International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 10 Issue 11, November-2021

Fetal Health Prediction using Classification Techniques

Naveen Reddy Navuluri
Department of Electronics and communication
Maulana Azad National Institute of Technology
Bhopal, India

***Abstract*—A fetal is basically an unborn offspring which is in the embryo stage until it comes to the world. During the pregnancy process, each three month period is known by a name called trimester. During this process the fetus grows and develops and along with it the regular checkups are very important. As we all know that a pregnancy lasts for 9 months and in this long period there may be various reasons which may cause disability or mortality in the newborn which is a very severe case and this needs to be avoided. One of the main tool to analyse the health of the fetal in the womb is by doing a CTG(Cardiotacagraphy) which generally is used to evaluate the heart beat and the uterine contractions hence the data generated is used by the doctor to analyse the health and give his wording. But there is a room for error hence the doctors are not reliable to analyse the data hence different machine and deep learning algorithms have been there which can analyse the data and predict the fetals health based on it. The main motive of the paper is to prove the prediction accuracy using the different classification models and compare which model performs better.**

Keywords—Comparative analysis, machine learning,Heart disease prediction, random forest, Classification

I. INTRODUCTION

Fetals are delivered from the women's womb and before that during the pregnancy the information about the fetus is tough to get. We can only get the information that there is a fetus and it would be delivered. So, as the information is not readily available the obstetricians who check the condition of the fetal rely on the indirect information. One of the most dependent information is about the fetal heart rate and there is a restriction of electronic fetal heart monitoring that this variation is hindering the records of the accurate communication and time management.There is an another component known as the cardiotocogram which contains distinct signals and is mainly used for recordings of the fetal heart rate which is the main way through which the obstetricians rely on the information. But the trend seen in these days by the doctors is that there is very high intra and inter observer fluctuation in FHR patterns. But there is a risk in which a falsely diagnosed fetal pain may lead to unnecessary interventions. Hence, the main motive of the research is to employ machine learning algorithms to classify the methods as there can be a room for error by the doctor but the prediction algorithm may perform really well in this case and also help in monitoring the results and give a proper analysis than the doctor could get by his own or someone's observation.

II. LITERATURE OVERVIEW

Some related work has been done in this following topic. In [1] it says that the models lightbgm and bayesian were successful and also gave a small amount of performance gain to the prediction.An adaptive neuro-fuzzy inference system (ANFIS) [2] demonstrated its performance by predicting normal and diseased states from CTG data with 97.2 percent and 96.6 percent accuracy, respectively.In [3] it is saying that comparing the classifiers, random forest and XGBoost are performing well but the dataset used is imbalanced as some modified version is to be used to get the better output in terms of accuracy.In [4] the dataset that has been is used is the CTG data which is observed to be beneficial to identify the abnormalities. The visual analysis along which the decision support system focuses has been made on the machine learning models that have been used. As the machine learning doesn't perform well on the basis of accuracy the ensemble model has been used which has bagged an accuracy of 99.02% after the 10-fold cross validation has been employed. Hence, this can be used to classify the normal and the pathological cases of the ctg data.While in [5] Artificial Neural Network(ANN) along with simple logistic models are used and base on the 10 fold cross validation which has been used to test the data the best results that they have been obtained are 98.47 with the Artificial Neural Network(ANN) model and 98.74 with the Logistic model and hence logistic model has performed better by a minute difference in the accuracy than the Artificial Neural Network(ANN).

III. METHODOLOGY

A. DATA PREPARATION

The data came from the Fetal Health stroke dataset, which has been utilised in various research. Because the data is scarce, the only way to run the model and produce a forecast was to take data from a trusted source. As a consequence, the Fetal dataset was used to create the dataset. This dataset is used to predict foetal health by utilising multiple characteristics such as baseline value,accelerations,fetal movement,uterine contraction,light decelerations, and so on. For Machine Learning and Data Visualization applications, the filtering approach is used to choose a subset of the original train data.The glimpse of the features can be seen in figure 1.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 10 Issue 11, November-2021

```
baseline value
accelerations
fetal_movement
uterine_contractions
light_decelerations
severe_decelerations
prolongued_decelerations
abnormal_short_term_variability
mean_value_of_short_term_variability
percentage_of_time_with_abnormal_long_term_variability
mean_value_of_long_term_variability
histogram_width
histogram_min
histogram_max
histogram_number_of_peaks
histogram_number_of_zeroes
histogram_mode
histogram_mean
histogram_median
histogram_variance
histogram_tendency
fetal_health
```

Figure 1: Dataset Overview

B. DATA PREPROCESSING

A dataset is made up of patterns or entities. A set of characteristics that characterise data items captures an item's basic features, such as the mass of a physical object or the time at which an event happened. So, in order to start improving the dataset quality, the first step would be to remove the null values which probe a problem while retrieving the accuracy of the dataset.After that the description of the dataset is required just to get an idea on it and to get a head start for the feature engineering process which can be seen in the figure 2.

	count	mean	std	min	25%	50%	75%
baseline value	2126.0	133.303857	9.840844	106.0	126.000	133.000	140.00
accelerations	2126.0	0.003178	0.003866	0.0	0.000	0.002	0.006
fetal_movement	2126.0	0.009481	0.046666	0.0	0.000	0.000	0.003
uterine_contractions	2126.0	0.004366	0.002946	0.0	0.002	0.004	0.007
light_decelerations	2126.0	0.001889	0.002960	0.0	0.000	0.000	0.003
severe_decelerations	2126.0	0.000003	0.000057	0.0	0.000	0.000	0.000
prolongued_decelerations	2126.0	0.000159	0.000590	0.0	0.000	0.000	0.000
abnormal_short_term_variability	2126.0	46.990122	17.192814	12.0	32.000	49.000	61.000
mean_value_of_short_term_variability	2126.0	1.332785	0.883241	0.2	0.700	1.200	1.700
percentage_of_time_with_abnormal_long_term_variability	2126.0	9.846660	18.396880	0.0	0.000	0.000	11.000
mean_value_of_long_term_variability	2126.0	8.187629	5.628247	0.0	4.600	7.400	10.800
histogram_width	2126.0	70.445908	38.955693	3.0	37.000	67.500	100.00
histogram_min	2126.0	93.579492	29.560212	50.0	67.000	93.000	120.00
histogram_max	2126.0	164.025400	17.944183	122.0	152.000	162.000	174.00
histogram_number_of_peaks	2126.0	4.068203	2.949386	0.0	2.000	3.000	6.000
histogram_number_of_zeroes	2126.0	0.323612	0.706059	0.0	0.000	0.000	0.000
histogram_mode	2126.0	137.452023	16.381289	60.0	129.000	139.000	148.00
histogram_mean	2126.0	134.610536	15.593596	73.0	125.000	136.000	145.00
histogram_median	2126.0	138.090310	14.466589	77.0	129.000	139.000	148.00
histogram_variance	2126.0	18.808090	28.977636	0.0	2.000	7.000	24.000
histogram_tendency	2126.0	0.320320	0.610829	-1.0	0.000	0.000	1.000
fetal_health	2126.0	1.304327	0.614377	1.0	1.000	1.000	1.000

Figure 2: Dataset Description

C. FEATURE ENGINEERING

Feature engineering plays an important role as everything from the data to the output of the same is dependent on the feature engineering which is being performed. Firstly a pie chart shown in figure 3 is being visualised which shows about the different health types of the fetal and get to know if it is a significant feature or not. While in figure 4 we are getting a correlation matrix of all the features to their significance. The correlation matrix is used in getting the relation of each feature or attribute with itself and also with the other features present out in the database.

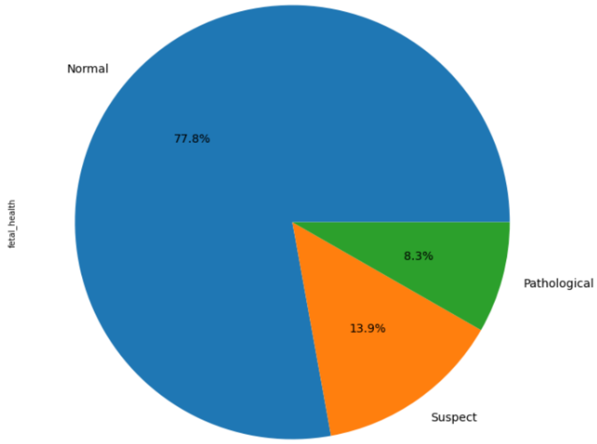

Figure 3: Fetal Health pie chart

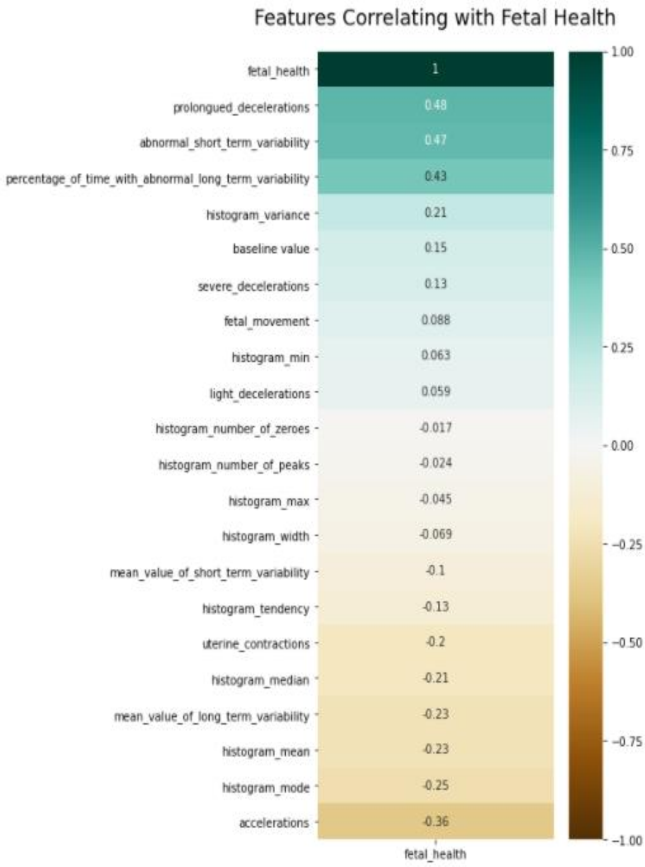

Figure 4: Correlation matrix

D. MODEL ARCHITECTURE

1) RANDOM FOREST CLASSIFIER

Random Forest Algorithm is a supervised learning algorithm which collects samples from different data sets and predicts the best solution. It is forming a Decision Tree like Structure. And It's more accurate than the Decision Tree. But quite slower in prediction and complex in constructing.

2) SUPPORT VECTOR MACHINE

In the Support Vector Machine algorithm the data sets are divided into different parts as support vectors by a Margin and between them there is a space which is called Hyperplane. It is a widely used algorithm and has many direct use cases out there.

3) NAIVE BAYES CLASSIFIER

The Naïve Bayes algorithm is based on Bayes Theorem and is used to solve Probabilistic ML Problems to predict the class of unknown data sets. It is one of the fastest and simplest machine learning algorithms for problem solving.

4) LOGISTIC REGRESSION

In Logistic Regression there are only 2 outcomes.
For Example: True or False, 1 Or 0 .Logistic Regression uses one or more Independent Variables to determine an outcome hence the following feature makes the algorithm faster and better to perform

IV. EXPERIMENTAL RESULTS

As we could see the health feature is the most important attribute and keeping that in mind from the feature engineering results. Keeping that in mind the different classification algorithms are performed such as logistic regression, random forest, svm and naive bayes. After all these models have provided us with the accuracy the authors have come to the results that logistic regression gives about 99.5 percent accuracy while random forest gives 98 percent , svm gives 96.5 percent and Naive bayes gives 97 percent. Hence, after seeing all the accuracy one can know that logistic regression is performing the best.

	Model Name	Accuracy
0	Logistic Regression	0.99524
1	Random Forest	0.98333
2	SVM	0.96543
3	Navive Bayes	0.97320

Figure 5: Accuracy of all models

V. CONCLUSION

Finally, after performing all the steps needed to get the results from preparation to preprocessing to feature engineering and finally performing the models(SVM, random forest, logistic regression and naive bayes) the authors have concluded that the model which performs the best out of all these is the logistic regression model with 99.5 percent accuracy

VI. FUTURE SCOPE

As there is a lot of possibility of improvement in this based on the data as modern real time data can be collected which can be used to test all the different models that are present and to create a new accuracy based on this. Another thing that can be done is to test the model made by the authors and also create a comparison on the new data that is there. The data collection would take a long time hence till then multiple times the data should be collected from different sources.

VII. REFERENCES

[1] medRxiv 2021.06.03.21255808; doi: https://doi.org/10.1101/2021.06.03.21255808
[2] Ocak, Hasan, and Huseyin Metin Ertunc. "Prediction of fetal state fromthe cardiotocogram recordings using adaptive neuro-fuzzy inferencesystems." Neural Computing and Applications 23.6 (2013): 1583-1589.
[3] Piri, Jayashree & Mohapatra, Puspanjali. (2019). Exploring Fetal Health Status Using an Association Based Classification Approach. 166-171. 10.1109/ICIT48102.2019.00036.
[4] Şahin, Hakan & Subasi, Abdulhamit. (2012). Classification of Fetal State from the Cardiotocogram Recordings using ANN and Simple Logistic.
[5] M. Ramla, S. Sangeetha and S. Nickolas, "Fetal Health State Monitoring Using Decision Tree Classifier from Cardiotocography Measurements," 2018 Second International Conference on Intelligent Computing and Control Systems (ICICCS), 2018, pp. 1799-1803, doi: 10.1109/ICCONS.2018.8663047.
[6] A. H. Khandoker, M. Wahbah, R. Al Sakaji, K. Funamoto, A. Krishnan and Y. Kimura, "Estimating Fetal Age by Fetal Maternal Heart Rate Coupling Parameters," 2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), 2020, pp. 604-607, doi: 10.1109/EMBC44109.2020.9176049.
[7] J. Kolarik, L. Soustek and R. Martinek, "Examination and Optimization of the Fetal Heart Rate Monitor : Evaluation of the effect influencing the measuring system of the Fetal Heart Rate Monitor," 2018 IEEE 20th International Conference on e-Health Networking, Applications and Services (Healthcom), 2018, pp. 1-4, doi: 10.1109/HealthCom.2018.8531168.
[8] M. Wahbah, R. Al Sakaji, K. Funamoto, A. Krishnan, Y. Kimura and A. H. Khandoker, "Estimating Gestational Age From Maternal-Fetal Heart Rate Coupling Parameters," in IEEE Access, vol. 9, pp. 65369-65379, 2021, doi: 10.1109/ACCESS.2021.3074550.
[9] S. Modi and M. H. Bohara, "Facial Emotion Recognition using Convolution Neural Network," 2021 5th International Conference on Intelligent Computing and Control Systems (ICICCS), 2021, pp. 1339-1344, doi: 10.1109/ICICCS51141.2021.9432156.
[10] S. Mazumdar, R. Choudhary and A. Swetapadma, "An innovative method for fetal health monitoring based on artificial neural network using cardiotocography measurements," 2017 Third International Conference on Research in Computational Intelligence and Communication Networks (ICRCICN), 2017, pp. 265-268, doi: 10.1109/ICRCICN.2017.8234518.
[11] D. D. Fong et al., "Design and In Vivo Evaluation of a Non-Invasive Transabdominal Fetal Pulse Oximeter," in IEEE Transactions on Biomedical Engineering, vol. 68, no. 1, pp. 256-266, Jan. 2021, doi: 10.1109/TBME.2020.3000977.
[12] V. Shah and S. Modi, "Comparative Analysis of Psychometric Prediction System," 2021 Smart Technologies, Communication and Robotics (STCR), 2021, pp. 1-5, doi: 10.1109/STCR51658.2021.9588950.
[13] M. B. I. Reaz and Lee Sze Wei, "An approach of neural network based fetal ECG extraction," Proceedings. 6th International Workshop on Enterprise Networking and Computing in Healthcare Industry - Healthcom 2004 (IEEE Cat. No.04EX842), 2004, pp. 57-60, doi: 10.1109/HEALTH.2004.1324471.
[14] L. H. Lee and J. A. Noble, "Automatic Determination of the Fetal Cardiac Cycle in Ultrasound Using Spatio-Temporal Neural Networks," 2020 IEEE 17th International Symposium on Biomedical

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 10 Issue 11, November-2021

Imaging (ISBI), 2020, pp. 1937-1940, doi: 10.1109/ISBI45749.2020.9098613.

[15] P. Dwivedi, A. A. Khan, S. Mugde and G. Sharma, "Diagnosing the major contributing factors in the classification of the fetal health status using cardiotocography measurements: An AutoML and XAI approach," 2021 13th International Conference on Electronics, Computers and Artificial Intelligence (ECAI), 2021, pp. 1-6, doi: 10.1109/ECAI52376.2021.9515033.

[16] 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