Predicting the Fetal Health Status from CTG features using Machine Learning: A Dual Approach

Muntaqim Ahmed Raju
MuntaqimAhmed_Raju@student.uml.edu
University of Massachusetts Lowell

Priyanka Siddappa
Priyanka_Siddappa@student.uml.edu
University of Massachusetts Lowell

Raul Olivares
Raul_Olivares@student.uml.edu
University of Massachusetts Lowell

Kush Patel
Kush_Patel3@student.uml.edu
University of Massachusetts Lowell

Nicholas Miceli
Nicholas_Miceli@student.uml.edu
University of Massachusetts Lowell

Abstract

Cardiotocography (CTG) is a technique used for data collection in fetal health. It is employed to analyze the health of a fetus during pregnancy and determine whether the fetus requires medical attention. CTG mainly yields two results fetal health rate (FHR) and uterine contractions (UC). In total, there are 21 attributes in the measurement of FHR and UC on CTG. These attributes can help obstreticians to clasify whether the fetus health is normal, suspected, or pathological. This study implements six different algorithms: Artifical Neural Networks, ADABoost, XGBoost, LGBoost,Random Forest and SVM in order to predict the health of a fetus and decide on a classification between Normal, Suspect, and Pathological. The implementation includes a Multiclass classifier using those three classes and a Binary classifier (Normal or not Normal) to demonstrate that a binary classifier can achieve a higher accuracy on this dataset and therefore allow doctors to prioritize patients in need with more confidence.By employing three scenarios, this paper reports the performance measurements among those algorithms. The study shows that 5 out of 6 algorithms perform very well using SMOTE Technique (85-98% accurate) and not as well using Ensemble Technique (58-77%accurate).

Keywords: Cardiotocography; Fetal health; ANN; XGBoost; LightGBM; SVM; AdaBoost; Random Forest Classifier; Multi-class Classifier; Binary-class Classifier, SMOTE, Ensemble

I. INTRODUCTION

The phase of pregnancy is a unique stage in a woman's life that demands thorough health monitoring to ensure the optimal growth of the unborn child. A variety of trimester-specific laboratory tests, including the cardiotocogram (CTG), are recommended. Cardiotocography (CTG) is one method used to do so which involves collecting data from the fetus regarding its Fetal Heart Rate (FHR) and the number of Uterine Contractions (UC). Typically, the CTG data must be analyzed by an Obstetrician in order to detect any abnormal values that may suggest an issue with the pregnancy. In many parts of the world, the number of Obstetricians is limited, and they may not have the available resources needed to meet with each patient individually. Therefore, any techniques which allow doctors to streamline the process and prioritize the patients with the direst needs can be invaluable to ensuring the highest possible rate of successful pregnancies.

II. Background

Cardiotocograph (CTGs) can be used as a monitoring tool to identify high-risk women during labor. The data instances from the CTG are classified as normal, suspicious, or pathological. The purpose of our dual approach is to demonstrate that the accuracy of machine learning algorithms can be enhanced by consolidating the categories of 'suspicious' and 'pathological' into a single class, termed 'not normal'. We can increase the accuracy of machine learning algorithms by implementing a binary classifier rather than a multi-class one. The improved accuracy of this binary classifier could be a valuable tool for doctors in resource-limited countries. It would enable them to prioritize patients classified as 'not normal' over those deemed 'normal', thereby ensuring that those most in need receive the highest level of attention.

III. Problem Statement

Our research aims to address the issue of perinatal mortality by predicting fetal health status using cardiotocography (CTG) features and machine learning. Our objective is to provide highly accurate and consistent diagnoses for assessing fetal health complications during pregnancy. This effort aims to reduce both fetal mortality and maternal mortality rates. The main challenge we encountered was balancing the imbalanced dataset. We explored traditional methods like under sampling, oversampling, and combinations of these techniques along with ensemble methods. However, due to the small size of the dataset, under sampling was not feasible. So, we opted for oversampling using the SMOTE technique, a conventional method to generate more data. Also, we compared SMOTE with ensemble techniques to understand how performance varied. Our analysis involves predicting fetal health across multiple classes such as normal, pathological, and suspect -and a binary classification of normal versus not normal fetal health statues.

IV. Related Work

In recent years, studies have been completed utilizing this dataset along with machine learning to classify new instances. Notably, "Comparison of machine learning algorithms to classify fetal health using cardiotocogram data" by Pahlawan et al. [2] explored several different algorithms to classify data, but only using a multi-class classifier. For the deep learning methods tested in their study (ANN and Long-Short Term Memory) the accuracy achieved was low and thus they concluded that deep learning was not suitable for this high-dimensional dataset.

Another study titled "Use of machine learning algorithms for prediction of fetal risk using cardiotocographic data" by Chowdhury et al. [3] implemented their own multi-class classifier using various algorithms and for the algorithms of Support Vector Machine and ADABoost achieved a testing accuracy of approximately 90 percent. Therefore, we aim to improve upon these accuracies used in related works by implementing a binary classifier.

V. Methodology

The dataset utilized for our study is the Cardiotocography Dataset from the University of California Irvine Machine Learning Repository [1]. This dataset contains 2126 records of Cardiotocogram data collected during pregnancies. This dataset contains 21 attributes used in the measurement of FHR and UC on CTG in Figure 1. The data has been reviewed by three expert Obstetricians and each fetus has been classified. 70% of them were Normal (No issues found), 20% were Suspicious (Some abnormalities, may need intervention), and 10% had a pathological fetal state (Significant abnormalities, needs intervention). The dataset was used in a Multiclass classifier using those three classes and a Binary classifier (Normal or not Normal) to demonstrate that a binary classifier will have a higher prediction accuracy on the data.

1	Baseline Value (FHR)
2	Accelerations
3	Fetal Movement
4	Uterine Contractions
5	Light Decelerations
6	Severe Decelerations
7	Prolongued Decelerations
8	Abnormal Short Term Variability
9	Mean Value of Short Term Variability
10	% of Time with Abnormal Long Term Variability
11	Mean Value of Long Term Variability
12	Histogram Width
13	Histogram Min
14	Histogram Max
15	Histogram Number of Peaks
16	Histogram Number of Zeros
17	Histogram Mode
18	Histogram Mean
19	Histogram Median
20	Histogram Variance
21	Histogram Tendency

Figure 1: 'Dataset Features'

Since the dataset used in this project was imbalanced, preprocessing was necessary. Machine learning models built on a highly skewed class distribution, where one majority class dominates the minority classes, can lead to several problems, and one of the alarming problems is bias towards the majority class. The following figure 2 shows the distribution of classes (normal, suspect, and pathologic) in the dataset before it was balanced with SMOTE.

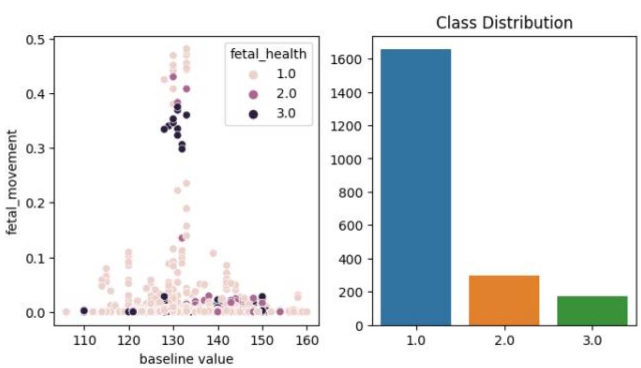

Fig 2: Class distribution before oversampling

To solve the imbalance in class distribution, we employed the SMOTE oversampling technique since it was effective previously on this dataset [4]. We performed 100% oversampling to create duplicate sets of data for classes 2 and 3. The following figure 3 shows the class distribution after oversampling.

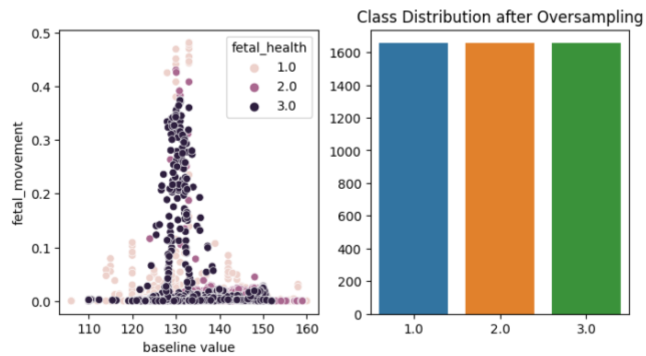

Fig 3: Class distribution after oversampling

For binary class prediction, we transformed multi-class problems into binary-class problems by merging suspect and pathologic classes into a single class labeled as 'not-normal'. Thereafter, we were only left with two classes: 'normal' and 'not-normal'.

We also took the ensemble technique into consideration to assess the differences in performance between the formal and later. In the ensemble technique for multi-class, we split class 1, which is normal, into 7 different chunks, with each chunk containing 206 instances. Following this, we merged each chunk with the other two classes (suspect with 155 instances and pathologic with 100 instances) to form seven different mini batches. The remaining 213, 140, and 76 instances were used to form a testing batch from the normal, suspect, and pathologic classes, respectively. Similarly for binary-class, we split class 1, which is normal, into 5 different chunks, with each chunk consisting of 280 instances. Next, we merged each chunk with the other class (not-normal with 260 instances) to form five different mini batches. The remaining 255, 211 instances were used to form a testing batch from the normal and not-normal classes, respectively.

The machine learning algorithms used in this study were Artifical Neural Networks, ADABoost, XGBoost, LightGB, Random Forest, and SVM in order to predict the health of a fetus and decide on a classification between Normal, Suspect, and Pathological. The implementation includes a Multiclass classifier using those three classes and a Binary classifier (Normal or not Normal) to demonstrate that a binary classifier can achieve a higher accuracy on this dataset and therefore allow doctors to prioritize patients in need with more confidence.

VI. Experiments

In our experimental approach, we aimed to classify fetal health states using Cardiotocograms (CTG) based on fetal heart rate (FHR) and uterine contraction (UC) features. The objective was to develop predictive models for both multi-class (Normal, Suspicious, and Pathological) and binary-class (Normal, Not-normal) scenarios. To achieve this, we commenced with a comprehensive exploratory data analysis, identifying patterns, detecting outliers, and establishing variable relationships within the dataset. Following the exploratory phase, we conducted an extensive review of CTG literature, gathering insights from blogs, peer-reviewed articles, and past analyses to uncover key factors influencing fetal well-being.

Our proposed approach involved feature engineering to extract relevant features from raw data, which were subsequently scaled using normalization techniques to ensure consistency. Different feature selection techniques were employed to identify correlations between input and output features. Multiple machine learning algorithms, including Support Vector Machine (SVM), Extreme Gradient Boosting (XGB), Light Gradient Boosting (LGBM), Adaptive Boosting (ADA Boosting), Artificial Neural Network (ANN), and Random Forest, were utilized to develop individual predictive models, as well as ensemble models based on individual algorithms. The hyper-parameter optimization technique was employed to fine-tune each model, ensuring optimal performance. To evaluate model efficiency, we used various metrics, such as accuracy, precision, recall, and F1-score. Finally, a detailed comparative analysis was conducted to identify the most effective models.

VII. Results

In comparing our results with existing works, our multi-class model with SMOTE showcased an accuracy of 84.89%, while the binary-class model with SMOTE achieved 89.53% accuracy. Noteworthy variations were observed when compared to other studies. For instance, Chowdhury et al. [3], utilized SMOTE but did not employ ensemble methods or binary adaptations, leading to lower accuracy rates, particularly in testing scenarios. Chidambaram and Joy [5], despite employing ensemble learning, utilized different algorithms for training, resulting in an 88% accuracy for the support vector classifier. In contrast, our ensemble models, whether with or without outliers, demonstrated competitive accuracy rates in both multi-class and binary-class scenarios.

Bhowmik et al. [6], achieved a high accuracy using ensemble methods, but their approach did not incorporate SVM or binary adaptations. Similarly, Rayhan et al. [7] achieved accuracies of 84.1% and 83.3% using selected and all features, respectively, without employing ensemble methods or binary adaptations. The differences in our approach, especially the utilization of binary-class scenarios and ensemble methods, showcase the robustness and versatility of our predictive models.

When utilizing Artificial Neural Networks our model was able to achieve accuracy 95% and 97% for multiclass with SMOTE oversampling and binary class with SMOTE oversampling respectively (Figure 1). Compared to related works where a maximum accuracy of 37% was achieved on the dataset [2], our results are much more suitable classifying the data.

Using ANN with ensemble techniques, we received an accuracy of 70% multiclass and 64% for binary class (Figure 2). Overall, these results are similar to results achieved in related works when using ensemble techniques to handle balancing data, but not accurate enough for a medical application.

For the Random Forest algorithm, our model was able to get an accuracy of 97% for the multiclass and 96% for the binary class using the SMOTE technique. The related works also has very similar results of around 97% meaning that the model is suitable for classifying the data.

For the ensemble techniques, the Random Forest got 72% and 77% for the multi class and binary class, respectively. The related works we reviewed got much higher accuracies with some being around 95%. This means the ensemble technique in our model is not suitable for doctors to use it.

Using XGBoost and LightGBM provided us with the highest accuracy of any of the models. Both delivered 98% and 99% accuracy respectively for Multiclass with SMOTE and Binary class with SMOTE. Using the ensemble method, XGBoost achieved 65% for multiclass and 67% for binary class. LightGBM with the ensemble method yielded 63% and 62% for multiclass and binary class.

Finally, our ADABoost implementation received 90% accuracy for multiclass which is comparable to the results achieved by Chowdhury et al. in their study [3]. However, for the binary classifier we achieved 95% which shows significant improvement. Using the ensemble method, ADABoost received 58% accuracy for multiclass and 61% for binary class.

Overall, our results suggest that utilizing a binary classifier on the dataset rather than a multiclass one can help to increase accuracy, and with a sensitive domain such as medical research, the highest possible accuracy is imperative. Using SMOTE oversampling, 5 out of 6 models showed an accuracy increase with a binary classifier.

On the other hand, using the ensemble method to balance the data gave accuracies in the range of 58% to 77%, none of which are suitable for a medical domain. This could be due to the size of the dataset, and each individual weak learner not having enough data to achieve high training accuracy on their own.

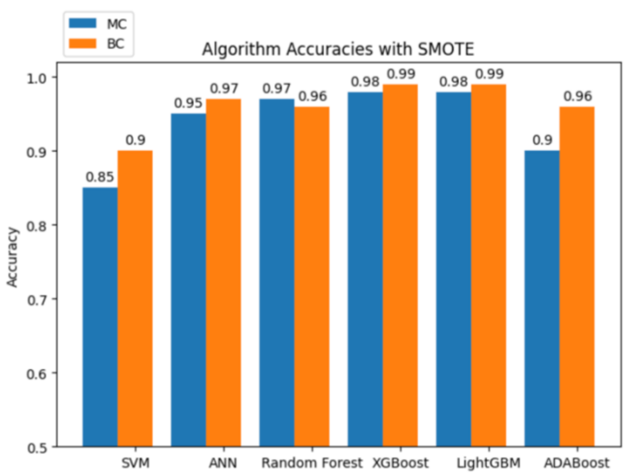

Figure 2

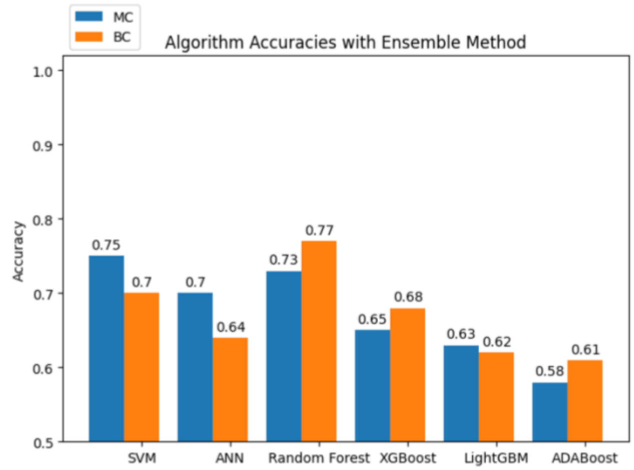

Figure 3

	SMOTE-MC	SMOTE-BC	Ensemble-MC	Ensemble-BC
SVM	0.85	0.9	0.75	0.69
ANN	0.95	0.97	0.7	0.64
Random Forest	0.97	0.96	0.73	0.77
XGBoost	0.98	0.99	0.65	0.68
LightGBM	0.98	0.99	0.63	0.62
ADABoost	0.9	0.96	0.58	0.61

Figure 4: "Comparison of Algorithm's Accuracies"

VIII. Future Work

These results emphasize the potential real-world applications of our model, particularly in low-resource settings where a finely tuned fetal health classification model could significantly expedite the diagnosis process, especially in regions with limited access to skilled obstetricians. The robustness and competitive performance of our models across various scenarios underscore the viability and potential contribution of our proposed approach to the field of fetal health prediction using machine learning.

One of the main issues encountered during the study was deciding how to balance the dataset. We used the SMOTE technique which yielded high accuracy but relies on synthesized data which is not ideal for a medical application. Using ensemble techniques does not synthesize data but also did not give good enough results; therefore, future work on the project would be to determine a novel method of balancing the dataset which will provide high accuracy while also not synthesizing data instances.

Additionally, future work will be done to add explainability to our models. This addition will help doctors reviewing the model's output because not only will it provide the classification but also indicate why the decision was made.

IX. Conclusion

To summarize, our study utilized the Cardiotocography dataset from UC Irvine [1] to implement classifier models to predict fetal health. We used the algorithms SVM, ANN, Random Forest, XGBoost, LightGBM, and ADABoost to implement our models as well as SMOTE and Ensemble methods to handle balancing the dataset.

Our goal was to implement a multi class and binary class classifier for each combination of algorithm and balancing technique to determine if a binary classifier can achieve higher accuracy on this dataset overall.

Our results indicated that a binary classifier was more accurate in predicting fetal health, which supports our original hypothesis. With these models having higher accuracy, doctors can have higher confidence levels in their predictions and can better utilize machine learning to prioritize patients in the most need.

References

[1] Campos, D. and Bernardes, J. (2010). Cardiotocography. UCI Machine Learning Repository. https://doi.org/10.24432/C51S4N.

[2] Pahlawan, M. Pradani, H. Rahmayanti, N. Vinarti, R. (2022). Comparison of machine learning algorithms to classify fetal health using cardiotocogram data, Procedia Computer Science, Volume 197, Pages 162-171, ISSN 1877-0509, https://doi.org/10.1016/j.procs.2021.12.130.

[3] Chowdhury, D. Hasan, B. Hoodbhoy, Z. Nasim, A. Noman, M. Shafique, A. (2019). Use of Machine Learning Algorithms for Prediction of Fetal Risk using Cardiotocographic Data. International Journal of Applied and Basic Medical Research, Volume 9, Issue 4, Pages 226-230, https://doi.org/10.4103/ijabmr.IJABMR_370_18

[4] Alam, M.T.; Khan, M.A.I.; Dola, N.N.; Tazin, T.; Khan, M.M.; Albraikan, A.A.; Almalki, F.A. Comparative Analysis of Different Efficient Machine Learning Methods for Fetal Health Classification. *Appl. Bionics Biomech.* **2022**, *2022*, 6321884.

[5] Chidambaram, M., & Joy, J. (2022). Comparative Analysis of Ensemble Learning Methods for Enhancing Fetal Health Prediction Using Cardiotocography.

[6] Bhowmik, P., Bhowmik, P. C., Ali, U. A. M. E., & Sohrawordi, M. (2021). Cardiotocography Data Analysis to Predict Fetal Health Risks with Tree-Based Ensemble Learning.

[7] Rayhan, M. T., Arefin, A. S. M. S., & Chowdhury, S. A. (2021). Automatic detection of fetal health status from cardiotocography data using machine learning algorithms.

Appendix

We all had individual tasks from the beginning of the study. Muntaqim Ahmed Raju worked on data pre-processing. The algorithms he worked on were XGBoost and LightGB. Nicholas Miceli worked on reviewing the literature papers and making the presentation slides. The algorithms he worked on were Artificial Neural Networks. Raul Olivares worked on reviewing the literature papers. The algorithm he worked on was Support Vector Machines. Priyanka Siddappa worked on literature review and comparative analysis. The algorithm she worked on was AdaBoost. Kush Patel worked on data processing. The algorithm he worked on was Random Forest. In addition to our individual tasks we all worked together on the methodology and algorithm development, writing this paper, and assisted Nick in creating the slides.