Comparative Analysis of Various Algorithms for Fetal Risk Prediction

Jay Mistry[1]

[1] Fourth Year Student, Information Technology Department, Thadomal Shahani Engineering College, Bandra(W), Mumbai - 400050

Abstract - *Fetal problems emerge while your unborn child grows within the womb. Congenital refers to the fact that a child is born with these conditions. Some fetal disorders are inherited genetically or from a parent. A machine learning approach for predicting if a woman is having a high risk fetus is needed. We have collected the data from an online repository. The data is been balanced and pre-processed for better prediction. This dataset is being used on machine learning algorithms like Random Forest, Bagging, AdaBoostM1, SMO, Kstar, Naïve Bayes, Hoeffding Tree and Classification via Regression to build a model which will help in predicting the fetal condition. All of these algorithms is compared with specific performance metrics and the best result is showcased.*

***Key Words*: Fetal, pre-processed, prediction, performance metrics**

1. INTRODUCTION

As included in NCI dictionary of Cancer terms, fetus is defined as the unborn baby that develops and grows inside the uterus in humans. The fetal period begins 8 weeks after fertilization of an egg by a sperm and ends at the time of birth. There might be cases where a pregnant lady could have pregnancy risks or distressing conditions for the fetus. This distress is may be due to low levels of amniotic fluids, high levels of amniotic fluids, placental abruption, uncontrolled diabetes and pregnancy lasting for more than 40 weeks. Fetal surgery may be advantageous for babies with specific birth abnormalities. Our specialists carry out these extremely difficult treatments while your unborn child is still in the womb. There are tests that are used to assess the fetal health such as fetal movement counts, biophysical profile, contraction stress test and Doppler ultrasound exam of the umbilical artery. A method for keeping track of uterine contractions and fetal heart rate during pregnancy is called Cardiotocography, or CTG. It is used to evaluate the health of the fetus and to spot fetal distress early.

To create a model that represents the various data classes and forecasts future data trends, classification and prediction are used. With the aid of prediction models, classification forecasts the category labels of data. We have the clearest knowledge of the data at a broad scale thanks to this analysis. Prediction models predict continuous-valued functions, whereas classification models predict categorical class labels. For instance, based on a person's income and line of work, we can create a classification model to classify bank loan applications as safe or risky, or a prediction model to estimate how much money a potential consumer will spend on computer equipment.

1.1 Motivation

Considering all the tests that used to detect fetal abnormalities, very few of them are actually reliable. Prenatal tests are not always perfect. The data for false-positives or false-negatives varies from test to test. These procedures might carry a real risk of miscarriage because an amount of amniotic fluid or tissue from around the fetus is needed.

1.2 Problem Statement

An analysis of various machine learning algorithms to select which algorithm can accurately predict the level of risk in fetus based on few performance metrics.

2. Related Work

In [10] J. Li and X. Liu have implemented twelve machine learning algorithms on CTG dataset. The proposed model has performed brilliantly in different classification model evaluations. The four top models are then combined to create the Blender Model using the soft voting integration method, which is then contrasted with the stacking integration method. The model described in this study outperformed the conventional machine learning models in a variety of Classification Model evaluations, achieving accuracy rates of 0.959, AUCs of 0.988, recall rates of 0.916, precision rates of 0.959, F1s of 0.958, and MCCs of 0.886.

In [7] R. Chinnaiyan and S. Alex have used Machine Learning algorithm to build a predictive classifier to forecast the fetal health and growth state from a set of pre-classified patterns knowledge. The majority of this evaluation of the literature focuses on fetal anomalies that occur during the first trimester of pregnancy. The major goal of this article review is to investigate the various machine learning processes for accurate diagnosis and prognosis of abdominal anomalies in order to lower the incidence rate. Tested machine learning algorithms save time and effort while delivering more precise results. Segmentation, Image Enhancement, Feature Extraction, and Image Classification are used to accomplish this.

In [6] the results of the tests display a classifier model to be 83% and 84% accurate, before and after feature selection, respectively. This research proposes a classification based on association (CBA) that is a rule-based method to the Cardiotocographic analysis. A useful indicator of fetus health confirmation is fetal improvement. One of the main causes of foetal death is irregular fetal movement. Therefore, early diagnosis is necessary to promote the foetal health state. The method under consideration aims to develop a model employing the associative classification technique to study fetal movement in order to improve the accuracy of diagnoses for expectant women while minimising fetal movement (DFM)

3. IMPLEMENTED WORK

Based on the above research, we have implemented an analysis between different Machine Learning algorithms. After downloading the data from Kaggle (a free online repository), it was inserted on the data mining application WEKA. The dataset looks is shown in the Figure 1.

histogram	histogram	histogram	histogram	histogram	histogram	histogram	histogram	fetal_health
126	2	0	120	137	121	73	1	2
198	6	1	141	136	140	12	0	1
198	5	1	141	135	138	13	0	1
170	11	0	137	134	137	13	1	1
170	9	0	137	136	138	11	1	1
200	5	3	76	107	107	170	0	3
200	6	3	71	107	106	215	0	3
130	0	0	122	122	123	3	1	3
130	0	0	122	122	123	3	1	3
130	1	0	122	122	123	1	1	3
186	2	0	150	148	151	9	1	2
186	5	0	150	148	151	10	1	2
154	5	0	135	134	137	7	1	1
158	2	0	141	137	141	10	1	1
174	7	0	143	125	135	76	0	1
174	3	0	134	127	133	43	0	1
178	5	0	143	128	138	70	1	1
174	5	0	134	125	132	45	0	2
158	6	0	133	124	129	36	1	1
177	6	1	133	129	133	27	0	1
182	13	0	129	104	120	138	0	3
198	9	0	129	125	132	34	0	1

Figure 1. Dataset

For better prediction, data was preprocessed. If you work with large amounts of raw data or big data, you are aware of how crucial data preprocessing is to enhancing the quality of the data as the retrieved raw data is frequently unreliable, imperfect, and noisy. We must ensure that our training data is in the right format before using it to train an algorithm. Data pretreatment is a crucial stage in data mining since it helps to spot errors, outliers, noise, and missing important variables. These data mistakes would persist without data pretreatment in data science, lowering the calibre of data mining. After the data cleaning is finished, a number of processing steps are included in data pre-processing, such as data integration, data conversion, and other processing steps. The filters available in WEKA tool helped in transforming numerical attributes to nominal attributes which is showed in the Figure 2.

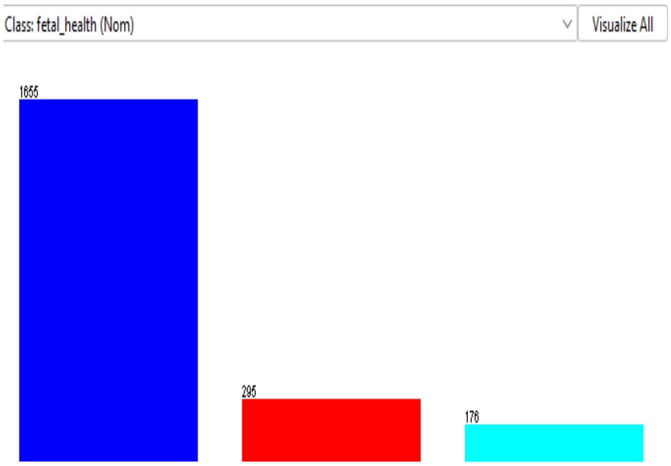

Figure 2. Nominal representation

Risk was divided into three categories namely Low, Medium and High level risk. 1 (Blue color) being the lowest and 3 (Sky blue color) being the highest.

Class-balancing was also performed simultaneously. The aim of Normalization is to scale down features to a similar scale. This increases training stability. The nominal data now consists of values ranging between 0 to 1. The screen in WEKA provided description about the data in the dataset. Missing values and redundant data were removed so that it won't hinder the accuracy of the model. The dataset consists of 11 columns and 1 column for prediction. The attributes are baseline value, fetal_movement, light_decelerations, severe_decelerations and prolongued_decelerations, uterine_contractions etc [4]. The graphs between all of those 11 columns were visualized using WEKA.

Open the classify tab in WEKA and go on selecting different algorithms for receiving a better accuracy. Each of those algorithms also include hyper-parameter tuning. For this, RandomForest, Bagging, AdaBoostM1, SMO, Kstar, NaiveBayes, HoeffdingTree and ClassificationviaRegression algorithms were used.

The selected algorithms are then analyzed and compared with three performance metrics to evaluate the model. An accurate model with good accuracy and performance measure is then picked as the preferred algorithm. The flow of the procedure is constructed using a block diagram in the Figure 3.

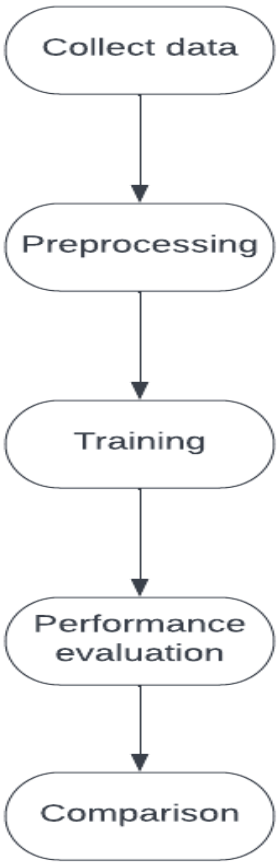

Figure 3. Block diagram

4. RESULTS AND ANALYSIS

In the classification process, all the algorithms were trained with the training datasets as well as with k-fold cross validation (k=10). A summary of each model was noted which is displayed in Table 1.

Algorithms	Training (%)	Precision	Recall	F-score	Time (secs)
NaiveBayes	89.5	0.96	0.92	0.94	0.01
AdaBoostM1-DecisionStump	77.8	0.77	1	0.87	0.08
Bagging-J48	91.8	0.92	0.97	0.95	0.01
RandomForest	99.9	0.99	1	0.99	0.06
HoeffdingTree	89.5	0.96	0.92	0.94	0.01
Classificationvia Regression	97.4	0.97	0.93	0.98	0.04
SMO	98.2	0.98	0.99	0.98	0.04
Kstar	99.9	0.99	1	0.99	5.49

Table 1. Analysis

According to the result displayed, we can see that RandomForest and Kstar algorithm gives the same accuracy. However when time complexity is included RandomForest is preferred more. Precision, Recall and F-measure are super standard way to evaluate a model. Precision helps us to measure the ability to classify positive samples in a dataset. Recall helps us to measure positive samples that are correctly classified by the model. F-measure is the harmonic mean of Precision and Recall. If the F-score is higher than the model is considered better. Formula for each of the metric is shown in Figure 4.

$$
\begin{aligned}
precision &= \frac{TP}{TP + FP} \\
recall &= \frac{TP}{TP + FN} \\
F1 &= \frac{2 \times precision \times recall}{precision + recall}
\end{aligned}
$$

Figure 4. Formula

Parameters that assisted RandomForest to get high accuracy are batchSize=100, numIterations=100, seed=1, macDepth=0. These were set by the WEKA tool by default.

5. CONCLUSIONS

A healthy birth comes from a healthy pregnancy. Prenatal care improves the chances of a healthy and risk-free pregnancy and birth. This begins with pre-pregnancy care. A pre-pregnancy check and prenatal care can help in the prevention of complications and help women in understanding how they can keep the baby healthy while taking care of themselves. Doctors can solve the issue by predicting if the fetus is under distress or risk-free.

To do the prediction, first data is downloaded, pre-processed. Various algorithm were trained on that model. The accuracy we received was highest for RandomForest algorithm. Precision, Recall and F-score also comments that RandomForest would be the preferred algorithm among the selected ones.

In future, we can create a system which predicts abnormalities in fetus and which procedure needs to be carried out even though it may require extensive dataset.

REFERENCES

[1] P. N. Tan, M. Steinbach, Vipin Kumar, “Introduction to Data Mining”, Pearson Education.

[2] Han, Kamber, "Data Mining Concepts and Techniques", Morgan Kaufmann 3nd Edition.

[3] WEKA: https://www.cs.waikato.ac.nz/ml/weka/

[4] Dataset:- https://www.kaggle.com/datasets/andrewmvd/fetal-health-classification

[5] A. K. Pradhan, J. K. Rout, A. B. Maharana, B. K. Balabantaray and N. K. Ray, "A Machine Learning Approach for the Prediction of Fetal Health using CTG," 2021 19th OITS International Conference on Information Technology (OCIT), 2021, pp. 239-244, doi: 10.1109/OCIT53463.2021.00056.

[6] J. Piri and P. Mohapatra, "Exploring Fetal Health Status Using an Association Based Classification Approach," 2019 International Conference on Information Technology (ICIT), 2019, pp. 166-171, doi: 10.1109/ICIT48102.2019.00036.

[7] R. Chinnaiyan and S. Alex, "Machine Learning Approaches for Early Diagnosis and Prediction of Fetal Abnormalities," 2021 International Conference on Computer Communication and Informatics (ICCCI), 2021, pp. 1-3, doi: 10.1109/ICCCI50826.2021.9402317.

[8] A. Chowdhury, A. Chahar, R. Eswara, M. A. Raheem, S. Ehetesham and B. K. Thulasidoss, "Fetal Health Prediction using neural networks," 2022 8th International Conference on Advanced Computing and Communication Systems (ICACCS), 2022, pp. 256-260, doi: 10.1109/ICACCS54159.2022.9784987.

[9] K. Agrawal and H. Mohan, "Cardiotocography Analysis for Fetal State Classification Using Machine Learning Algorithms," 2019 International Conference on Computer Communication and Informatics (ICCCI), 2019, pp. 1-6, doi: 10.1109/ICCCI.2019.8822218.

[10] J. Li and X. Liu, "Fetal Health Classification Based on Machine Learning," 2021 IEEE 2nd International Conference on Big Data, Artificial Intelligence and Internet of Things Engineering (ICBAIE), 2021, pp. 899-902, doi: 10.1109/ICBAIE52039.2021.9389902.