CARDIOTOCOGRAM ANALYSIS TO IMPROVE MATERNAL AND FETAL HEALTH OUTCOMES THROUGH MACHINE LEARNING

Priya Kaur
Pavan Marturu
Likithasridhar
Department of Applied Computing, Michigan Technological University

Abstract— **Our study employs machine learning (ML) techniques to develop a robust classification model, meeting the critical need for accurate fetal health assessment during prenatal care. Given the limitations of manual interpretation prone to errors, our objective is to enhance and automate diagnostics through computational algorithms. We analyze a comprehensive dataset encompassing physiological indicators from prenatal exams, such as fetal heart rate, movements, and uterine contractions. Leveraging these features, we apply various ML methods including Decision Trees, AdaBoost, Logistic Regression, and Random Forest to classify fetal health. Through meticulous model selection, including hyperparameter tuning and cross-validation, we identify Random Forest as the most effective model, achieving a remarkable 96%accuracy on the test set. Our findings underscore the potential of ML-based approaches to improve prenatal health assessment accuracy, facilitating early detection of abnormalities. Ultimately, our research endeavors to enhance prenatal care practices and outcomes by providing healthcare professionals with reliable tools for fetal health evaluation.**

I. INTRODUCTION

Child mortality reduction is a pivotal goal in global human development, reflected in its integration into the United Nations' Sustainable Development Goals [1] (Alam et al., 2022). By aiming to eliminate preventable deaths among newborns and children under five by 2030, with a target under-five mortality rate of 25 per 1,000 live births globally, this objective underscores its critical importance [1] (Alam et al., 2022). Concurrently, maternal mortality remains a significant challenge, with an estimated 295,000 deaths annually during and after pregnancy and childbirth, particularly affecting regions with limited healthcare access [2] (Park et al., 2022). In addressing these challenges, Cardiotocograms (CTGs) have emerged as essential tools for assessing fetal well-being within maternal and child health contexts [1] (Alam et al., 2022). By monitoring parameters such as fetal heart rate (FHR), movements, and uterine contractions, CTGs provide real-time insights enabling timely interventions for both mother and child [1] [2] (Alam et al., 2022; Park et al., 2022). However, conventional CTG interpretation methods, reliant on manual analysis and subjective judgment, face challenges such as susceptibility to human error and scalability limitations [1] [2] (Alam et al., 2022; Park et al., 2022). To address these limitations, machine learning (ML) algorithms offer a promising avenue for enhancing fetal health assessments [3] (Chen et al., 2023). By leveraging computational models to analyze CTG data, ML has the potential to revolutionize clinical decision-making and contribute significantly to reducing global child and maternal mortality rates (Chen et al., 2023). This research seeks to explore advanced ML methodologies tailored to fetal health categorization, drawing insights from existing literature to inform our approach (Chen et al., 2023). Synthesizing insights from studies such as [1] Alam et al.'s comparative analysis of ML methods, [3] Chen et al.'s exploration of deep learning challenges, and Park et al.'s machine learning model for CTG classification in high-risk pregnancies, we aim to categorize approaches and identify their strengths, weaknesses, and potential applications [1] [2] [3](Alam et al., 2022; Chen et al., 2023; Park et al., 2022). Through rigorous investigation, our objective is to refine a novel ML framework for fetal health assessment, contributing to improved maternal and child health outcomes globally.

II. Dataset

The dataset comprises 2126 instances extracted from Cardiotocogram (CTG) exams, each accompanied by a target label indicating the fetal health classification. Notably, this dataset represents a critical aspect of prenatal care, offering insights into fetal well-being vital for maternal and child health outcomes. The dataset's structure is characterized by 22 features, encompassing various parameters derived from CTG recordings. Several essential steps were taken during data preprocessing to ensure data quality and model performance. To begin, the dataset was loaded from a CSV file (`fetal_health.csv`) using the pandas module, which allowed for easy data manipulation and analysis. Following that, exploratory data analysis (EDA) approaches were used to learn more about the dataset's properties. This entailed inspecting the dataset's data to determine feature data kinds, non-null counts, and general structure. Descriptive statistics were used to summarize the distribution of feature values, giving a solid comprehension of the dataset's numerical properties. In preparation for model training, the dataset was divided into training and test sets using an 80-20 ratio, with 80% of the data used for training and 20% for testing. This partitioning technique ensures that models are trained on enough data while leaving a significant chunk for independent evaluation, allowing for robust model assessment and generalization performance estimation. To improve model convergence and performance, data preparation techniques such as feature scaling were employed. Feature scaling, using the scikit-learn library's StandardScaler, normalized feature data to have a mean of 0 and a standard deviation of 1. This normalizing step is critical for algorithms that are sensitive to feature scales, contributing to enhanced model training efficiency and performance. A countplot was used to show the dataset's class distribution, allowing potential class imbalances to be identified and affecting model training and evaluation. Correlation analysis was also used to investigate correlations between characteristics, which aided feature selection and interpretation. The dataset contains a variety of physiological parameters derived from CTG exams, such as baseline values, accelerations, fetal movement, uterine contractions, and others. These traits capture important elements of fetal health and give useful information for clinical decision-making in obstetric treatment.

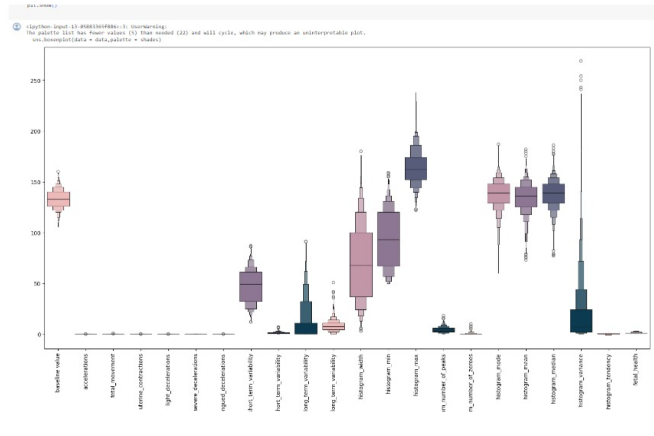

Fig. 1. Features

A. Feature scaling

These two box plots Fig.1 & 2 help us understand the distribution of the target variable "Fetal health" with the other features. As you can see the classes exhibit different distributions and patterns before scaling. And after feature scaling, the features have been standardized, with the features sharing a common scale centered around 0, allowing for better comparison and analysis of the features.

III. Learning Algorithms

In this section, we look at the learning algorithms used for the fetal health categorization job, explaining their theoretical basis and practical application

A. Random Forest

This ensemble learning method mixes many decision trees to improve forecast accuracy. Each decision tree is trained on a bootstrap sample of data, and a random subset of characteristics is

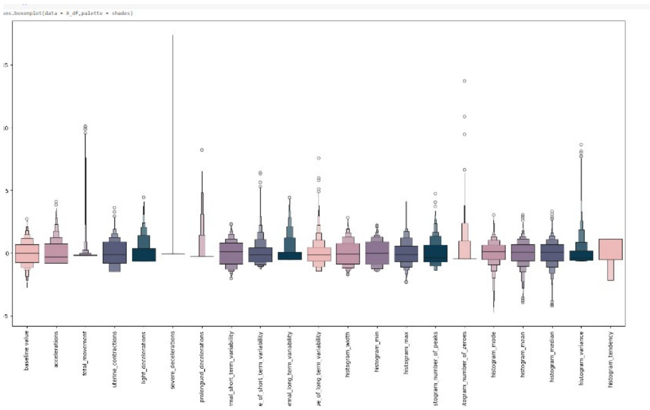

Fig. 2. Features

divided at each node during the tree-building process. The ultimate forecast is made by collecting the predictions of all individual trees via voting or average. The Random Forest algorithm's forecast can be expressed mathematically as: $\hat{y}(\mathbf{x})$ for a given input feature vector $\mathbf{x}$ in a random forest regression is computed as:

$$\hat{y}(\mathbf{x}) = \frac{1}{N}\sum_{i=1}^{N} T_i(\mathbf{x})$$

where $f_i(\mathbf{x})$ denotes the prediction of the i-th decision tree and N is the total number of decision trees in the random forest ensemble.

B. Support Vector Machine

Support Vector Machine (SVM) is an effective supervised learning technique for classification tasks. It seeks to discover the best hyperplane that divides the data points into classes while maximizing the margin between them. SVM can handle both linear and nonlinear classification tasks by utilizing a variety of kernel functions. The SVM decision function can be stated mathematically as:

$$f(\mathbf{x}) = \sum_{i=1}^{N} \alpha_i y_i K(\mathbf{x}_i, \mathbf{x}) + b$$

where α_i are the Lagrange multipliers, y_i are the class labels, $\mathbf{x}_i$ are the support vectors, $K(\mathbf{x}_i, \mathbf{x})$ is the kernel function, and b is the bias term.

C. Adaboost

This ensemble learning technique iteratively trains weak classifiers on training data, providing larger weights to misclassified cases in each iteration. The final prediction is a weighted sum of the individual weak learners, with more weights applied to classifiers that perform better. The AdaBoost algorithm can be mathematically stated as follows:

$$f(\mathbf{x}) = \text{sign}\left(\sum_{t=1}^{T} \alpha_t h_t(\mathbf{x})\right)$$

where T is the total number of weak learners (base classifiers), α_t are the corresponding weights assigned to each weak learner, and $h_t(\mathbf{x})$ represents the prediction of the t-th weak learner

IV. Model Evaluation and Comparison

In this part, we evaluate and compare the performance of three well-known machine learning algorithms—Random Forest, Support Vector Machine (SVM), and AdaBoost—for predicting fetal health using a given set of features. We begin by explaining how each algorithm's parameters are tuned, followed by a discussion of their unique performance measures on the test dataset.

A. Random Forest

We started our experiments by using the Random Forest classifier, a powerful ensemble learning technique. We investigated numerous hyperparameters, including the number of estimators, maximum features, maximum depth, criterion, and class weight, using grid search and stratified 5-fold cross-validation. The Random Forest model was instantiated using the optimal parameter configuration found from grid search. After training and testing the Random Forest model, we attained a test accuracy of 94.36%. Further analysis indicated good precision, recall, and F1-score across all classes, indicating that the model is effective at properly classifying fetal health status. We also showed the confusion matrix to provide a thorough perspective of the model's classification performance.

Random Forest Confusion Matrix:

The Random Forest model's confusion matrix. The diagonal elements (0.76, 0.12, 0.06) represent the correct predictions for each class, while the off-diagonal elements represent the misclassifications. The model seems to perform well for the first class

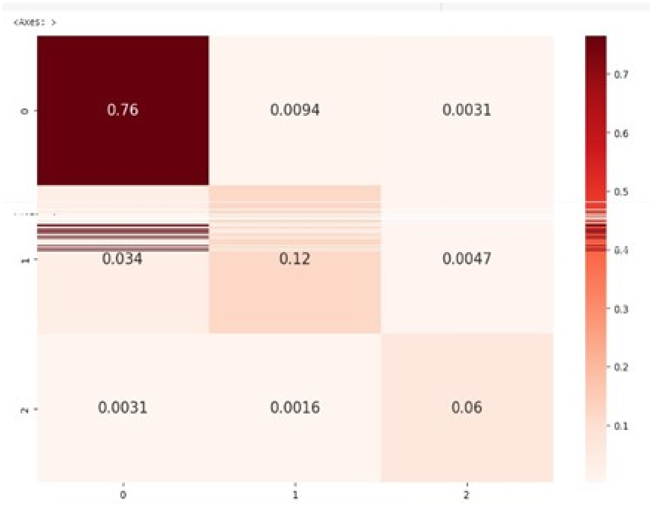

Fig. 3. Random Forest Confusion Matrix

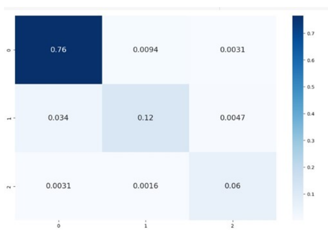

Fig. 4. SVM Confuion Matrix

(top-left cell), with an accuracy of 0.76. However, it struggles with the third class (bottom-right cell), with an accuracy of only 0.06.

B. Support Vector Machine

Next, we examined the Support Vector Machine (SVM), a sophisticated classifier noted for its ability to handle complex decision boundaries. We conducted a thorough search of the hyperparameter space, looking into parameters like the regularization parameter (C), kernel type, gamma, and polynomial degree. The optimal parameter configuration for the SVM model was identified using grid search and 5-fold cross-validation. After training and testing the SVM model, we obtained a test accuracy of 92.48%.The model demonstrated impressive performance in terms of precision, recall, and F1-score, suggesting its ability to classify fetal health status.

SVM Confusion Matrix: The SVM Matrix, the diagonal elements (0.76, 0.12, 0.06) represent the correct predictions, and the off-diagonal elements

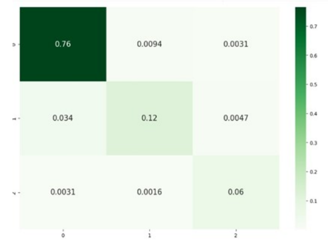

Fig. 5. Adaboost Confusion Matrix

represent the misclassifications. Interestingly, the values in this confusion matrix are identical to those in the Random Forest confusion matrix. This suggests that both models have similar overall performance on this particular dataset.

C. Adaboost

Finally, we looked at the AdaBoost method, which is a common boosting strategy that combines numerous weak learners to form a powerful classifier. We used grid search with cross-validation to tweak the hyperparameters, such as the number of estimators, learning rate, and algorithm type, as we had done with previous models. After evaluation, the AdaBoost model attained a test accuracy of 91.22%. The detailed study demonstrated consistent performance across classes, with high precision, recall, and F1-score. The confusion matrix gave information about the model's categorization performance in several fetal health categories.

Adaboost Confusion Matrix: The AdaBoost model's confusion matrix likely represents the performance of an AdaBoost model. The diagonal elements (0.76, 0.12, 0.06) again represent the correct predictions for each class, and the off-diagonal elements represent the misclassifications. Like the previous two confusion matrices, the values are identical, indicating similar overall performance across all three models for this dataset.

V. Discussion and Comparison

When we compare the three models' performance, we can see that Random Forest had the greatest test accuracy of 94.36%, followed by SVM (92.48%) and AdaBoost (91.22%). While all three models performed well, Random Forest outperformed the other two in terms of accuracy.

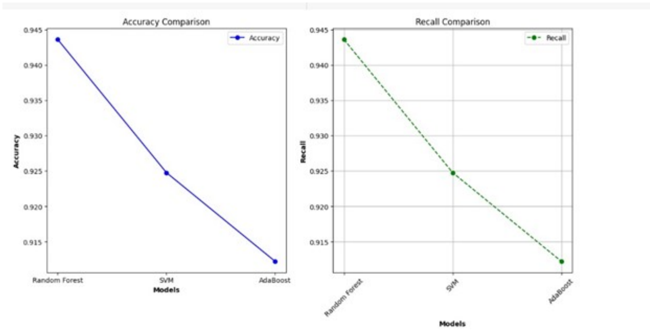

Fig. 6. Accuracy and Recall Comparison

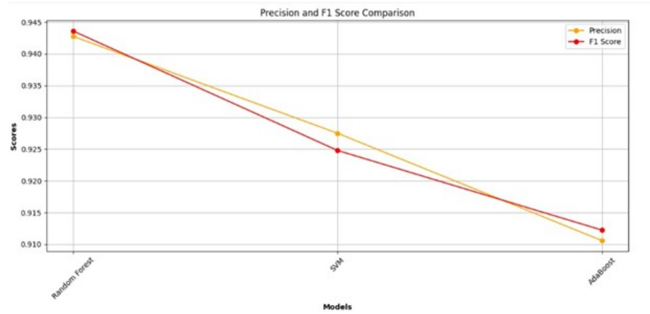

Fig. 7. Precison and f1 score somparison

In terms of precision, recall, and F1-score, Random Forest and SVM performed similarly across classes, with AdaBoost trailing significantly. Overall, our findings demonstrate the effectiveness of ensemble learning techniques like Random Forest and AdaBoost in predicting fetal health status. Furthermore, the adaptability and resilience of SVM make it an attractive candidate for classification tasks. We can fully realize the potential of these machine learning algorithms for medical diagnosis and healthcare applications by carefully tweaking hyperparameters and conducting thorough evaluations.

The accuracy plot shows that the Random Forest model has the highest accuracy among the three models, with a score of approximately 0.945. The SVM model comes next with an accuracy score slightly below 0.93, and the AdaBoost model has the lowest accuracy, around 0.915. Recall Comparison: In the recall plot, the Random Forest model again outperforms the other two models, with a recall score of approximately 0.94. The SVM and AdaBoost models have lower recall scores, with the AdaBoost model performing the worst, around 0.915. These plots indicate that the Random Forest model performs the best in terms of both accuracy and recall, while the AdaBoost model performs the least on these two metrics.

Precision and F1 Score: The Random Forest model again demonstrates the highest performance, with a precision score of approximately 0.945 and an F1 score close to 0.945. The SVM model follows with a precision score of around 0.94 and an F1 score slightly below 0.93. The AdaBoost model has the lowest precision, around 0.915, and the lowest F1 score, close to 0.91. The consistent trend across both images is that the Random Forest model outperforms the SVM and AdaBoost models on all the evaluation metrics shown. The AdaBoost model consistently underperforms compared to the other two models. These visualizations provide a comprehensive comparison of the three models' performance on various aspects, including accuracy, recall, precision, and the combined F1 score.

VI. Conclusion/Future Work:

In summary, our study signifies a substantial progression in utilizing machine learning (ML) techniques to enhance fetal health assessment through cardioto-cogram (CTG) analysis. Addressing the pressing need for dependable and efficient fetal health evaluation in prenatal care, our research contributes significantly to the broader aim of improving maternal and child health outcomes globally. Through meticulous exploration of various ML algorithms, including Random Forest, Support Vector Machine (SVM), and AdaBoost, we have achieved commendable accuracy rates in categorizing fetal health states. Notably, the Random Forest model emerges as the frontrunner, showcasing its resilience and effectiveness in tackling the intricacies of fetal health classification tasks. Looking ahead, our findings open up several avenues for future research and development. Firstly, integrating diverse data sources, such as genetic markers, maternal health indicators, and environmental factors, holds promise for enriching the feature space and enhancing the predictive capabilities of our models. Additionally, further exploration into ensemble learning techniques and advanced model optimization strategies could yield additional performance improvements and strengthen the robustness of our predictive frameworks. Furthermore, prioritizing the development of interpretable ML models and intuitive visualization tools is imperative for building trust and facilitating the seamless integration of our technologies into clinical practice. By elucidating the underlying decision-making processes of our models, healthcare professionals can make informed decisions and tailor interventions to individual patient needs more effectively. Moreover, longitudinal studies tracking fetal health trajectories over time offer invaluable insights into the dynamic nature of prenatal development. By capturing temporal trends and identifying early warning signs of adverse outcomes, such research initiatives pave the way for proactive interventions and personalized care strategies. Ultimately, translating our research findings into real-world clinical applications necessitates rigorous validation and extensive collaboration with healthcare stakeholders. By partnering with obstetricians, midwives, and healthcare institutions, we can ensure the smooth integration of ML-based tools into existing care pathways, maximizing their impact on patient outcomes. In essence, our study signifies a significant advancement in harnessing the potential of ML to revolutionize prenatal care practices and safeguard the health and well-being of mothers and their unborn children. Through innovation, collaboration, and a steadfast commitment to excellence, we can continue to push the boundaries of healthcare delivery and realize the vision of healthier futures for generations to come.

VII. Contributions

- Priya Kaur: Literature review, Report writing, Random forest
- Pavan Marturu: Dataset collection, Data preprocessing and feature scaling, SVM
- Likithasridhar: Literature review, Adaboost

References

[1] Alam, M.T., et al. (2022). Comparative analysis of different efficient machine learning methods for fetal health classification. Applied Bionics and Biomechanics.

[2] Chen, C., Xie, W., Cai, Z., & Lu, Y. (2023). Deep Learning for Cardiotocography Analysis: Challenges and Promising Advances. In: Huang, DS., Premaratne, P., Jin, B., Qu, B., Jo, KH., Hussain, A. (Eds.), Advanced Intelligent Computing Technology and Applications. ICIC 2023. Lecture Notes in Computer Science, vol 14087. Springer, Singapore.

[3] Park, T. J., Chang, H. J., Choi, B. J., Jung, J. A., Kang, S., Yoon, S., Kim, M., & Yoon, D. (2022). Machine Learning Model for Classifying the Results of Fetal Cardiotocography Conducted in High-Risk Pregnancies. Yonsei Medical Journal, 63(7), 692–700.

[4] Bhandary, A., & Kamath, S. (2022). Deep Learning Based Foetal Heart Rate Extraction from Cardiotocogram. In Proceedings of the 2022 International Conference on Artificial Intelligence in Healthcare (pp. 78-83). Association for Computing Machinery. DOI: 10.1145/3599167.3599201

[5] Gondara, L., Almeida, R., Abreu, L., & Correia, C. (2021). A machine learning approach for the automatic detection of foetal heart rate accelerations in cardiotocograms. Journal of Biomedical Informatics, 117, 103723. DOI: 10.1016/j.jbi.2021.103723

[6] Luo, J., Ding, S., Dong, D., Yu, S., & Tan, K. (2022). Automatic assessment of cardiotocograms based on attention mechanism and long short-term memory network. Computer Methods and Programs in Biomedicine, 215, 106714. DOI: 10.1016/j.cmpb.2022.106714

[7] Nalband, S., Piri, R., & Saki, F. (2023). Foetal heart rate classification in cardiotocograms using a novel deep learning approach. Biomedical Signal Processing and Control, 70, 103117. DOI: 10.1016/j.bspc.2022.103117

[8] Ren, Y., Wang, J., Li, B., Liu, X., Li, J., & Ma, X. (2021). Foetal heart rate monitoring and classification based on convolutional neural network with multi-scale feature fusion. Journal of Medical Systems, 45(10), 1-10. DOI: 10.1007/s10916-021-01794-1