Fetal Health Classification Using Machine Learning

Het Patel
(3rd yr.) B.tech in CSE
Inst. of Tech., Nirma University
Ahmedabad, India
ORCID ID. : Γ

Jal Patel
(3rd yr.) B.tech in CSE
Inst. of Tech., Nirma University
Ahmedabad, India
ORCID ID. : Γ

Meet Goti
(3rd yr.) B.tech in CSE
Inst. of Tech., Nirma University
Ahmedabad, India
EMAIL ID.

***Abstract*—A CTG (cardiotocography), which is typically used to assess the heartbeat and uterine contractions, is one of the primary methods used to assess the health of the fetus while it is still inside the mother. The data generated by the CTG is then used by the doctor to assess the fetus's condition and provide his or her opinion. But because there is space for error, multiple machine and deep learning algorithms have been developed that can analyze the data and make health predictions for foetuses based on it. The primary goal of this literature is to demonstrate prediction accuracy using various classification models and to assess which model is more effective.we have used three classification techniques to make the classification.They are:1)Logistic regression 2)K-Nearest Neighbour and 3)Random Forest classifier. We have pre-processed the data by scaling and generalising it.The grid search is used to identify the optimal hyper-parameters for a model.The model is trained and the classification report is generated.Then the final classification technique is chosen to predict the state of seriousness.**

***Index Terms*—CTG(cardiotocography), Measles, Mumps and Rubella(MMR), FHR(Fetal Health Rate), KNN(K-Nearest Neighbour), RF(Random Forest), F1 Score, Learning Curve, Grid-Search CV.**

I. Introduction

A fetus is an unborn child that develops from an embryo. There are three stages of fetal growth: 1) Germinal, 2) Embryonic, 3) Feta [1]. In the domain of prenatal health, the extraction of valuable information from health data such as CTG is a key problem. The heart rate of the fetus (known as FHR) and uterine contractions are monitored by CTG during labor to assess the health of the mother and child [1]. Here, we employed a multi-objective evolutionary approach to pick the traits that had the greatest influence on fetal health. Multiple theories and optimum solutions are produced as a result of multi-objective feature selection techniques operating on a population of solutions instead of a single one [1]. In order to break a tie and select one subset of features as the ideal set of features, there must be a third criterion in addition to the number of features and accuracy when there is a tie among all non-dominated solutions [1].

One of the most difficult and intricate procedures in medicine is tracking fetal growth during pregnancy. Even if preventative precautions have been implemented, approximately 810 pregnant women still pass away every day, according to the World Health Organization (WHO, 2021) [2]. Monitoring of the fetus's health is done during pregnancy. The health of the mother has no bearing on fetal growth. Cardiotocography is used to continuously measure the health and growth rate of the fetal to prevent such issues. The goal of the cardiotocography is to monitor the fetal heartbeat and gauge the mother's hormonal changes simultaneously [2]. Cardiotocographic (CTG) results will show the mother's uterine contractions as well as the fetus's heart rate, acceleration and deceleration patterns, and other intricate measurements. The normal, suspicious, and phases of the fetal can be classified using a variety of machine learning (ML) techniques. The findings indicate that the automated system used to analyse the fetal period health will be built around the ML technique [2]. Medical professionals can reduce the MMR and high labor complications by using machine learning (ML) techniques to aid in early decision-making in difficult situations like diagnosis [2]. Although classifying the stages of fetal health is a difficult task, ML classification techniques excel at handling it (Arif, 2015) [2]. Logistic Regression, KNN, and random forests are some of the common classification techniques (Quilligan Paul, 1975) [2]. The RF classifier performs better and with greater accuracy when classifying the stages of fetal health [2].

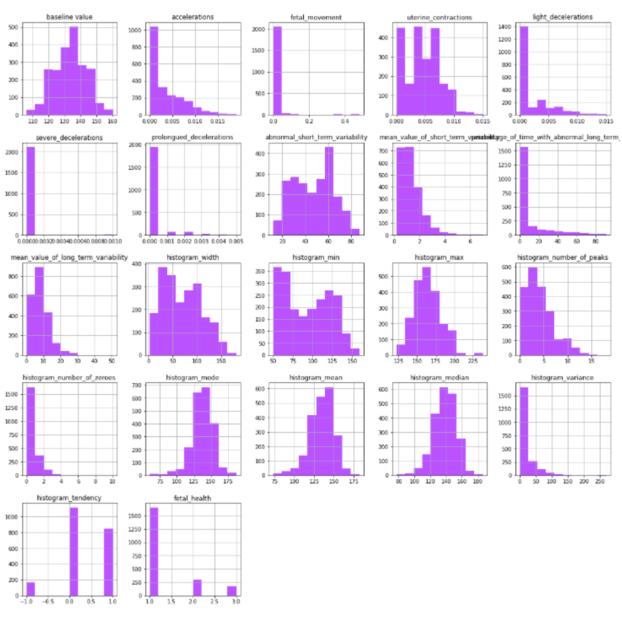

Fig. 1. Histogram plot between all Features and Frequencies

II. Dataset Analysis

The histogram for the entire dataset is displayed in Fig. 1 where the relation between frequency and features shown for each using a histogram [3]. The fetal data is provided for 2126 women, with more than 455 uterine contractions, and other significant distributional characteristics on which the model will be trained on further [3].

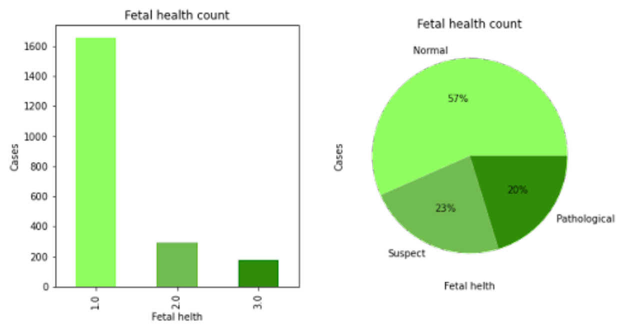

Fig. 2. The Model Flow used for Classification using ML models

Fig 2 shows the number of entries related to each class. 1212 are of class 1(Normal), 487 of class2 (Suspect) and 427 of class3(Pathological).

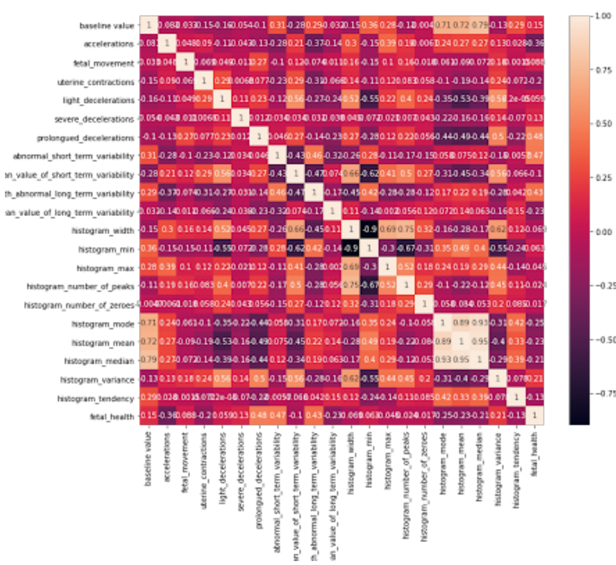

Fig. 3. Correlation Heatmap Among Features of Dataset

Fig. 4 depicts the flow of the entire Machine Learning model which will be used for fetal health classification and the flow of this paper.

III. Data Pre-processing

A. Feature Scaling

Feature scaling in pre-processing of data is very critical task before working on a machine learning model. Feature scaling change the strengthen of machine learning model extensively. Normalisation and Standardisation are two frequent methods used for feature Scaling. Normalisation is used when we want

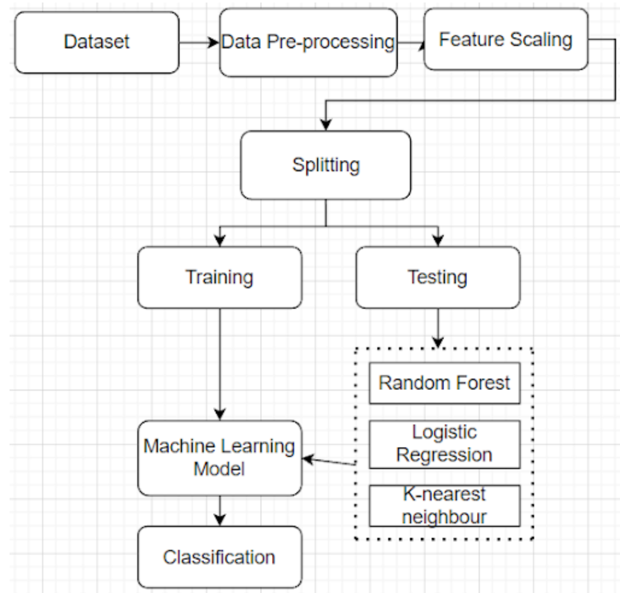

Fig. 4. The Model Flow used for Classification using ML models

to limit the range of our data to two numbers, often [0, 1] and [-1, 1] while Standardisation changes our information so that it has an average of 0 and a variance of 1, which makes it unitless. Refer to Fig. 5, which depicts how data appears after being scaled in the X-Y plane.

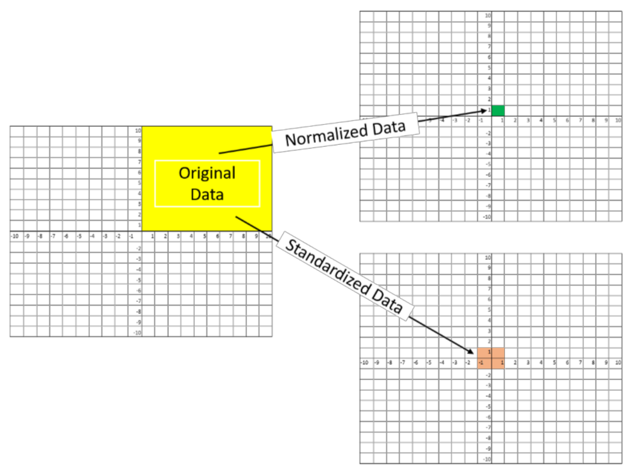

Fig. 5. Scaling Data by two ways :Normalisation and Standardisation [4]

B. Data Train-Test-Split

Train Test Split : We must separate a dataset into test and training sets in order to evaluate how well our algorithm for ML performs. It is possible to use the train set's statistics, which are used to make the model fit. Only predictions are made using the second batch of data, the testing data.

Cross-validation : A technique for analysing ML models that entails training several ML models on portions of the data input before contrasting the outcomes with another portion of the data. Widely, there are two types of Cross Validation(CV) techniques available for Hyper-parameter tuning: GridSearchCV and RandomSearchCV. For the current models the cross validation technique used here is GridSearchCV. The reason behind this is because GridSearchCV tries out all possible combinations of Hyper-parameters passed and chooses the best possible combinations of Hyper-parameters based on CV score. While, random search is faster but uses random combinations within a particular range provided for parameters and does not guarantee best parameters and is suitable for very large datasets.

IV. Classification Techniques

In this model, we have used basically 3 kinds of Classification Techniques:

1) Logistic Regression
2) K Nearest neighbours
3) Random Forest

A. Logistic Regression

Fig. 6. Logistic Regression Flow Chart [3]

Logistic regression is a method of forecasting which makes use of a number of independent factors to determine the result of a categorical dependent variable [3]. Moreover, output must therefore be discrete or categorical such as: true or false, yes or no, 0 or 1, and so on are all possible outcomes. However, probabilistic values (between 0 and 1) are provided rather than exact values between 0 and 1 [3]. Instead of fitting a regression line as in linear regression, logistic regression should be used to forecasts two maximum values (0 or 1) [3]. The probability of anything, such as whether or not cells are malignant or whether or not a rodent is overweight, is represented by the logistic function's curve. It is a well-liked machine learning technique since it can provide predictions and categorise new information using both continuous and discrete datasets. [3].

Hyper-parameters associated with this model are : "tol"(Tolerance of Stopping Criteria), "C", "InterceptScaling".

B. K-Nearest Neighbours

One such supervised learning method that may be applied to both regression and classification analyses is KNN. By calculating the distance between the test data and all of the training points, KNN tries to predict the correct class for the test data. Then choose the K points that are closest to the test data. The KNN algorithm determines the likelihood that the test data fall into each of the "K" classes of training data, and the class with the highest similarity is chosen.

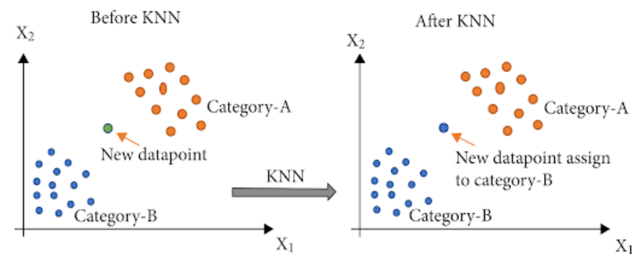

Fig. 7. KNN Flow Chart [3]

A supervised learning method named K-nearest neighbours (KNN), may be applied to both classification and regression analysis. KNN aims to predict the right class for the test data by estimating the distance between the testing data and each of the training points. then decide which K points are most near the test data. The KNN method calculates the probability that the test data will belong to each of the "K" training data classes, and it then selects the group with the highest degree of similarity as illustrated in figure 7.

Hyper-parameters associated with this model are : "leaf-size", "nneighbors", "p"(Euclidean or Manhattan distance).

C. Random Forest

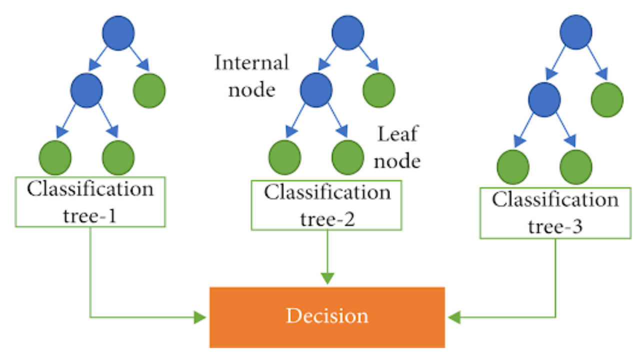

Fig. 8. Random Forest Flow Chart [3]

One of the machine learning method for supervised learning is the random forest. It constructs a "forest" out of a collection that have primarily been prepped for the "bagging" method. Fundamentally, the bagging approach is acceptable since integrating many learning models enhances the result. To produce a more precise and trustworthy representation, the random forest generates a vast number of distinct trees before combining them. It benefits from resolving the arrangement and relapse problems that most existing ML frameworks experience. Another interesting aspect of the random forest approach is how simple it is to determine the general importance of each estimate component. Adaptability is one of random forest's most appealing qualities. It can be used for relapse detection as well as for grouping tasks, it is clear how much weight is generally given to information features. It is also a good strategy because the default hyper parameters it uses frequently produce clear expectations. Since there aren't many hyper parameters to begin with, understanding them is essential. Although,Overfitting is a common problem in ML, yet when employing the randomised random forest classifier, it seldom occurs. If the forest has a sufficient number of trees, the classifier won't overfit the model.

Hyper-parameters associated with this model are : "minsamplessplit","minsamplesleaf", "nestimators", "criterion"(entropy or gini).

V. Comparison of Models

A. Classification Report Analysis

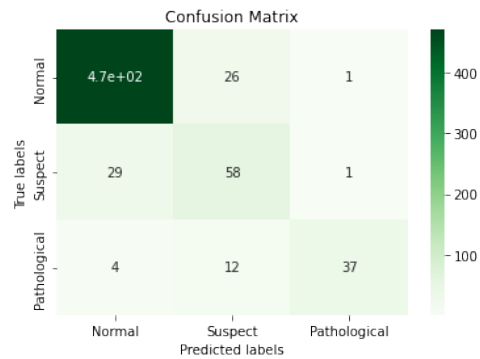

Fig. 9. Logistic Regression Confusion Matrix

Classification Report

	precision	recall	f1-score	support
1.0	0.93	0.95	0.94	497
2.0	0.60	0.66	0.63	88
3.0	0.95	0.70	0.80	53
accuracy			0.89	638
macro avg	0.83	0.77	0.79	638
weighted avg	0.89	0.89	0.89	638

Fig. 10. Logistic Regression Classification Report

1) Logistic Regression: Fig. 10 displays the classification report of LR model's . In this instance, an overall F1 score of 79% was attained. An individual F1-score of 94%, 63%, or 80% is assigned to a pathogenic diagnosis. The confusion matrix 9 also displays the anticipated outcome in addition to the computed performance of the model. While 73 predictions were incorrect, 565 were totally correct.

2) K-Nearest Neighbours: The classification report from the K-nearest neighbour model is displayed in Fig. 12. The overall F1 score in this case is 75.33%. Individual F1-scores range from 94% for normal to 57% for abnormal, with 75% being suspicious. The confusion matrix 11 shows both the anticipated outcome and the model's computed performance. There were 78 inaccurate guesses out of a total of 560 right ones.

3) Random Forest: As per the Fig. 14. The overall F1-score in this case is 86.33%. The individual F1-score for normal is 96%, suspected is 76%, and abnormal is 87%. As the number of trees increases, the error%age decreases. In this scenario, 100 trees provide the most accuracy with entropy as best criteria. The prediction produced by the random forest model

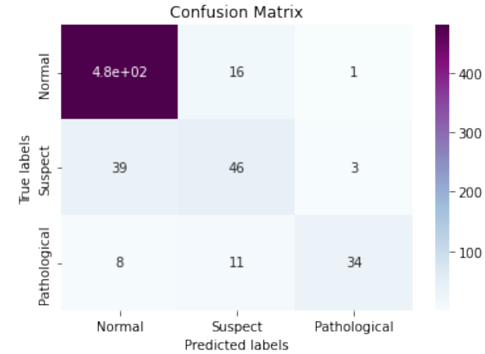

Fig. 11. K Nearest Neighbour Confusion Matrix

Classification Report

	precision	recall	f1-score	support
1	0.91	0.97	0.94	497
2	0.63	0.52	0.57	88
3	0.89	0.64	0.75	53
accuracy			0.88	638
macro avg	0.81	0.71	0.75	638
weighted avg	0.87	0.88	0.87	638

Fig. 12. K Nearest Neighbour Classification Report

are displayed in 13 The confusion matrix shows both the anticipated outcome and the model's computed performance. There were 45 inaccurate forecasts, for a total of 593 correct predictions.

B. Learning Curve Analysis

Learning Curve is a relationship that appears as a direct proportion on a graph between learner's success on a task and the number of iterations or tries needed till we get optimum result. The model probably won't benefit from additional data if the training and cross-validation scores converge when more data is supplied. The model definitely needs more training examples in order to more effectively generalise if the training score is significantly higher than the validation score. The curves are drawn using the average scores, while variation during cross-validation is displayed using the shaded areas, which stand for a standard deviation above and below the average for all cross-validations. More fluctuation around the training score curve is probably to be expected if the model has bias-related inaccuracy. There will be even more variance in the cross-validated score if the model has errors caused by variance.

The Learning Curves for the above worked upon models are provided as:

From the study of models we finally come to a verdict that based upon the score chart we can clearly see that Random Forest has provide the best possible accuracy among the three 18.

VI. Conclusion

Obstetricians can identify fetal defects and choose a course of treatment using CTG data, which enables doctors to prevent

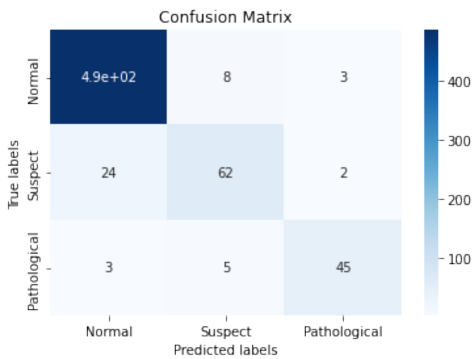

Fig. 13. Random Forest Confusion Matrix

Classification Report

	precision	recall	f1-score	support
1	0.95	0.98	0.96	497
2	0.83	0.70	0.76	88
3	0.90	0.85	0.87	53
accuracy			0.93	638
macro avg	0.89	0.84	0.87	638
weighted avg	0.93	0.93	0.93	638

Fig. 14. Random Forest Classification Report

lasting injury to the unborn child. The visual analysis of the CTG results, however, sometimes cannot not be fair, hence it is becoming more and more common practise in medicine to use decision support systems to recognise and foresee abnormal circumstances. In this study, the identification of prenatal risks was the primary focus using CTG data. In order to identify prenatal malformations, ML models may be employed in conjunction with the CTG dataset as a decision support system. On the other hand, in our study, we incorporated a number of popular ML techniques. We have used balanced accuracy as scoring measure due to imbalanced dataset. The most accurate algorithms with balanced accuracy as a scoring measure were the random forest, logistic regression and K-nearest neighbour with 92.94%, 88.56% and 87.78% accuracy respectively. The robustness of the models has been demonstrated by several model comparisons, and the research analysis can be used to infer the scheme. To strengthen this system, various intricate machine learning models may be used in the future.

REFERENCES

[1] Automatic Classification of Fetal Heart Rate Based on Convolutional Neural Network Jianqiang Li , Zhuang-Zhuang Chen, Luxiang Huang, Min Fang, Bing Li, Xianghua Fu, Huihui Wang , Senior Member, IEEE, and Qingguo Zhao

[2] Mehbodniya, A., Lazar, A. J. P., Webber, J., Sharma, D. K., Jayagopalan, S., K, K., Singh, P., Rajan, R., Pandya, S., Sengan, S. (2021). Fetal health classification from cardiotocographic data using machine learning. Expert Systems, e12899. https://doi.org/ 10.1111/exsy.12899

[3] Md Takbir Alam, Md Ashibul Islam Khan, Nahian Nakiba Dola, Tahia Tazin, Mohammad Monirujjaman Khan, Amani Abdulrahman Albraikan, Faris A. Almalki, "Comparative Analysis of Different Efficient Machine Learning Methods for Fetal Health Classification", Applied Bionics and Biomechanics, vol. 2022, Article ID 6321884, 12 pages, 2022. https://doi.org/10.1155/2022/6321884

[4] Image Source

Fig. 15. Logistic Regression Learning Curve

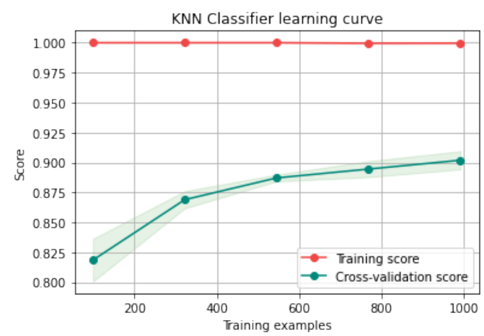

Fig. 16. KNN Learning Curve

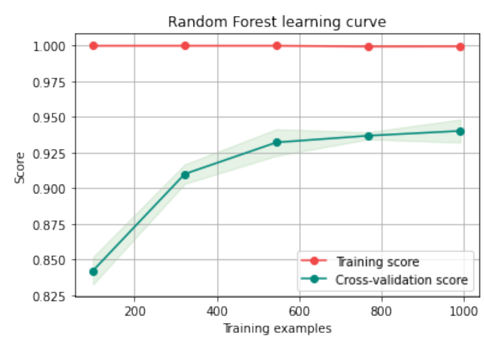

Fig. 17. Random Forest Learning Curve

	Model
Score	
0.929467	Random Forest
0.885580	Logistic Regression
0.877743	KNN

Fig. 18. K Nearest Neighbour Confusion Matrix