IJSRD - International Journal for Scientific Research & Development| Vol. 7, Issue 01, 2019 | ISSN (online): 2321-0613

# Effective Diagnostic System of Fetal Heart Disease Based on Cardiotocography using AdaBoost

**D.Rubiya Sweetlin[1] Dr.I.Muthulakshmi[2]**
[1]PG Scholar [2]Professor
[1,2]Department of Computer Science & Engineering
[1,2]VV College of Engineering, India

*Abstract*— Fetal health is commonly surveilled by cardiotocography (CTG) monitor, which documents fetal heart rate (FHR) signal. The fetal health monitoring is very important to predict the growth of the fetal as well as improvement in its each stage. For yielding the status of the fetal more accurately, fetal heart rate (FHR) signal is the best well known source. From this gained information, embryo’s unhealthiness can be visualized and it provides a chance to take action before an irremediable harmness happening to the fetal. Previous work uses Naïve Bayes and SVM classifier, which was used in the olden days, to classify the CTG data, produced low accuracy. To improve the classification accuracy, the proposed work addresses Adaptive Boosting algorithm for classification of congenital heart disease in fetus. Fetal state is commonly grouped into normal and pathological class. The results of these classifiers are to be compared and the best classification technique is to be identified.

***Key words:*** AdaBoost, Cardiotocography, FHR, Machine Learning, Naïve Bayes, SVM

## I. Introduction

In the recent days, computers can be used in many sectors and it is an unavoidable part of human life. Computers have gained maximum significance since they have improved the efficiency and productivity.

Communication technologies and information stored in computers widely used in the healthcare sector. Also, the computers have been employed in the administrative department of the healthcare sector to give efficient services to the patients. Computers are used to store information regarding the patients, which help doctors to help patients more effective.

### A. *Machine Learning on Medical:*

In healthcare industry, massive amounts of complex data about patient records, hospitals contrivance, disease anticipation, medical devices etc, are generated. This data is the key source for knowledge extraction which enables decision making.

Machine Learning techniques are applied to this processed data to identify hidden patterns which provide additional source of knowledge for decisions making. These decisions are handled by healthcare professionals. Machine learning algorithms are used to filtering most relevant information from the dataset.

### B. *Cardiotocography (CTG):*

For retaining the new born from birth ailment that results in death or lifelong blight, the CTG screen is utilized to store fetal pulse (FHR) ceaselessly. By scrutinizing cardiotocograms, the non-reassuring fetal state can be determined in a computerized way. But most of the techniques consume more time for determining the fetal state and so forth many new machine learning techniques are needed to make it less time consuming.

The paper has been composed as pursues. In section II surveyed paper is discussed. In section III, proposed methodology and architecture is explained. In section IV, comparative result is analyzed. Finally, in section V, the conclusion and future work is displayed.

## II. Literature review

N. Chamidah and I. Wasito presented a fetal state classification system to detect the heart disease in the fetus [1]. For evaluating the heart disease condition in fetus, CTG data were examined. In this research, K-Means algorithm was used to extract useful information from the CTG data. Using SVM, the cardiotocograms are skilled. Before classifying these CTG samples, it has been split into training and testing samples which improves the classification accuracy to 90.64%. Hybrid K-SVM was shown the excellent results while comparing it with the SVM.

S. A. A. Shah, W. Aziz, M. Arif, and M. S. A. Nadeem presented a decision tree based classification system which utilizes a new ensemble method called bagging approach with the combination with three conventional decision tree algorithms [2]. The proposed system generated a better classification results.

S.Mazumdar, R.Choudhary and A. Swetapadma were proposed an ANN based method for fetal heart rate monitoring [3]. Cardiotocography data can be classified through Artificial Neural Network (ANN) for predicting the fetal state. The classification performance can be reached to 99.9% using ANN.

M. Huang and Y. Hsu proposed a fetal distress prediction using discriminant analysis, decision tree, and artificial neural network to evaluate fetal distress [4]. If the fetal lacks oxygen in uterine, fetal death could happen. The DA, DT and ANN’s resulted accuracies are 79.1%, 82.36% and 92.78%, respectively.

H. Ocak and H. M. Ertunc presented adaptive neurofuzzy inference system (ANFIS) [5]. An ANFIS system predicted the state of fetal by classifying the CTG data. The system was tested using cardiotocograms that consist of 1,831 CTG recordings. It yielded 95.2% accuracy.

The surveyed paper of fetal heart disease prediction system is listed in table 1.

 

| TOPIC | ALGORITHM | ADVANTAGES | DISADVANTAGES |
|---|---|---|---|
| Fetal state classification from cardiotocography based on feature extraction using hybrid K-Means and support vector machine | K-Means SVM | Feature Extraction can improve the accuracy (High accuracy) | Missing value problem |
| Decision Trees based classification of cardiotocogram using Bagging Approach | Decision Tree Bagging Approach | Best Feature Selection Easy classification without any computation | Poor performance due to overfitting problem Prone to sampling error |
| An Innovative method for fetal health monitoring based on Artificial Neural Network using cardiotocography measurements | Artificial Neural Network | Better Performance | Overfitting Problem |
| Fetal distress prediction using discriminant analysis, decision tree, and artificial neural network | Discriminant Analysis Decision Tree Artificial Neural Network | Easy Classification | Time Consuming Missing value problem |
| Prediction of fetal state from the cardiotocogram recordings using Adaptive Neuro-Fuzzy Inference Systems | Adaptive Neuro-Fuzzy Inference Systems | High Accuracy Speed | Missing value problem |

Table 1: Literature Survey

## III. METHODOLOGY

### A. *System Overview*

Regarding the fetal pulse, the cardiotocograms contains many attributes and it is obtained from the UCI Machine Learning Repository. Cardiotocograms are partitioned into training and testing data using cross validation (CV) technique. Cross validation procedures are used to analyze the performance of various machine learning techniques. CTG data are trained and tested using Naïve Bayes, Support Vector Machine (SVM) and AdaBoost classification algorithm. Finally these classifiers outcomes are compared.

The system architecture is shown in the Fig 1. The architecture comprises of CTG Data, Cross validation, principal component Analysis and classification (Naïve Bayes, Support vector machine (SVM) and AdaBoost).

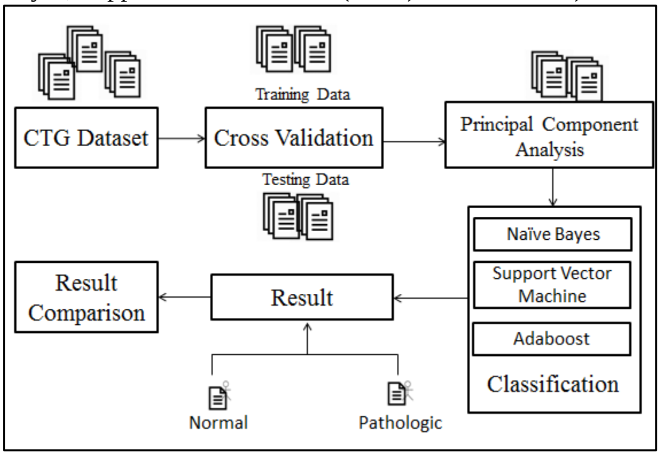


Fig. 1: Work flow of Proposed System

### B. *Cross Validation*

Before classifying the given samples, the cardiotocograms are partitioned into training and testing data using cross validation (CV) technique. For evaluating and comparing the performance of various machine learning techniques, cross-validation (CV) becomes a powerful choice. [4].

### C. *Principal Component Analysis*

In a preprocessing phase before classification in ML, Dimensionality reduction has played an important role and become greatly useful method to improve classification performance and speed up computation through detecting correlation between input features. Reducing number of features in the given dataset is known as dimensionality reduction. For this purpose, the proposed system uses Principal Component Analysis ((PCA).

Principal Component Analysis (PCA) is a way to perform the dimensionality reduction by linear combination of the original features in such a way that preserves as much of the relevant information as possible. This method computes eigenvalues of the correlation matrix o the input data vector and then projects the data orthogonally onto the subspaces spanned by the eigenvectors (principal component (PC)) corresponding to the dominant eigenvalues. Even if the whole set of eigenvectors is retained, this may also lead to an improvement of the classification performance, because the new set has features that are uncorrelated and this improves the classification capabilities of a classifier.

### D. *Classification*

Building a model from the given input data for classifying the data are called as Classifiers. The main objective of any classifier is to build a new model from the input data which can predict the class labels. The famous classifiers are support vector machines and naive bayes classifiers used for classification.

The classification performance model is computed based on the counts of number of correctly and incorrectly predicted classes which are tabulated in a ‘confusion matrix’. A confusion matrix of a classification problem is given in Table 2.

 

| | | Predicted Class | |
|---|---|---|---|
| | | Class 'X' | Class 'Y' |
| Actual Class | Class 'X' | $d_{xx}$ | $d_{xy}$ |
| | Class 'Y' | $d_{yx}$ | $d_{yy}$ |

Table 2 Confusion Matrix

### E. *Naïve Bayes Algorithm*

The Naive Bayes classifier is one of the probabilistic approaches which classify the data using Bayes' theorem with independent considerations between predictors.

Bayes theorem characterizes the probability of event based on prior knowledge of conditions which might be related to the event. It is computed as

$$P(m\ /\ n) = \frac{P(n\ /\ m) \times P(m)}{P(n)}$$

$$P(m/N) = P(n_1 \vee m) \times P(n_2 \vee m) \times P(n_3 \vee m) \ldots \ldots \ldots \ldots \times P(n_n \vee m) \times P(m)$$

$P(m/n)$ is the posterior probability of class (target) given predictor (attribute).

$P(m)$ is the prior probability of class.

$P(n\ /\ m)$ is the likelihood which is the probability of predictor given class.

$P(N)$ is the prior probability of predictor.

### F. *Support Vector Machine (SVM)*

For solving the problem of classification, Support Vector Machine (SVM) can be used. One of the best known supervised learning techniques is SVM. Each and every data points are plotted in dimensional space of size n. Then, it groups the data points by determining the hyper-plane with the maximum margin. The data points around the corner of hyperplane are called as support vectors. The distance between hyper-plane and the support vectors are called margin.

The idea of using SVM is rearranging the objects using a set of mathematical relations. The process of rearranging is called as 'mapping' and the rearranged space is called as 'feature space'.

### G. *Adaptive Boosting Algorithm*

The Adaptive Boosting Algorithm (AdaBoost) is a general ensemble method that creates a strong classifier from a number of weak classifier. The survey shows that AdaBoost algorithm is a wise choice for many applications.

AdaBoost uses only one classification algorithm to construct diverse weak base classifiers, which are trained on the data sets selectively sampled from an initial training dataset. Each sample is given a weight representing its possibility to be selected as a training sample and all examples share an equal weight in the first round of iteration. In the next iteration, the samples correctly classified by the base classifier of the last iteration will get lower weights and the samples misclassified will get higher weights. Thus the base classifier will focus more and more on the "difficult" data points that may locate near the classification margin and finally improve the classification. After certain number of iterations, the base weak classifiers are combined to be a strong classifier and the classification result is output.

## IV. Results & Discussion

### A. *Performance Evaluation*

To compare the classification performance of various classification models, there are many performance metrics are developed. One of the important metrics is accuracy. The following parameters are used to calculate the accuracy of the proposed scheme:

1) The *recall* or *true positive rate* (*TP*) is the proportion of positive cases that were correctly identified
2) The *false positive rate* (*FP*) is the proportion of negatives cases that were incorrectly classified as positive
3) The *true negative rate*(*TN*) is defined as the proportion of negatives cases that were classified correctly
4) The *false negative rate*(*FN*) is the proportion of positives cases that were incorrectly classified as negative
5) The *accuracy (AC)* is the proportion of the total number of predictions that were correct which can be calculated by the following formula.

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

| METHOD | ACCURACY | PRECISION |
|---|---|---|
| Naive Bayes | 55.9878 | 56.8756 |
| Support Vector Machine | 87.9999 | 75.4532 |
| AdaBoost | 99.9998 | 98.5673 |

Table 3: Performance analysis of SVM, Naive Bayes and AdaBoost Classifiers using complete features set

The results of experiments have been tabulated in table 3 which shows that the adaptive boosting algorithm performs better in classifying CTG data in terms of accuracy and other performance metrics. The performance measures listed in Table 3 have been shown graphically in Figure 2 for improved illustration.

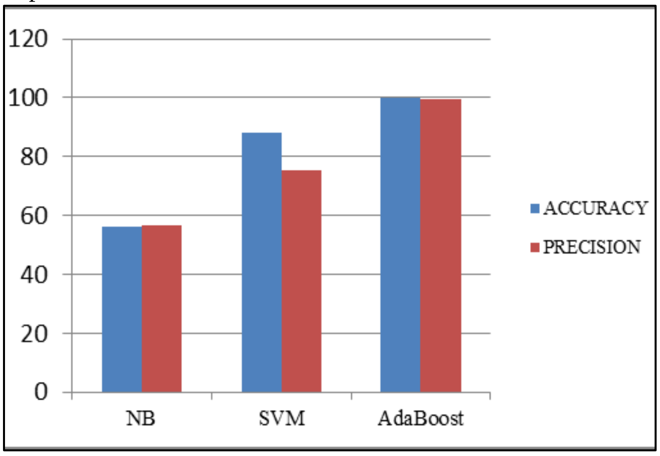


Fig. 2: Graphical representation of performance analysis of SVM, NB and AdaBoost Classifiers for complete features set

The comparison is performed on the following algorithms: Naïve Bayes, Support Vector Machine (SVM) and Adaptive Boosting Algorithm. By analyzing the experimental results, it shows that the AdaBoost classifier outperformed the other algorithm in both accuracy and other metrics. That is, AdaBoost classifier turned out to be the best classifier for the prediction of fetal heart disease.

| METHOD | ACCURACY | PRECISION |
|---|---|---|
| Naive Bayes | 65.9878 | 56.8756 |
| Support Vector Machine | 78.9999 | 83.4532 |
| AdaBoost | 96.9998 | 97.5673 |

Table 4: Performance analysis of SVM, Naive Bayes and AdaBoost Classifiers using reduced features set

Table 4 shows the performance analysis of three classifiers results using reduced features set (PCA). The performance measures listed in Table 4 have been shown graphically in Figure 3 for improved illustration.

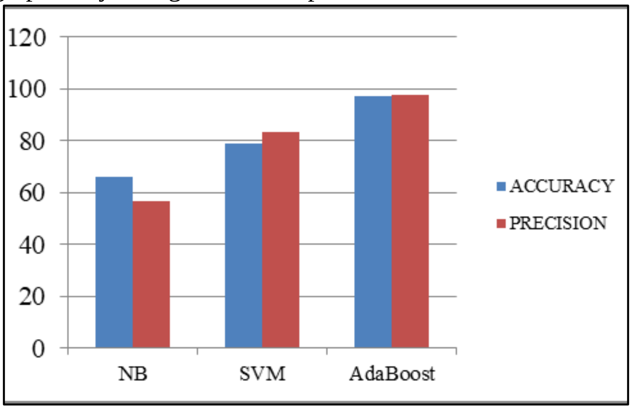


Fig. 2: Graphical representation of performance analysis of SVM, NB and AdaBoost Classifiers for reduced features set

## V. Conclusion & Future Enhancement

For obstetricians those who, experts to identify the fetal condition, CTG recordings are well helped. And it provides better medical treatment before happen any undesirable heart disease on baby. The automated analysis of CTG recordings detects prenatal pathologies in a computerized way. Several machine learning methodologies are employed to classify CTG signals.

In this study, Naive Bayes, Support Vector Machine (SVM) and Adaptive Boosting Algorithms are used to predict the state of fetal. By analyzing the experimental results, it shows that AdaBoost classifier turned out to be a best classifier.

The future study will tend to improve performance efficiency by applying other machine learning techniques. In addition, the feature selection technique can also be considered in data preprocessing stage.

## References

[1] N. Chamidah and I. Wasito, "Fetal state classification from cardiotocography based on feature extraction using hybrid K-Means and Support Vector Machine", in International Conference on Advanced Computer Science and Information Systems, ICACSIS 2015, Depok, Indonesia, pp. 37-41, Oct 2010.

[2] P.A. P.A.Warrick, R. E. Kearney, D. Precup and E. F. Hamilton, “System identification noise suppression for intra-partum cardiotocography to discriminate normal and hypoxic fetuses”, IEEE Computers in Cardiology, 2008

[3] S. A. A. Shah, W. Aziz, M. Arif, and M. S. A. Nadeem, "Decision Trees Based Classification of Cardiotocograms Using Bagging Approach," in 13th International Conference on Frontiers of Information Technology, Islamabad, Pakistan, pp. 12-17, 2015.

[4] S.Mazumdar, R.Choudhary and A. Swetapadma, “An innovative method for fetal health monitoring based on artificial neural network using cardiotocography measurements”, IEEE Transaction on Neural Network, pp. 265 – 268, Dec 2017.

[5] M. Huang and Y. Hsu, "Fetal distress prediction using discriminant analysis, decision tree, and artificial neural network," Journal of Biomedical Science and Engineering, vol. 05, pp. 526-533, 2012.

[6] H. Ocak and H. M. Ertunc, "Prediction of fetal state from the cardiotocogram recordings using adaptive neuro-fuzzy inference systems," Neural Computing and Applications, vol. 23, pp. 1583-1589, 2013.

[7] G. Georgoulas, D. Stylios and P. Groumpos, “Predicting the risk of metabolic acidosis for new born based on fetal heart rate signal classification using support vector machines”, IEEE Transaction in Biomedical Eng. 53, pp. 875–884, 2006.

 