RESEARCH ARTICLE OPEN ACCESS

Prediction of Fetal Health State during Pregnancy: A Survey

Tadele Debisa Deressa [1], Kalyani Kadam [2]
M.Tech Scholars [1], Assist. Professor [2]
Department of Computer science and Engineering
Symbiosis International University, Pune412115
India

ABSTRACT

Fetal Health is the indicator of fetal wellbeing and regular contact in the uterus of pregnant women during pregnancy. Most pregnancy period complication leads fetus to a severe problem which restricts proper growth that causes impairment or death. Safe pregnancy period by predicting risk levels before the occasion of complications encourage proper fetal growth. Prediction of fetal health state from a set of pre-classified patterns knowledge is a common important in developing a predictive classifier model using data mining algorithms. In this paper the University of California Irvine (UCI) Cardiotocogram (CTG) dataset is the source of information required for building a model which predicts whether the fetal is healthy or unhealthy. It consists of 2126 instances with 22 attributes obtained from FHR and UC which are multivariate datatypes. The major goal of this study is to choose efficient data mining algorithm which develops the accurate predictive model to manage knowledge of fetal health during pregnancy.

Keywords:- Data mining techniques, Classification, Cardiotocograph, Fetal health, Pregnancy, Prediction

I. INTRODUCTION

Pregnancy is a period during which offspring develop inside a woman [1]. Fetal health is the indication proper growth of the fetus in the pregnant woman's uterus during the gestation period. Physiological and morphological changes in pregnant women are at serious risk of pregnancy that affect fetal growth and movement [1] [2]. Most Maternal related factors that affect fetus health and leads to premature, stillbirth, gestational diabetes, heart disease etc. [2]. In case of Physiological parameter abnormalities; inconsistent fetal movement may cause maternal and fetal death [3]. Although decreasing of fetal movement due to internal and external factors is the crucial causes that adverse birth outcome unless monitored consistently. In this study fetal heart rate (FHR) and uterine contraction (UC) signal are the two selected attribute which reflect variability in fetal health sate abnormality [3]. In order to examine the fetal well-being UCI CTG dataset is the standard source of information and an electronic fetal state monitoring that used for predictive classification purpose [3] [5]. The FHR has been obtained from a Doppler ultrasound and mother's UC by pressure transducer were recorded on cardiotocograph (CTG) [5].

Mothers related high blood pressure abnormality during the gestation period suffer 10% of pregnant women around the world that leads to stress, impairment, and death of both lives as the World Health Organization (WHO) studies reflect [6]. This affects the babies in uterus to get insufficient blood circulation, which reduces fetus movement [6]. Data mining and machine learning algorithms were used for maximizing performance of choosing classifier that builds an accurate learning model for predicting the risk based on CTG dataset [7]. The baby's Fetal Heart Rate and UC are collected on CTG techniques that is highly instrumental in the previous abnormalities identification and provides the obstetrician to predict future risks [16]. Due to the high opportunity of fetal movement detection is more possible in late gestation the CTG data at the third trimester stage was preferable to predict using data mining algorithms. Therefore, this paper proposed building accurate predictive model for fetal health state management using the data mining algorithm by applying on medical dataset collected during pregnancy.

II. METHODS

2.1 Dataset description

The UCI CTG dataset is the source of information required for the analysis and development of fetal health predictive model [3]. CTG information also provides a visualized unhealthiness of the fetus that helps for early intervention before the risk happening. It consists of 2126 instances with 22 chosen attributes which are multivariate datatypes [3]. To do the analysis attributes of the fetus, such as FHR and maternal related UC are required. Most of the attributes are numerical (the combination of discrete and continuous data). The Attributes from 1to 9 are discrete and 10 to 20 are continuous whereas 21to 22 are nominal variable.

Table 1: Dataset description

Dataset Name	No Instance	No attributes	Datatypes
CTG	**2126**	**22**	**Multivariate**

2.1.1 Fetal Data Set Characteristics

Table 2: CTG instances attribute

No	Attributes	Data types
1.	FHR baseline (beats per minute)	Numerical
2.	Number of accelerations per second	Numerical
3.	Number of fetal movements per second	Numerical
4.	Number of uterine contractions per second	Numerical
5.	Number of light decelerations per second	Numerical
6.	Number of severe decelerations per second	Numerical
7.	Number of prolonged decelerations per second	Numerical
8.	Number of highest histogram peaks	Numerical
9.	Number of lowest histogram zeros	Numerical
10.	Percentage Of Time With Abnormal Short Term Variability	Numerical
11.	Mean Value Rate Of Short Term Variability	Numerical
12.	Number of lowest histogram zeros	Numerical
13.	Percentage Of Time With Abnormal State Long Term Variability	Numerical
14.	Mean Value Rate Of Long Term Variability	Numerical
15.	Width Of Fetal Heart Rate Histogram	Numerical
16.	Minimum Of Fetal Heart Rate Histogram	Numerical
17.	Histogram Mode Value	Numerical
18.	Histogram Mean Value	Numerical
19.	Histogram Median Value	Numerical
20.	Histogram Variance Value	Numerical
21.	Histogram Tendency	Nominal
22.	Class Label (Normal, Suspect, Pathological))	Nominal

2.2 Data mining Techniques Concepts

Data mining technique provides the ability to extract value knowledge as well as hidden information from large databases in the medical system. Smart tools and technology has been designed with various built-inefficient algorithms for developing efficient medical decision making system

2.2.1 Classification

Classification is one of the supervised data mining techniques that employ a pre-classified dataset to develop a predictive model and perform diagnosis [15] [17]. Classification is a necessary supervised learning task used to predict an unborn baby health state from CTG dataset. Most of the data mining algorithms that have been developed for data classification in medical diagnosis are artificial neural network (ANN), Opposition Based Firefly Algorithm (OBFA), decision tree (DT), C4.5 algorithm, K-nearest neighbor (KNN), random Forest (RF), naive Bayesian (NB), Support vector machine (SVM) and Genetic algorithm (GA) etc. [3] [7] [8] [9] [10] [12] [13] [17].

2.2.1.1 Genetic Algorithm (GA)

The importance of feature selection [16] is for optimized, accurate prediction and increasing speed and minimizing predictors cost. The Major Feature selection steps:

a) Initial population
b) Subset input Selection by fitness value
c) Crossover and mutation
d) Obtaining the best new population
e) Checking the maximum accuracy
f) Obtain the optimal features

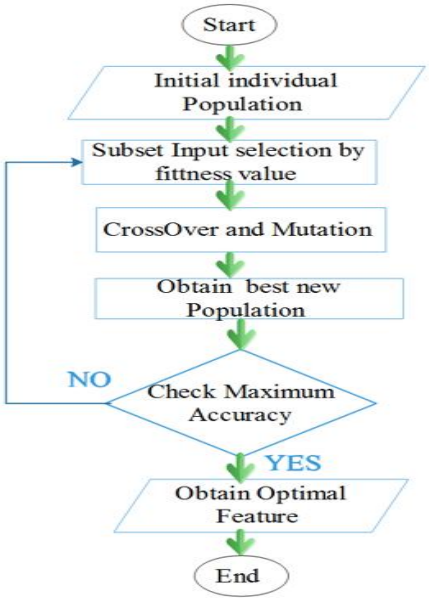

Fig 1: GA feature selection

2.2.1.2 Artificial Neural Network (ANN)

ANN classifier was composed of a number of input layers, hidden layer and outputs with respect to their weight that transfer signal between neurons [3] [7]. In this algorithm the model will be trained by backpropagation learning methods to predict complex human health related nonlinear data.

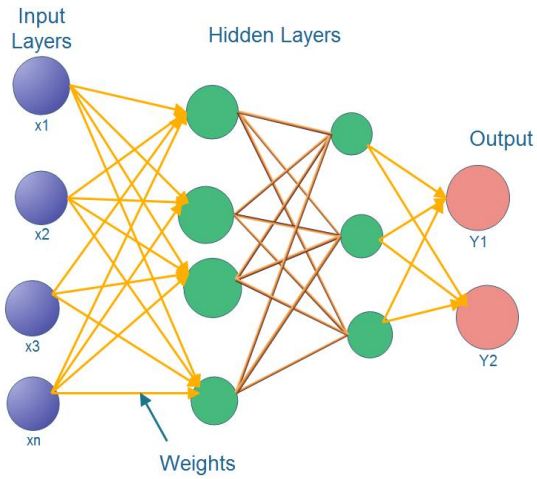

Fig 2: ANN layers

$$Yi = f\sum_{i=1}^{n} wixj \quad \ldots\ldots\ldots\ldots\ldots\ldots (1)$$

The above figure 2 shows the role of input data, hidden process and Weight over the learning output.

2.2.1.3 Support Vector Machine

SVM is a popular classier that performs classification tasks by constructing hyperplanes in a multidimensional space that separates cases of different class labels. Both regression and classification tasks are supported by SVM that consists multiple numerical variables [5] [7]. Nonlinear kernel functions help in order to integrate data into a suitable form that tends to split the data [16]. Due to a better generalization capability and low computational cost, RBF kernel has been applied in this work for separating the optimal hyperplane

$$\begin{pmatrix} Wx+b \leq -1 \\ Wx+b = 0 \\ Wx+b \geq -1 \end{pmatrix} \quad \ldots\ldots (2)$$

Splitting multidimensional data into two or more classes requires hyperplane which convert to linearly separable data [3].

2.2.1.4 K-Nearest Neighbor algorithm (KNN)

KNN is an algorithm which is based on distance between items. K-NN was the parameter free method which uses a Euclidean distance measure. This algorithm expressed in terms of the resemblance measure among the pair of data in N-dimension, the number of nearest data that are trained for

classification, and vector of training data applied by the classifiers. K-NN can be applied using proper procedures [3].

Strategy I

1. Associate weights with the attributes
2. Assign weights according to the relevance of attributes
3. Assign random weights and calculate the classification error
4. Adjust the weights according to the error
5. Repeat till acceptable level of accuracy is reached

k-Nearest Neighbor
Classify (X, Y, *x*) // X: training data, Y: class labels of X, *x*: unknown sample
for *i* = 1 **to** *m* **do**
  Compute distance $d(X_i, x)$
end for
Compute set *I* containing indices for the *k* smallest distances $d(X_i, x)$.
return majority label for $\{Y_i \text{ where } i \in I\}$

$$de = \sum_{i=1}^{n} \sqrt{xi^2.yi^2}$$ **…… Equation** (3)

Whereas de is the Euclidean distance computed, xi attribute size and yi is the result size

2.2.1.5 Random Forest algorithm (RF)

RF algorithm was a particular synthesis of classification accuracy and give chance to succeed besides optimal generalizations based on bagging methods [3].

Some of considerable features of RF are:-

- Giving the chance of reliability classification any techniques ,
- Allow to examine necessity of the classifiers and
- Examine Trained classifier that allows to identify correlations among selected data's

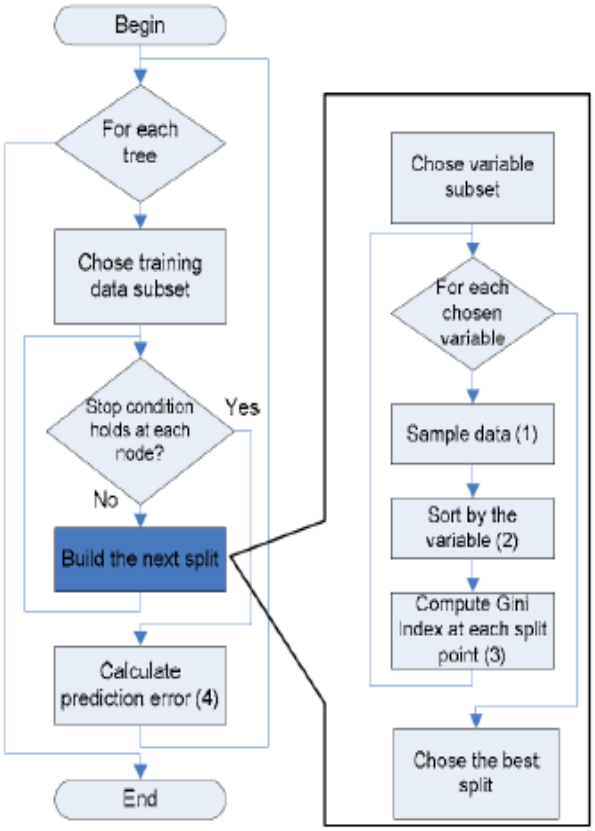

Fig 3: Random Forest flow

2.2.1.6 C4.5 Decision Tree Classifier

C4.5 algorithm consists vector of attribute values, mapping function correspond from attribute for classes and needs to know the types of classes in the existing dataset. Information gain was utilized to drive Gain ration which defined as follows [2] [15]. Information gain has been utilized as attributes reduction to produce Gain ratio.

$$Gainratio(E, Tf) = \frac{Gain(EP, Tf)}{SplitInfo(EP, ft)}$$ ….**Eqn** (4)

E –stands for entropy, Tf -represent training feature, EP-represent Entropy probability, ft.-represent feature test

2.3 Cross Validation (CV)

CV is the most critical statistical method for examining the efficiency of the model and judge the algorithm by performing a crossover check in successive rounds. The method provides ignoring to choose a specific part of training and testing set [3]. In k-fold CV the dataset is divided into k teams of uniform length, and repeat k times that each k teams reserved for validation and the other $k-1$ has been used for building the model [13].

III. LITERATURE REVIEW

In this paper several data mining algorithms have been studied for predicting fetal health state and grasped knowledge to provide good accuracy.

The abdominal ECG measurement tracking system studies had been proposed that the pre - process amplitude of movement in order to prevent fetal movement disorder [1]. The paper also points that decreasing in fetal movement is the sign of fetal healthiness that exposes to death unless detected early [1]. However, another author argued that the importance of hybrid data mining algorithm to a build model for pregnant women, health risk prevention, which caused by parameter inconsistency changes during pregnancy [2]. The C4.5 algorithm provided accurate performance of 98% [2]. The proper dataset consisting of relevant number of parameters and applying the hybrid approach my help for better prediction of fetal growth during pregnancy. Similarly, in other studies eight machine learning algorithms result over CTG dataset using weka tools has been reported [3]. The accurate prediction response of all the algorithms was examined by partitioning the dataset into ten (10) equal size for validating. The classifier model performance, highest accurate classification has been scored 99.2%. ANN drawbacks were solved by feedforward NN with non-linear functions which composed a number of weighing inputs, hidden layers followed by initiation function, a bias that provides output for the next layers [3]. ANN is applied with high accuracy for very large and complex data set. The theoretical concepts of fetal movement through summative context analysis method have been declared to solve inconsistent decreasing of fetal movement [4]. The Oral conversation between obstetricians and patient about the gestational state of the women cannot properly exert the factors that cause complications during the pregnancy. However, a new method has been explored to improve genetic algorithm feature selection for clinical features from the CTG dataset to diagnosis the fetal well-being [5]. The best average classification accuracy, performance has been achieved through Extreme Learning Machine (ELM) classifier was 93.61%. Also SVM was the popular acknowledged classifier to solve the nonlinear and binary or multiclass classification problem.

In similar way new algorithm which estimates an average of features intra-dependence has been demonstrated to improve the performance of the Doppler uterine artery experiment in high-risk of blood supplying to the fetus [6]. UC fired from Gestational Diabetes Mellitus affects only pregnant women that cause of maternal and infant death during labor [2] [6]. Henceforth, stated future work deals with the correlation between features to evaluate other classifier that positively affects the performance. In the studies [7] ANN method has been preferred for prediction of hypertension disease. High computation cost and long learning rate of ANN enforce extending the model to deep learning networks. Therefore, low computational cost and learning rate classifier required for such problems. In order to improve fetal risk prediction response hybrid approaches of SVM with the other attributes reduction method has been discussed [8]. SVM has been one of the most over-optimistic classifier methods and the genetic algorithm (GA) was demonstrated to reduce the number of features that maximize the classifier performance [9] [10]. SVM is a latest classification method that classifies any data into binary or multi classes by searching the best hyperplane that classify the data under proper category [11].

List of Metrics Equations	Eqn
$Sensitivity = 100 * \frac{TruePosistive}{Truepositive + FalseNegative}$	**(5)**
$Specificity = 100 * \frac{TrueNegative}{TrueNegative + FalsePositive}$	**(6)**
$Accuracy = 100 * \frac{TrueNegative + TruePositive}{TrueNegative + FalsePositive + TruePositive + FalseNegative}$	**(7)**

Table 3: Confusion matrix equation

Accuracy, F-measure, G-mean, and Kappa statistic whereas for more than two class label results are described in terms of Accuracy and G-mean. True positive (TP), true negative (TN), false positive (FP), and false negative (FN) are the four basic elements of a confusion matrix metric [10] [14]. ROC analysis was a standard tool for the design, optimization, and evaluation of the class classifiers by Cross-validation [13] [14].The studies [15] has been suggested that different data mining techniques to carry out accurate achievement that comfort pregnancy outcomes by predicting degree of risk. In [16] the author has been evaluated application of the algorithm on the different stages of pregnancy data may provide an objective measurement of fetal health condition. GA was preferred for giving a better feature subset whereas a linear SVM was chosen as the classifiers to investigate the relationship between features and adverse outcome [16]. Predicting the risk mitigation based on the model and for particular patient trend was the limitation of the paper. Moreover, GA and ANN has been examined for high power of optimal feature selection and global model respectively [17].The main goal of the study was employing a new hybrid technique for developing an accurate classifier model to predict fetal health during pregnancy period.

3.1 Limitation of previous studies

No	Title of the paper	Limitation
1	Feasibility Study of a new method for low-complexity fetal movement detection from abdominal ECG recordings	❖ Only Heart rate Considered ❖ The diagnosis lacks accuracy ❖ The analysis requires Latest technique for accuracy
2	Women and Birth, Fetal Movement :What are we telling women ?	❖ Lack of consistency advice and Absence evidence based guide
3	Accuracy of Fetal Kicks Detection During Pregnancy Using a Single Wearable Device	❖ No trust worthy of maternal annotation ❖ Only single machine learning tools investigated to explore diagnosis ❖ No proper and sufficient dataset
4	Fetal distress prediction using discriminant analysis, decision tree, and artificial neural network	❖ Discrete attribute was excluded ❖ Requires techniques which can be fit in discrete attributes ❖ Feature selection technique did not applied

Table 4: Limitation of previous studies

3.2 Summary of Methods

NO	Title of Paper	Techniques	Accuracy	Dataset
1.	Classification of the CTG For anticipation Fetal risk using Machine Learning techniques	RF	99.18	CTG dataset
		C4.5	99.13	
		SVM	98.96	
		CART	98.91	
		ANN	98.63	
		K-NN	98.42	
2	Using Machine Learning to Predict Hypertension from a Clinical Dataset	ANN model	*82*	Clinical dataset
3	A hybrid Filter-Wrapper Attribute Reduction Approach For Fetal Risk Anticipation	SVM with OBFA	92.85	CTG dataset
		Hybrid IG-OBFA	96.24	
4	Genetic Algorithm based feature subset selection for fetal state classification	SVM with GA	91.35	CTG dataset
5	Variable-length Accelerometer Features and Electromyography to Improve Accuracy of Fetal Kicks Detection During Pregnancy Using a Single Wearable Device	Random Forest	Increased by %11	Clinical record dataset
6	Opposition-Based Firefly Algorithm Optimized Feature Subset Selection Approach For Fetal Risk Anticipation	SVM with OBFA	92.85	CTG dataset
7	A Survey of Various Classification Techniques for Clinical Decision Support System	ANN with GA	Varied dataset	Clinical dataset

Table 5: Summary of the Methods

IV. PROPOSED SYSTEM

Fetal movement is the major sign of fetus healthiness confirmation during pregnancy. In order to promote fetal health state; inconsistent fetal movement complications that
may causes fetal death and stillbirth must be diagnosed early.

The proposed system addresses to design *a model for* fetal movement which improve the diagnosis quality for pregnant women suffering from decreasing fetal movement (DFM) using ensemble Methods. Genetic Algorithm, Artificial neural Network and Support vector machine are the recommended and suitable algorithm for the prediction of fetal movement problem treatment. Data mining algorithms are powerful to perform biomedical data regression, classification and visualization tasks, in order to facilitate decision-making for healthcare.

Problems addressed to be solved in further studies:-

- Complication and risk during pregnancy
- Absence of fetal movement at third trimester leads fetus and pregnant to high risk.
- An Inconsistent fetal movement that may cause fetal death maximize the stillbirth rate

4.1 System Architecture

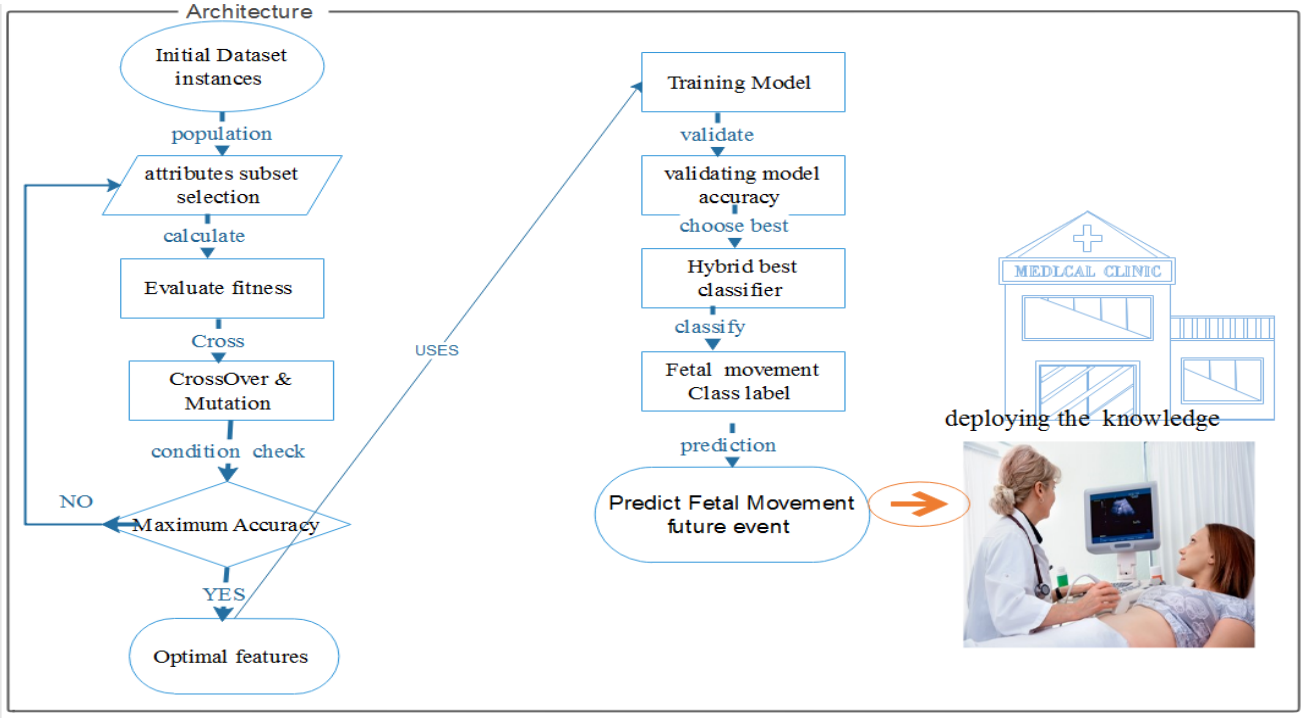

Fig 4: system architecture

V. CONCLUSION AND FUTURE WORK

This paper discusses the review of the various data mining algorithms and dataset used in fetal healthy diagnosis during pregnancy. CTG data were identified as an objective data for managing fetal health which is essential during the gestation period using data mining techniques. Random Forest, C4.5 (J48), Support vector machine, and artificial neural network through Genetic algorithm are the preferred methods to classify correctly fetal health state with Up-to-date pregnancy dataset. The study suggested that the hybrid of more than two data mining algorithm can improve the performance of the classifier to build an accurate fetal health state predictive model. In Future improvement Evidence-based analysis of fetal movement consistency model will be developed to manage a knowledge patterns of baby's movement in the uterus. Also Ensemble techniques that may help to provide better outcomes for fetal movement classification can be considered in the further work.

REFERENCE

[1] M.J. Rooijakkers, C.Rabbotti, H.de Lau, S.G.Oei, J.W.M Bergmann's, and M.mischi. "Feasibility Study of a new method for low-complexity fetal movement detection from abdominal ECG recordings." *Journal of Biomedical and Health Informatics*, IEEE, 2015, pp.2168-2194.

[2] Lakshmi.B.N, Dr.Indumathi.T.S, Dr.Nandini Ravi. "A hybrid Approach for Prediction Based Health Monitor in Pregnant Women." *(ICETEST),* Elsevier.2016.

[3] Hakan Sahani, Abdulhamit subasi. "Classification of the CTG For anticipation fetal risk using Machine Learning techniques." *Journal of Applied Soft computing*, Elsevier, 2015, pp.232-238

[4] Jane Warland, Pauline Glober. Women and Birth, Fetal Movement: What are we telling women? Elsevier, 2016, pp.2-6.

[5] Sindhu Ravindran, Asral Bahari Jambek, Hariharan Muthusamy, and Siew-ChinNeoh. "A Novel Clinical Decision Support System Using Improved Adaptive Genetic Algorithm for the Assessment of Fetal Well-Being." *Computational and Mathematical Methods in Medicine*, Hindawi. 2015, pp.2-11.

[6] Mario W. L. Moreira, Joel J. P. C. Rodrigues,Antonio M. B. Oliveira, Kashif Saleem, Augusto V. Neto . "An Inference Mechanism Using Bayes-based Classifiers in Pregnancy Care." *18th International Conference on e-Health Networking, Applications and Services*, IEEE, 2016.

[7] Daniel LaFreniere, Farhana Zulkernine, David Barber, Ken Martin. "Using Machine Learning to Predict Hypertension from a Clinical Dataset." *IEEE, 2016.*

[8] V. Subha; D. Murugan; A. Manivanna B. "A Hybrid Filter-Wrapper Attribute Reduction Approach for Fetal Risk Anticipation." *Asian Journal of Research in Social Sciences and Humanities* Vol. 7, No. 2, February 2017, pp.1094-1106.

[9] Hasan Ocak. "A Medical Decision Support System Based on Support Vector Machines and the Genetic Algorithm for the Evaluation of Fetal Well-Being." *J .Med System*, Springer, 2013, pp.2-9.

[10] V.Subha, D.Murugan and S.Prabha, A. Manivanna Boopathi. " Genetic Algorithm based feature subset selection for fetal state classification." *Journal of Communications Technology, Electronics and Computer Science*, 2016, pp.13-19.

[11] Marco Altini, Elisa Rosetti, Michiel Rooijakkers, Julien Penders, Dorien Lassens, Lars Grieten, Wilfierd Gyselaers . "Variable-Length Accelerometer Features and Electromyography to Improve Accuracy of Fetal Kicks Detection during Pregnancy Using a Single Wearable Device." IEEE, 2017, pp.221-224.

[12] Mei-Ling Huang, Yung-Yan Hsu. "Fetal distress prediction using discriminant analysis, decision tree, and artificial neural network." *J. Biomedical Science and Engineering*, 2012, pp.526-533.

[13] V.Subha (Assistant Professor), D.Murugan (Associate Professor). "Opposition-Based Firefly Algorithm Optimized Feature Subset Selection Approach for Fetal Risk Anticipation." Machine Learning and Applications: An International Journal (MLAIJ), 2016, pp.55-64.

[14] ErsenYJlmaz, CaLlarKJlJkçJer. "Determination of Fetal State from Cardiotocogram Using LS-SVM with Particle Swarm Optimization and Binary Decision Tree." *Hindawi Publishing Corporation Computational and Mathematical Methods in Medicin*e, 2013, Pp.1-9.

[15] Lakshmi.B.N, Indumathi.T.S, Nandini Ravi."A Comparative Study of Classification Algorithms for Predicting Gestational Risks in Pregnant Women." International Conference on Computers, Communications, and Systems, EEE, 2015, pp.42-46.

[16] Liang Xu, Antoniya Georgieva, Christopher W. G. Redman, Stephen J. Payne."*Feature Selection for Computerized Fetal Heart Rate Analysis using Genetic Algorithms, 2013.*

[17] Chaitali Vaghela, Changa, Nikita Bhatt and Changa Darshana. "A Survey on Various Classification Techniques for Clinical Decision Support System." *International Journal of Computer Applications*, 2015, pp.14-17.

AUTHOR

Tadele Debisa Deressa received his Bachelor Science degree in Computer science at Wollega University, Ethiopia in June, 2013. He has been teaching as assistant lecturer since July, 2013 in Wollega University. Currently he is pursuing Master's degree in CSE.