International Journal of Psychosocial Rehabilitation, Vol. 24, Issue 08, 2020
ISSN: 1475-7192

DIAGNOSING ABNORMALITY OF FOETUS USING MACHINE LEARNING ALGORITHMS

[1]Abhijith Ragav, [2]P.Surya, [3]Shivabhinav.S.G, [4]P.Akilandeswari

Abstract-The primary objective of the paper is to utilise machine learning algorithms for diag- nosing sufferance caused to the foetus by parameters such as Fetal Heart Rate (FHR) recordings (section). The significance of this work is to accurately predict the foetus condition at a much earlier stage since it is very important to analyse the issue at the right time to avoid complications. Concluding results from Cardiotocography (CTG), a test which is used widely for estimating fetal distress poses a major challenge. This work benefits the medical community, by detecting the fetal distress conditions using features derived from CTG results at a preliminary stage of 30-35 gestational weeks. The acute state can be classified by a few indications such as decrease in oxygen content due to reduction of haemoglobin count of fetal unit and is usually a complication of labour. The paper deals with the prior prediction of foetal conditions so as to provide early and effective treatment.

Keywords **:** Fetal Heart Rate, Cardiotocography, K Nearest neighbours, Support Vector, Machine, Radial Basis Function , Extreme Gradient Boosting.

I INTRODUCTION

The assessment of the fetal development by obstetricians is a major issue during antepartum period. Proper diagnosis has to be made at the correct time to assess whether the foetus is in distress so that corrective actions to avert issues like Fetal Asphyxia can be taken [Divya17]. Delay in diagnosis normally has severe consequences both on the infant and the mother, at times even causing death. In order to predict sudden problems like cord accidents or placental breakage, methods such as testing and monitoring need to be developed. Before and during child birth, fe- tal heart rate is monitored generally using CTG. From the fetal heart rate data, metabolic acidosis and hypoxic injury are determined. Disability in neural devel- opment, cerebral palsy or even death can be caused by hypoxic condition. Mothers life is risked by caesarean delivery which is performed as the normal intervention method. FHR and Uterine contractions (UC) data is very difficult to obtain and the obstetricians find it difficult to analyse it for the entire time period. Success and accuracy of the monitoring systems have to be confirmed so that a clinical or an optimal solution support system that

[1] *Information Technology, SRM IST, Chennai - 603203, Tamil Nadu, India abhijith.ragh@gmail.com*

[2] *Computer Science and Engineering, SRM IST, Chennai - 603203, Tamil Nadu, India suryaprabhakar41@gmail.com*

[3] *Computer Science and Engineering, SRM IST, Chennai - 603203, Tamil Nadu, India shivabhinav1999@gmail.com*

[4] *Assistant Professor,Computer Science and Engineering, SRM IST, Chennai - 603203, Tamil Nadu, India akilandeswari.p@ktr.srmuniv.ac.in*

International Journal of Psychosocial Rehabilitation, Vol. 24, Issue 08, 2020
ISSN: 1475-7192

provides superior visual estimation of signals can be proposed. If not properly analysed, low risk pregnancies can lead to increase of caesarean delivery. Misdiagnosis generally takes place due to misinterpretation of obstetricians, which results in improper treatment, which can lead to death.The main advantage of this concept is that it classifies fetal rate into three categories or results which includes a suspect category, a normal category and a pathological category.

II Algorithms

2.1. Logistic Regression for predicting abnormality of the foetus

Logistic Regression is one of the go-to algorithms when it comes to Binary Classifi- cation problems. It predicts the probabilities using the sigmoid function. The values are mapped to 0 and 1 using a threshold classifier. Logistic Regression predicts the probability of an event occurring with the

$$S(a) = 1/(1 + e^{(-aX)^T}) \qquad (1)$$

TABLE I. CONFUSION-MATRIX (LOGREG)

Predicted	Actual		
	Normal State	**Suspect State**	**Pathological State**
Normal State	0.9401	0.0479	0.0119
Suspect State	0.3709	0.5968	0.0323
Pathological State	0.1333	0.2	0.6667

As can be observed from TABLE I, Logistic Regression predicts Normal cases with 94% accuracy, Suspect cases with 59.6% accuracy and Pathological cases with 66.6% accuracy.

2.2. *K-Nearest Neighbours for predicting abnormality of the foetus*

The K-Nearest Neighbours algorithm (KNN) is an algorithm which can be used for classification and regression. It is a non-parametric algorithm. As the number of neighbours increase, the decision boundary is observed to smoothen out.

TABLE II. CONFUSION-MATRIX (KNN)

Predicted	Actual		
	Normal State	Suspect State	Pathological State
Normal State	0.9701	0.0269	0.0029
Suspect State	0.3871	0.5967	0.0161
Pathological State	0.2333	0.0667	0.7

• KNN classifier deals with maximum votes given by its neighbours using which an object is classified, where it is allocated to the class which is most common among its k nearest neighbours. The output in this is in the form of class membership.

• In KNN regression, the average value of the nearest neighbours is the output which in turn is the property value of the object.

As can be observed from TABLE II, KNN predicts Normal cases with 97% accuracy, Suspect cases with 59.6% accuracy and Pathological cases with 70% accuracy.

2.3. *Support Vector Machine (SVM) for predicting abnormality of the foetus*

SVM classifies the different classes based on a linear line or Hyper-Plane. The aim of the SVM is to find the most suitable Hyper-Plane that minimizes the distance between the plane and the Training Examples.

$$S(a_i, a_j) = (\varphi(a_i).\varphi(a_j))F(a) = sign(\bar{\ } S(a_i, a_j) + c) \quad (2)$$

Two SVM Kernels have been compared for solving the Classification problem in hand.

- Gaussian Radial Basis Function Kernel (RBF):-

$$S(a_i, a_j) = e^{(-\Gamma ||a_i - a_j||^2)} \quad (3)$$

Where, $\Gamma = 1/(\sigma\sqrt{2})^2$

- Linear Kernel

TABLE III. CONFUSION-MATRIX (SVM-RBF KERNEL)

Predicted	Actual		
	Normal State	**Suspect State**	**Pathological State**
Normal State	0.9731	0.0239	0.0029
Suspect State	0.1452	0.8548	0
Pathological State	0.1	0.1	0.8

As can be observed from TABLE III, the SVM RBF kernel predicts Normal cases with 97.3% accuracy, Suspect cases with 85.4% accuracy and Pathological cases with 80% accuracy.

2.4. *Decision Tree for predicting abnormality of the foetus*

Decision Tree is one of the Machine Learning Algorithms which follows a Tree structure to determine the statistical probability. Each and every problem or a decision is represented as a branch of the tree. The tree begins with a single node which gives a set of possible outcomes, then each of these outcomes results into another set of possibilities. The internal nodes represent the decisions taken and the leaf node represents the final outcome. Hence it forms a tree like structure. The decision tree takes plenty of parameters, but the most important parameter is 'max depth' which is used to adjust the depth of the tree, the accuracy of the model improves as the depth increases, but we must ensure that the assignment of the depth is not very large as it results in overfitting.

TABLE IV. CONFUSION-MATRIX (DECISION TREE)

Predicted	Actual		
	Normal State	Suspect State	Pathological State
Normal State	0.9551	0.0449	0
Suspect State	0.1774	0.8226	0
Pathological State	0.0667	0.0333	0.9

As can be observed from TABLE IV, Decision Tree Classifier predicts Normal cases with 95.5% accuracy, Suspect cases with 82.2% accuracy and Pathological cases with 90% accuracy.

2.5. *Nave Bayes for predicting abnormality of the foetus*

Naive Bayes is an algorithm that is primarily based on Bayes Theorem. The algo- rithm is naive and assumes low dependence among the features. Therefore, it may not perform well in situations where certain features have higher importance than the others. Given below are two types:-

- Bernoulli Nave Bayes:- This is based on multivariate Bernoulli distribu- tions, where each feature is assigned to a binary value:-

$$P(a_i|b) = P(i|b)a_i + (1 - P(i|b))(1 - a_i) \quad (4)$$

- Gaussian Nave Bayes:- This implements the Gaussian form of algorithm for classification and performs better in this case.

$$P(x_i|y) = (1/(sqrt(2\sigma^2)))e^{(-(x_i - \mu_y)^2 / 2\sigma_y^2)} \quad (5)$$

As can be observed from TABLE V, Gaussian Naive Bayes produced slightly better results when compared to the Bernoulli Naive Bayes. The Gaussian Naive Bayes Classifier predicts Normal cases with 78.% accuracy, Suspect cases with 91.9% ac- curacy and Pathological cases with 66.6% accuracy.

TABLE V. CONFUSION-MATRIX (GAUSSIAN NAÏVE BAYES)

Predicted	Actual		
	Normal State	Suspect State	Pathological State
Normal State	0.7844	0.1796	0.0359
Suspect State	0.0645	0.9194	0.0161
Pathological State	0.0333	0.3	0.6667

2.6. *Extreme Gradient Boosting (XGBOOST) for predicting abnormality of the foetus*

Boosting is an ensemble learning technique which builds a strong classifier from bunch of weak classifiers. Boosting is sequential, unlike bagging which is parallel, i.e. first the model is trained by one learner and then the next learner trains the model according to the errors made by the previous learner. XGBoost is a tree boosting algorithm that solves many problems in an efficient manner.

TABLE VI. CONFUSION-MATRIX (XGBOOST)

Predicted	Actual		
	Normal State	Suspect State	Pathological State
Normal State	0.9760	0.0209	0.0029
Suspect State	0.0806	0.9193	0
Pathological State	0	0.0333	0.9667

As can be observed from TABLE VI, XGBOOST predicts Normal cases with 97.6% accuracy, Suspect cases with 91.9% accuracy and Pathological cases with 96.7% accuracy.

III Dataset

The CTG data was obtained from Kaggle and classified into 3 classes by expert obstetricians[Kaggle18].

- ALTV- Abnormal Long-Term Variability (%).
- ASTV- Abnormal Short-Term Variability (%).
- FHR- Baseline FHR (BPM)
- Max- FHR histogram's maximum
- Mean- Histogram's mean
- Median- Histogram's median
- Min- FHR histogram's minimum
- MLTV- Long-term variability's mean value
- Mode- Histogram mode
- MSTV- Short-term variability's mean value
- NMax- # of peaks in histogram
- NSP- (Normal, Suspect, Pathological)
- Nzeros- # of histogram zeroes
- Tendency- Histograms tendency
- Variance- Histograms variance
- Width- FHR histograms width

All the below mentioned parameters are expressed on per-seconds basis:-

- DP- Prolonged Deceleration
- FM- Fetal movements
- DS Severe Deceleration
- DL Light Deceleration
- UC- Uterine contractions
- AC- Acceleration

IV Feature Importances (XGBoost)

Feature importance is used to determine the significance of each and every feature for a given Machine learning model. Feature importance is a numerical value (0 - 1) given to every feature, depending upon the impact each feature has during prediction. Consequently, it becomes easy to perform feature engineering and re- move redundant features. Since Extreme Gradient Boosting (XGBOOST) is the best performer of all the algorithms used in the case study, its feature importances are extracted and represented in TABLE VII. MSTV has the significantly higher feature importance when compared to other features.

Received: 21 Dec 2019 | Revised: 18 Jan 2020 | Accepted: 05 Feb 2020

International Journal of Psychosocial Rehabilitation, Vol. 24, Issue 08, 2020
ISSN: 1475-7192

SNO	FEATURE	FEATURE IMPORTANCE
1	LBE	0.078515
2	AC	0.048153
3	FM	0.033456
4	UC	0.066912
5	ASTV	0.135564
6	MSTV	0.24367
7	ALTV	0.097273
8	MLTV	0.074260
9	DL	0.013730
10	DP	0.046799
11	Width	0.038290
12	Mix	0.047380
13	Max	0.039257
14	Nmax	0.033262
15	Nzeros	0.005608
16	Mode	0.042932
17	Mean	0.059563
18	Median	0.055502
19	Variance	0.052021
20	Tendency	0.007155

TABLE VII

V Conclusion and Future Work

From the comparative study of the various classification algorithms used, XG- BOOST proved to be the best performer with a prediction accuracy of 97.6% for Normal cases, 91.9% for Suspect cases and 96.7% for Pathological cases. Figure 1 is a bar chart comparing the different Classification Algorithms and their respective Prediction Accuracy (in %). This work can be implemented into an IoT-Edge device which provides real time foetal monitoring during gestational period (30-35 weeks) to reduce the number of pathological cases during birth. Advanced Nuero Fuzzy Algorithmns may also be tried out[Knack99]. The accuracy of the device increases as more patients are diagnosed which is due to the increase in training data.

VI Motivation

Antepartum Foetal Monitoring can be used to predict the foetal condition only during labour whereas machine learning algorithms predict the foetal condition in just 30-35 gestational weeks, so proper medical assistance can be given to the foetus at an early stage if the state of foetus is diagnosed to be pathological. This prompted

Received: 27 Mar 2020 | Revised: 20 Apr 2020 | Accepted: 05 May 2020

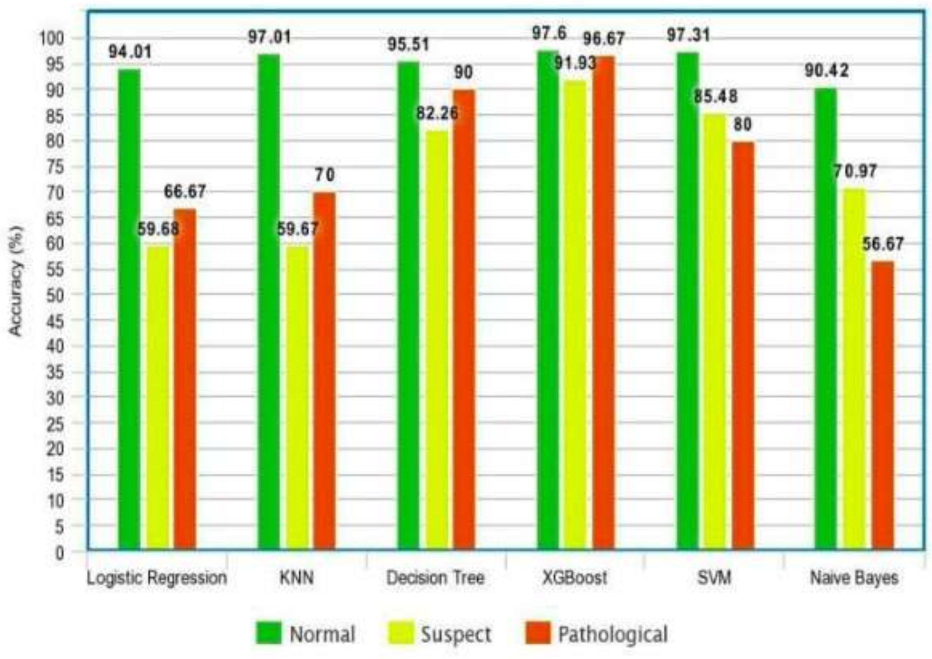

Figure 1the authors to figure out the most efficient method by comparing various algorithms to predict the foetal state using CTG data.

VII Bibliography

1. Divya Sampath V.N.H.G , Corns S.,Long S., Evaluation of Support Vector Machines and Random Forest Classifiers in a Real-time Fetal Monitoring System Based on Car- diotocography Data978-1-4673-8988-4/17/$31.00 2017 IEEE
2. Knack D, R Kruse, ”Obtaining interpretable fuzzy classification rules from medical data”, Artificial Intelligence in Medicine, vol. 16-2, pp. 149-169, June 1999.
3. Kaggle, ”UCI Cardiotocography”, Retrieved from the following link https://www.kaggle.com/propanon/ucicardiotocography