Fetal Risk Prediction Using Optimized Genetic Algorithm - Support Vector Machine Based Feature Selection Techniques

J. Jayashree[1], J. Vijayashree[2], N.Ch.S.N. Iyengar[3*]

[1,2]School of Computer Science and Engineering, Vellore Institute of Technology, Vellore, India

[3]Department of Information Technology, SNIST, Hyderabad, Telangana, India

E-mail: [1]jayashree.j@vit.ac.in, [2]vijayashree.j@vit.ac.in, *and* [3]srimannarayanach@sreenidhi.edu.in

Abstract— Improved feature selection methodology for fetal risk data collection defining important features. The aim is to improve the fetal risk prediction rate by using an optimized technique such as GA-SVM for feature selection. Then the selected features are given to various classifiers such as random forest, naïve bayes, multi- layer perceptron and support vector machine for prediction. As a result, the feature selected by optimized feature selection techniques provides higher accuracy, precision and recall when compared to non-optimized techniques.

Keywords— Fetal, optimization, features, prediction.

I. INTRODUCTION

About 25 babies out of 1000 births are affected by fetal asphyxia associated with metabolic acidosis, children's death in the first month of life is a significant contributor to underfive mortality. Intrapartum complications are one of the principal causes of perinatal mortality. Fetal cardiotocography (CTGs) can be used as a screening device to detect high-risk women during childbirth.

Around 795 women die from preventable causes of death Pregnancy and childbirth, 98 percent of which are infantile Developing pays. Maternal deaths widespread 84.3 per cent decrease after 1990, The ratio of global maternal mortality (no of deaths per 100090 Live birth) dropped by just 2,450 years per annum in 1989. And by 2015 gynaecologists say any the trimester is about 98 days, which add to a bout 41 weeks pregnant. About 21 million people every year, ill health are endured worldwide by Pregnant. In 1990, the impact was about 376,000 women who died of complication in pregnancy, which fell to 293,000 in 2013. Of those 288,000 people died during the childbirth process and most of the deaths where settings could have been avoided in low facilities and most of them.

A. Fetal heart rate

The General FHR fluctuates from 120-155 BPM during the utero cycle. Sonographically it is observable from about 6 weeks and the range varies during development, increasing to about 170 bpm at 70 days, and then to about 130 bpm at term. While myocardium starts to contract 20 days after giving birth (myocardial pace maker cells in the embryonic heart starts depolarizing), it is first visible after 6 weeks of sonography gestation. The FHR usually then beats 100 to 120 per minute (bpm). While heart rhythm is usually common in the normal fetus, a beat to beat variance of approximately 5 to 15 beats per minute is required.

Table 1: *Types of Fetal Diseases.*

	TYPE-1	TYPE-2	TYPE-3
Name	Fetal Bradycardia (Low heart rate in the fetus).	Fetal Tachycardia (Increase in heart rate in the fetal).	Fetal Bradyarrhythmia (Low heart rate of the fetus)
Causes	• poor uterine perfusion • maternal hypotension • umbilical cord prolapse • rapid fetal descent	• maternal fever • Dehydration • maternal ketosis • preterm fetus • maternal thyrotoxicosis	NILL
Symptoms	• An abnormally fast heart rate. • Abrupt decreases in heart rate. • Late returns to the baseline heart rate after a contraction.	• An abnormally slow heart rate	NILL

Heart rate	below 99 beats per minute (bpm)	> 158-176 beats per minute (bpm)	< 99-109 beats per minute.

II. RELATED WORK

1. The studies have favored an ANN method for predicting hypertension disease. The high computing costs of ANN and the long learning speeds drive the extension of the idea to a deep learnings network. Therefore, low computational cost and classification of learning rates are important for such an issue. To improve fetal risk predictions hybrid support vector machine response approach were addressed with other method for reductions of attributes (Subha.V et al. 2017).
2. This paper simultaneously studies the enhancement of data classification accuracy for Cardiotocogram in the collection of functions and classifiers based on ensemble learning. Two subset filtering techniques (Correlation-based Feature Collection; Consistency-based Filter) and two filter-based feature ranking techniques (Relief; Information Gain) are used in the collection of features, while the Help Vector Machine ranking technique is used in the classification (Silwattananusarn et al.2016).
3. The results performed significantly better than previous work [20] compared the predictive accuracy of normal & pathological classification (99.68%) and were 99.1% accurate using Random Forest (Nagendra.et al.2017).
4. Confusion matrix classification with minimal misclassification error (0.183383) using pruned decision tree to evaluate cardiotocogram data to assess fetal distres (Permanasari. et al.2017).
5. CTG monitoring allows obstetrician to recognise fetal conditions and to decide on medical intervention during pregnancy and childbirth before serious damage to the baby (Zhang. et al.2017).
6. Once validated, future uses for this method may include the use of lay physicians and nurse for remote critical treatment of pregnant women at high risk of extreme perinatal outcome based on CTG test, for clarification and further management. (Hoodbhoy. et al.2019).
7. The research work analyzes the same data and finds that ANN achieve 92.42 percent accuracy. The findings obtained in the paper from the CTG classification indicate that the most reliable findings are obtained from the decision tree-based algorithm (J48) with 0.0407 as MAE, 0.8706 as kappa statistics and 94.23% as reliable with the highest accuracy metric value. The Random Forest and Regression classification was near J48 (Bhatnagar. et al.2016).
8. Pre-processing of fetal heart monitoring data, sequential removal of fetal heart rate data for data set removal with a high proportion of deletions, back-to-back missing values, linear interpolation, smooth denotation and then data structure adjustment to obtain three different data formats (Tang et al.2018).
9. However, another author argued the value of hybrid data mining algorithms to a pregnant women's construct model, prevention of health risk induced by parameter inconsistencies during pregnancy. The C4.5 algorithm delivered 98.1 % accurate output (Lakshmi. et al.2016).
10. Predicting fetal growth during pregnancy is the correct data set consisting of the correct number of parameters and the implementation of the hybrid approaches. Similarly, in other experiments, eight machine learning algorithm were recorded through weka tools over the CTG datasets. For validation the exact prediction responses of all algorithms was analyzed by partitioning the datasets into ten equal scale. The classifier model's results, the highest accurate classification was rated at 99.1 per cent. Feedforward ANN solved NN drawbacks with non-linear functions that make up a range of weighing input, hidden layers followed by the initiation function, a bias that provide output for the next layers (Kalyani. et al.2018).

III. PROPOSED WORK

The proposed work uses genetic algorithm and SVM based genetic algorithm for feature selection. Features selected by both the methods were given to classifiers such as random forest, naïve bayes, multi-layer perceptron and support vector machine. for fetal risk prediction. The proposed method uses an optimized method for improving the prediction accuracy. The proposed architecture is shown in fig1.

3.1 Data Set Information:

2126 fetal cardiotocograms (CTGs) were processed automatically and the respective diagnostic properties calculated. Three expert obstetricians also classified the CTGs, and each was assigned a consensus classification label. Classification concerned both a morphological pattern (A, B, C. ...) and a fetal state (N, S, P). Thus the dataset can be used for experiments of either 10 or 3 classes.

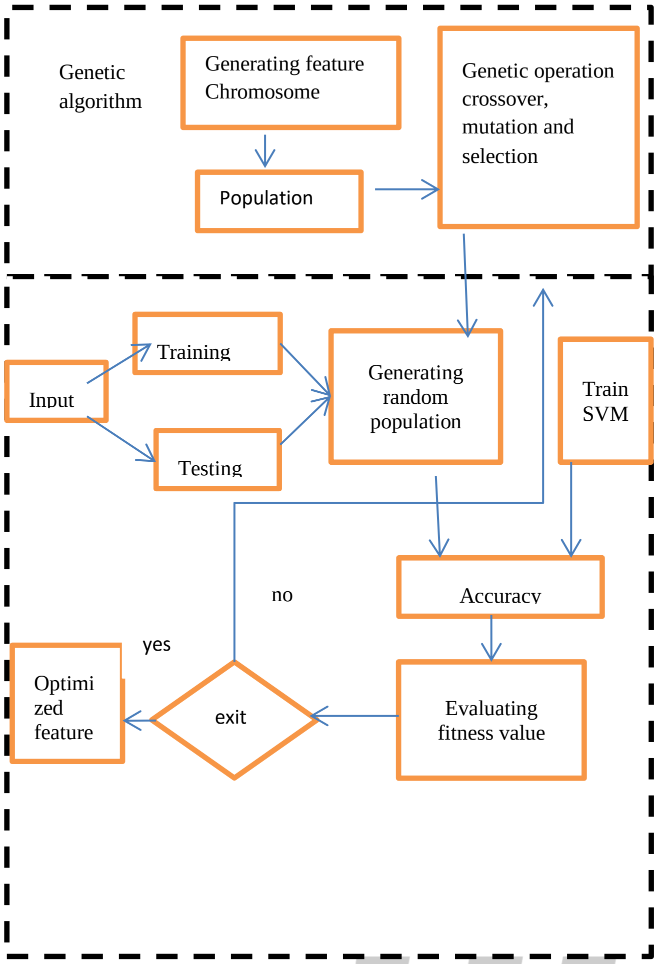

Figure 1: *Proposed Architecture*

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 prolongued 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
CLASS - FHR pattern class code (1 to 10)
NSP - fetal state class code (N=normal; S=suspect; P=pathologic)

3.2 Random Forest(RF):

Yet one thing to remember is that constructing the forest is not the same as setting up the decision using the information gain or index gain strategy. Random forests which are an system of ensemble learning for classification, regression & the other activities that functions by constructing a mixture of decision tree during training time & outputting class which would be the class mode (classification) (or) regression of individual trees and the

Random decision-making forest are perfect for decision trees' ability to overfit to their training set.

3.3 Support Vector Machine(SVM):

The aim of (SVM)support vector machine algorithm is to find a hyper-plane in an N-dimensional space which separately classifies the distinct data points.

SVM's are supervised learning models in machine learning which are associated learning algorithms which evaluate data which are used for the classification & regression analysis. Based on a collection of training instances, each of which is classified as belonging to either one or both of two categories, a Svm classification algorithm creates model which assigns the training examples to one or both category, rendering it a non -probability binary linear classification (though methods like Platt scaling occur to use Support vector machine in classification). The Support vector machine model is a description of the instances as spatial points, mapped in such a way as to split the examples of the different groups by as large a simple distance as possible.

3.4 Navies Bayes(NB):

Naive bayes classifiers are the family of basic "probabilistic classifiers" regarding the interpretation of bayes with clear (naive) assumptions of the characteristics of freedom.

Naive Bayes is a simple that construct & particularly useful for extremely big data sets and Naive- Bayes is considered to out perform also extremely advanced methods of classification. Naive Bayes Theorem gives a way for P(c), P(x) and P(x) to measure posterior likelihood. Look at the equation underneath:

$$\underbrace{P(c\mid x)}_{\text{Posterior Probability}}=\frac{\overbrace{P(x\mid c)}^{\text{Likelihood}}\,\overbrace{P(c)}^{\text{Class Prior Probability}}}{\underbrace{P(x)}_{\text{Predictor Prior Probability}}}$$

$$P(c\mid X) = P(x_1\mid c)\times P(x_2\mid c)\times\cdots\times P(x_n\mid c)\times P(c)$$

3.5 Multi-Layer Perceptron:

This is composed of at least 3 node layers which are the input layer, output layer and a hidden layer. In the exception of input nodes, each node becomes a neuron using a nonlinear activation mechanism. For teaching, MLP uses a supervised method of learning called backpropagation.

3.6 Genetic Algorithm:

Genetic algorithms seem to be useful for searching very general spaces and optimization problems. Each solution generated in Genetic algorithms is called a chromosome (individual). Each chromosome is made up of genes, which are the individual elements (alleles) that represents the problem. The collection of chromosomes is called a population. The internal representation of the chromosomes is known as its genotype.

IV. RESULTS AND DISCUSSION

The result of the proposed work should be evaluated using the following metrics: precision, accuracy and recall.

$$\textbf{Accuracy} = \frac{\textit{Number of instances classifed correctly}}{\textit{total number of instances}}$$

$$\textbf{Precision} = \frac{\textit{True Positive}}{(\textit{True Positive} + \textit{False Positive})}$$

$$\textbf{Recall} = \frac{\textit{True Positive}}{(\textit{True Positive} + \textit{Falsee Negative})}$$

Table 2: *Comparison of accuracy of classifiers using GA and GA-SVM based feature selection*

	Accuracy (%)	
Classifiers	**GA based Feature selection**	**GA-SVM based feature selection**
RF	89.6	93.1
NB	71.5	78.4
MLP	83.8	88
SVM	91.4	95.8

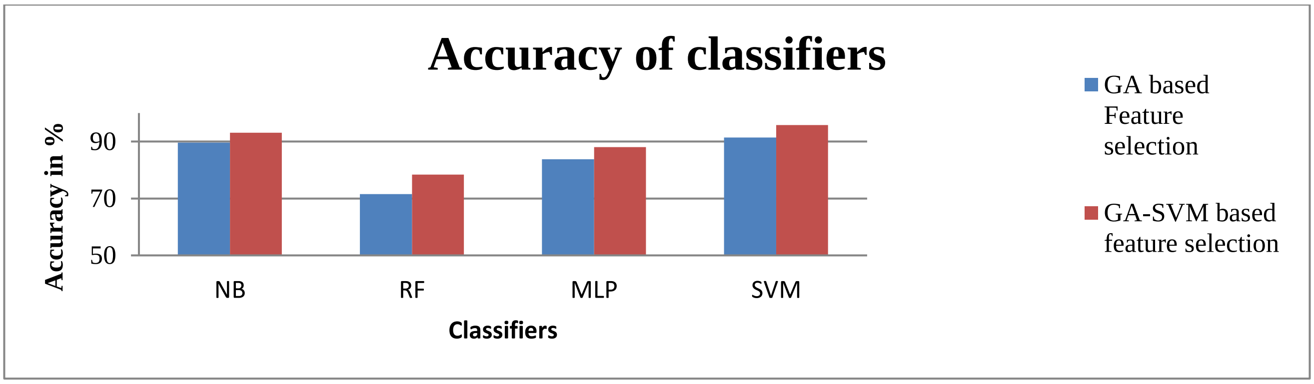

Figure 2: *Accuracy of classifiers*

Table 2: *Comparison of precision and recall of classifiers using GA and GA-SVM based feature selection*

Classifiers	Precision in %		Recall in %	
	GA based Feature selection	GA-SVM based feature selection	GA based Feature selection	GA-SVM based feature selection
RF	87.2	92.6	86.8	92.4
NB	72	77.5	72.7	78.1
MLP	84.4	87.8	81.3	88.4
SVM	91	94.9	89.7	95.1

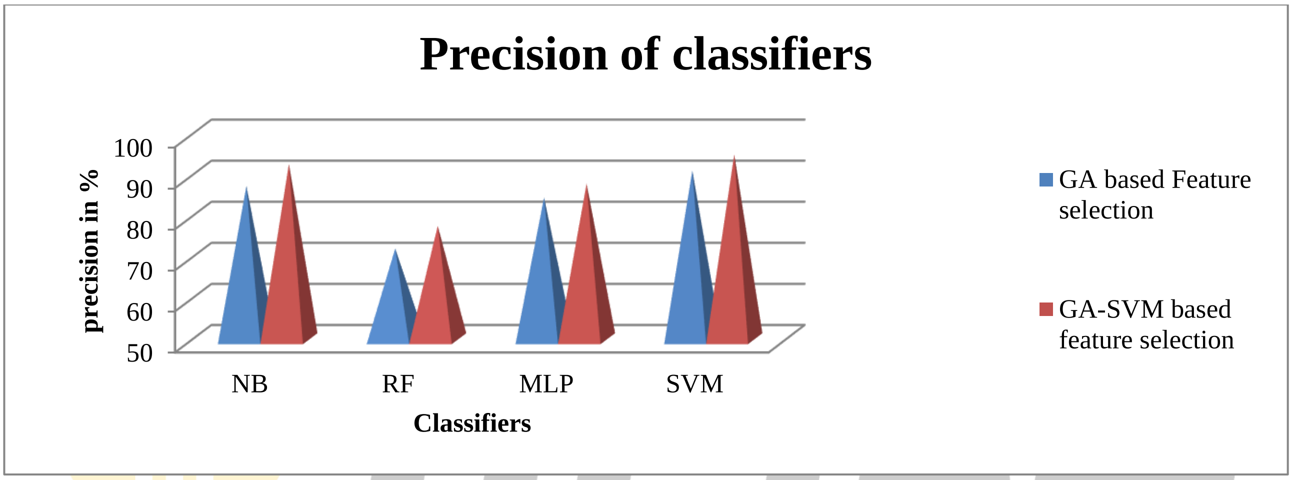

Figure 3: *Comparison of precision of classifiers*

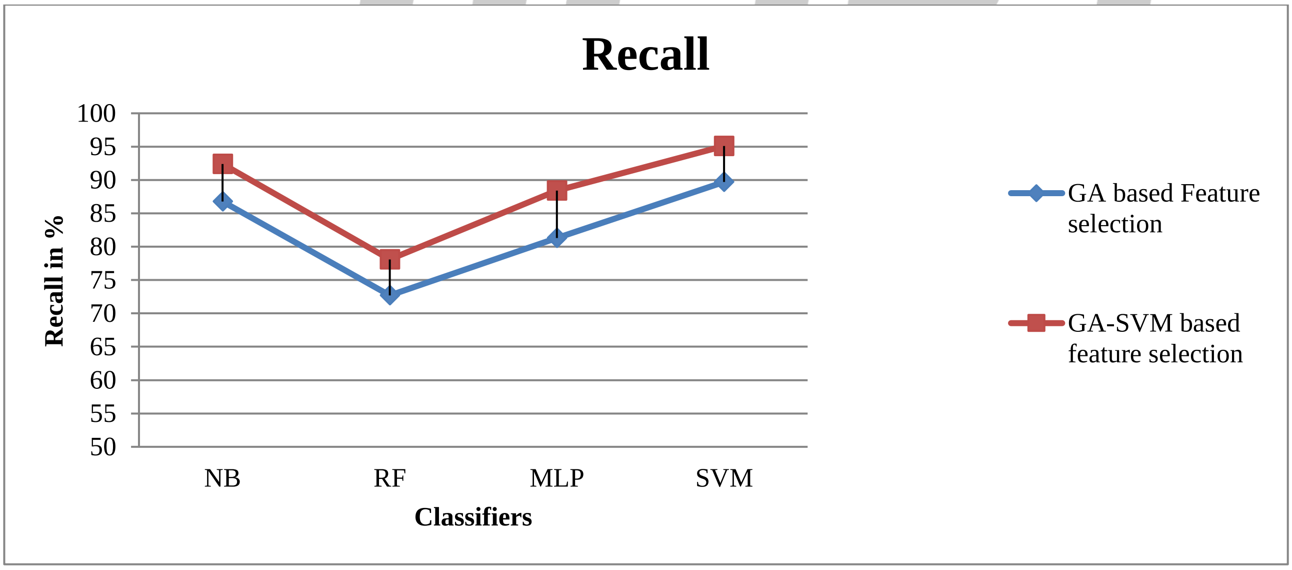

Figure 4: *Comparison of recall of classifiers*

V. CONCLUSION AND SUMMARY

In this research work, we analyzed the automatic prediction of fetal heart rate disease prediction from UCI fetal data set to improve the overall disease prediction accuracy. During the fetal heart rate data recognition process, fetal data is collected from patient using by optimized machine learning approaches. From the results it is evident that the features selected by optimized GA-SVM gives higher accuracy when compared to the features selected by GA. Also support vector machine classifier outperforms naïve bayes, random forest and multi-layer perceptron techniques.

REFERENCES

[1] Subha, V., D. Murugan, and A. Manivanna Boopathi. "A hybrid filter-wrapper attribute reduction approach for fetal risk anticipation."

Asian Journal of Research in Social Sciences and Humanities 7.2 (2017): 1094-1106.
[2] Silwattananusarn, Tipawan, Wanida Kanarkard, and Kulthida Tuamsuk. "Enhanced classification accuracy for cardiotocogram data with ensemble feature selection and classifier ensemble." Journal of Computer and Communications 4.4 (2016): 20-35.
[3] Nagendra, Vinayaka, et al. "Evaluation of support vector machines and random forest classifiers in a real-time fetal monitoring system based on cardiotocography data." 2017 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB). IEEE, 2017.
[4] Permanasari, Adhistya Erna, and Akhsin Nurlayli. "Decision tree to analyze the cardiotocogram data for fetal distress determination." 2017 International Conference on Sustainable Information Engineering and Technology (SIET). IEEE, 2017.
[5] Zhang, Yang, and Zhidong Zhao. "Fetal state assessment based on cardiotocography parameters using PCA and AdaBoost." 2017 10th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI). IEEE, 2017.
[6] Hoodbhoy, Zahra, et al. "Use of machine learning algorithms for prediction of fetal risk using cardiotocographic data." International Journal of Applied and Basic Medical Research 9.4 (2019): 226.
[7] Bhatnagar, Divya, and Piyush Maheshwari. "Classification of cardiotocography data with WEKA." International Journal of Computer Science and Network-IJCSN 5.2 (2016).
[8] Tang, Haijing, et al. "The Design and Implementation of Cardiotocography Signals Classification Algorithm Based on Neural Network." Computational and mathematical methods in medicine 2018 (2018).
[9] Huang, Mei-Ling, and Yung-Yan Hsu. "Fetal distress prediction using discriminant analysis, decision tree, and artificial neural network." (2012).
[10] Karabulut, Esra Mahsereci, and Turgay Ibrikci. "Analysis of cardiotocogram data for fetal distress determination by decision tree based adaptive boosting approach." Journal of Computer and Communications 2.9 (2014): 32-37.
[11] Yılmaz, Ersen, and Çağlar Kılıkçıer. "Determination of fetal state from cardiotocogram using LS-SVM with particle swarm optimization and binary decision tree." Computational and mathematical methods in medicine 2013 (2013).
[12] Sundar, C., M. Chitradevi, and G. Geetharamani. "Classification of cardiotocogram data using neural network based machine learning technique." International Journal of Computer Applications 47.14 (2012).
[13] Sharma, Pooja, and Asst Prof Rupali Bhatia. "Implementation of decision tree algorithm to analysis the performance." (2012).
[14] Krupa, Niranjana, et al. "Antepartum fetal heart rate feature extraction and classification using empirical mode decomposition and support vector machine." Biomedical engineering online 10.1 (2011): 6.
[15] Adhatrao, Kalpesh, et al. "Predicting students' performance using ID3 and C4. 5 classification algorithms." arXiv preprint arXiv:1310.2071 (2013).
[16] Area, Sonepat, and Ranchi Mesra. "Analysis of Bayes, neural network and tree classifier of classification technique in data mining using WEKA." (2012).
[17] Lakshmi, B. N., T. S. Indumathi, and Nandini Ravi. "An hybrid approach for prediction based health monitoring in pregnant women." Procedia Technology 24 (2016): 1635-1642.
[18] Jantan, Hamidah, Abdul Razak Hamdan, and Zulaiha Ali Othman. "Human talent prediction in HRM using C4. 5 classification algorithm." International Journal on Computer Science and Engineering 2.8 (2010): 2526-2534.
[19] Deressa, Tadele Debisa, and Kalyani Kadam. "Prediction of fetal health state during pregnancy: a survey." International Journal of Computer Science Trends and Technology (IJCST) 6.1 (2018).