describe the methodology and datasets used followed by the performance metrics used to evaluate the performance of the proposed system. This study systematically adopts random forest data mining technique for predictive data mining due to the impressive performance of random forest method. In [6][7] random forest method is evaluated in several statistical measures (sensitivity, specificity, accuracy, F-measure and ROC curve) and showed that the random forest method outperforms the other classifiers. It has been suggested that random forest is good in medical data classification. Random Forest is an ensemble method that uses randomization to produce a diverse pool of

individual classifiers. RFs combine the simplicity of decision trees with flexibility resulting in a vast improvement in accuracy. Attributes are selected randomly to determine the split for the decision tree formation. Most popular class is returned based on voting mechanism during the classification. Random forests are trained using bagging with random attribute selection. Given a training set D of d tuples. For developing an ensemble of classifiers k decision trees are generated. From training set D_i, d tuples is sampled with replacement for each iteration (I =1, 2, 3…..k). So we can say each D_i is a bootstrapped sample of D in which some tuples may occur more than once, while others may be excluded. To determine the split suppose S be the no. of attributes, where S is much smaller than the no. of available attributes. S attributes are randomly selected as candidates for split at node to construct a decision tree classifier M_i. Random forests are formed by CART methodology, Random input selection and by random linear combination of input attributes. Random forests have comparable accuracy to AdaBoost but are more robust to errors and outliers. As long as the number of trees in forest is large, the generalization error converges. Therefore over fitting is not a problem in random forest. Random forests are not sensitive to the no. of attributes selected for consideration at each split. Mostly $\log_2 d+1$ are chosen, because fewer attributes are selected for each split.

Datasets:

In our experiments, we use the Cardiotocography Dataset available at UCI machine learning repository (http://archive.ics.uci.edu/ml/datasets/cardiotocography/) that consists of measurements of fatal heart rate (FHR) and uterine contraction (UC) features on cardiotocograms classified by expert obstetricians. There are total of 2126 instances in the dataset with 22 attributes. Description of these attributes is available on the above given url address. Classification labels used in the dataset are Normal (represented by 1), Suspect (represented by 2), and Pathologic (represented by 3). And the thyroid dataset available at UCI machine learning repository (http://archive.ics.uci.edu/ml/datasets/thyroiddisease/). TDD is given by Garavan Institute, Australia by Ross Quinlin. There are total of 7200 instances in the dataset with 22 attributes. Classification labels used in the dataset are Normal (represented by 1), Hyperthyroidism (represented by 2), and hypothyroidism (represented by 3).

Performance Metrics

Accuracy, F-measure, and Cohen's Kappa statistics are used as performance measures to evaluate the performance of the proposed data mining model.

Accuracy (α) is calculated as the total number of correct predictions divided by the total number of instances in the dataset. Accuracy of a model is calculated using the formula given in equation 1.

$$\alpha = \frac{Number\ of\ instances\ correctly\ classified}{Total\ number\ of\ instances} \quad (1)$$

Accuracy **(α)** is a popular metric to evaluate the performance of a classifier and works well on balanced data. But, in case the data is imbalanced in the test dataset, accuracy may give misleading information regarding accuracy of models.

F-measure also called as F1-score or F-score is considered a more accurate way to measure

the performance of the classifier. F-measure is calculated.

$$F\text{- measure} = \frac{2(Precision\ X\ Recall)}{Precision + Recall} \quad (2)$$

Where

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

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

TP - True Positive; FP - False Positive; FN - False Negative

Therefore, F-measure is basically the harmonic mean of precision and recall and thus effective for handling the imbalanced data distribution in different classes. F-measure for each class is computed and F - measure of model is computed by averaging

the F-measures obtained for each class using equation 2.

Cohen's Kappa Coefficient (*k*) is an effective metric to evaluate the performance of a classifier and evaluate classifiers themselves as well. Measures like accuracy and precision/recall do not provide complete picture of performance of multi-class classifiers. *k* also performs comparatively better to evaluate performance of classifiers in case of imbalanced classes. *k* is evaluated using equation 2.

$$k = \frac{(Observed\ Accuracy - Expected\ Accuracy)}{(1 - Expected\ Accuracy)}$$

3. RESULTS, EVALUATION AND DISCUSSION

The model is trained using the training data obtained from datasets Thyroid and CTG respectively. Each dataset is partitioned into training and test datasets with probability of 0.7 and 0.3 respectively. Thyroid dataset used to train the model have 7200 instances, out of which 5040 are used for training and remaining 2160 are in the test dataset. The CTG dataset model have 2126 instances, out of which 1488 are used for training and remaining638 are in the test dataset. Accuracy, F-measure and Cohen's Kappa of model is obtained using the random forest configuration settings as follows. Number of trees, t = 300 in the random forest model and number of variables tried at each split is equal to 8. The model achieves an accuracy 99.7% for thyroid dataset and 93.1 for CTG dataset and the confusion matrix for model is given in table 1 below :

Table 1: Confusion Matrix of model obtained using test dataset from Thyroid dataset

	1	2	3
1	49	0	3
2	0	127	3
3	0	0	1958

It can be observed that there are 7200 test instances that are input to model with 166 instances belonging to class 1, 368 instances in class 2 and 6666 instances in class 3.

Table 2: Confusion Matrix of model obtained using test dataset from CTG dataset

	1	2	3
1	480	19	3
2	13	59	2
3	2	4	49

It can be observed that there are 2126 test instances that are input to model with 1655 instances belonging to class 1, 295 instances in class 2 and only 176 instances in class 3. The accuracy, F-measure and *k* value for model is given in Table 6.

Table 3: Performance measurement of model

Dataset	Accuracy	F measure	Cohens Kappa
Thyroid	0.9972	0.9619	0.9820
CTG	0.9319	0.8733	0.8071

Two important parameters that affect the performance of local random forest models are number of decision trees (t) in the random forest and number of attributes considered at each node of the decision tree.

In order to analyze the effect of number of trees on error rate, all the local models are trained with number of trees, t = 500. The plot of out of bag (OOB) error rate along with the error rates for each class at varying number of trees for model is given in Figure1. It can observed from the figure that OOB error rate and error rates for each independent class start to stabilize with t >= 300. Therefore, 300 trees for the model case is sufficient because more number of trees do not enhance the performance of these model further.

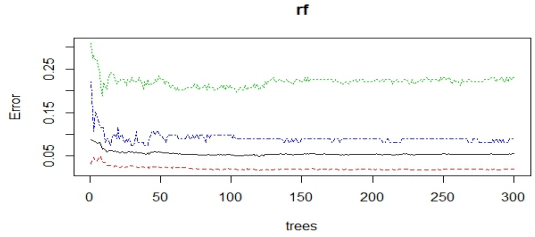

Figure 1: Error rate of model

Conclusion:

In this paper, a random forest based data mining technique is adopted to mine the sensitive health data maintained with different healthcare facilities without revealing any patient specific information in the process. Evaluation of results indicates that random forest ensemble method outperforms than bagging as well as boosting methods. In future, similar ensemble approaches can be applied on other disease datasets such as diabetes, hypertension, coronary heart disease and dementia. Moreover, different individual techniques like Naïve Bayes, SVM and neural networks etc. can be incorporated as base learners in ensemble framework.

References:

[1] Available from: http:// www.ias.ac.in /currsci/oct252000/n%20kochupillai. PDF [Last accessed on 2011 April 2].

[2] G. A. Macones, G. D. V. Hankins, C. Y. Spong, J. Hauth, and T. Moore, "The 2008 National Institute of Child Health and Human Development Workshop Report on Electronic Fetal Monitoring: Update on Definitions, Interpretation, and Research Guidelines," Journal of Obstetric, Gynecologic & Neonatal Nursing, vol. 37, no. 5, pp. 510–515, Sep. 2008.

[3] A. Ugwumadu, "Understanding cardiotocographic patterns associated with intrapartum fetal hypoxia and neurologic injury," Best Practice & Research Clinical Obstetrics & Gynaecology, vol. 27, no. 4, pp. 509–536, Aug. 2013.

[4] I. G. Freedman, A. Agarwal, N. Doilicho, E. Cohen, C. M. Pettker, and J. A. Copel, "Predicting Neonatal Pathologic Disposition From Cardiotocography Using Machine Learning [25A]," Obstetrics & Gynecology, vol. 133, p. 16S, May 2019.

[5] H. Sahin and A. Subasi, "Classification of the cardiotocogram data for anticipation of fetal risks using machine learning techniques," Applied Soft Computing, vol. 33, pp. 231–238, Aug. 2015.

[6] Z. Masetic and A. Subasi, "Congestive heart failure detection using random forest classifier," Computer Methods and Programs in Biomedicine, vol. 130, pp. 54–64, Jul. 2016.

[7] A. Subasi, E. Alickovic, and J. Kevric, "Diagnosis of Chronic Kidney Disease by Using Random Forest," CMBEBIH 2017, pp. 589–594, 2017.