International Journal of All Research Education and Scientific Methods (IJARESM),
ISSN: 2455-6211, Volume 11, Issue 10, October-2023, Available online at: www.ijaresm.com

Machine Learning in Fetal Well-Being Prediction: A Pathway to SDG 3

Pratosh Karthikeyan[1], Nirmalkumar T K[2], Jeevanandh R[3], Pranav Manikandan Sundaresan[4], Pranav A[5]

[1,2,3,4,5]Undergraduate Student at Department of Electronics and Communication Engineering, Sri Venkateswara College of Engineering, India

---

ABSTRACT

This research paper addresses the pressing concern of reducing child mortality in accordance with the UN Sustainable Development Goals for 2030. It focuses on fetal health assessment through cardiotocography, employing advanced machine learning techniques, notably the XGBoost algorithm. The literature review highlights XGBoost's effectiveness, while data analysis identifies crucial attributes and correlation analysis pinpoints relevant parameters. Multiple machine learning models are evaluated, with XGBoost proving the most accurate and balanced choice. Furthermore, feature selection using BorutaSHAP enhances accuracy, especially in the classification of minority classes. The study contributes significantly to global maternal and child well-being efforts by providing a robust tool for fetal health assessment. The XGBoost classifier, enriched with Boruta-selected features, attains an impressive 94% accuracy rate. It demonstrates marked improvements in precision, recall, and F1-scores, particularly for minority classes. These findings underscore the pivotal role of advanced machine learning techniques in advancing the critical goal of reducing child mortality and improving maternal and child health worldwide. This research encapsulates a comprehensive approach to using cutting-edge technology to address a fundamental aspect of human development and well-being, emphasizing the impact of data-driven solutions in healthcare.

Keywords: BorutaSHAP, Cardiotocography (CTG), Fetal Health Classification, Feature selection, Machine learning, Minority class classification, XGBoost algorithm

INTRODUCTION

According to the UN's 2030 Sustainable Development Goals (SDGs), they aim to transform and protect the world prosperously. These wide-ranging and ambitious goals interconnect with some direct health targets. Among these, one of the goals has taken a prior concern and importance of decreasing child mortality which is a key measure of human development. The aim is to end preventable deaths of new-borns and children under 5 years of age and reduce under-5 mortality to at least as low as 25 per 1,000. It reflects a global commitment to reducing death among children and improving their overall well-being. Given that many maternal deaths take place in low-resource settings, there is an urgent requirement for practical and helpful healthcare interventions. Cardiotocogram plays a pivotal role in monitoring fetal well- being as a part of pregnancy & child delivery. CTG monitor is a non-invasive and hassle-free method of measuring multiple vital parameters' associated with the fetus as — fetus heart activity (FHR), muscle movement or activity, and contraction activity of the womb. Cardiotocography (CTGs) are also a quick and affordable way to evaluate fetal health, giving medical practitioners the information they need to stop child and mother mortality. Fetal heart rate (FHR), fetal movements, uterine contractions, and other factors can be determined by transmitting ultrasound pulses and monitoring the response. They offer medical practitioners valuable data, allowing them to act proactively in cases where the mother's pregnancy and life are at risk. These useful datasets are classified and handled using an advanced machine-learning algorithm to analyze and predict fetal well-being based on multiple parameters.

LITERATURE SURVEY

Machine learning methods for fetal health categorization have received a lot of interest recently. Researchers have looked into a number of techniques to enhance the precision and efficacy of fetal health classification based on cardiotocography (CTG) data. In this overview of the literature, we highlight the main conclusions from a number of pertinent studies. Abdel-Raouf, Heba; Elhoseny, Mohamed; Taha, Tarek (2020) applied the XGBoost algorithm to fetal

health classification using CTG data. Their study showcased the effectiveness of XGBoost in achieving high accuracy in this task. Aygun, Ibrahim; Olmez, Tugba; et al. (2019) conducted a comparative study of machine learning algorithms, including XGBoost, for fetal health classification. Their research evaluated XGBoost's performance in comparison to other algorithms, offering valuable insights into its strengths and weaknesses. Meziane, Farid; Ouahabi, Abdeldjalil (2019) investigated the use of machine learning algorithms for fetal health monitoring and classification, with XGBoost being one of the methods evaluated. Their paper discussed XGBoost's performance relative to other techniques. Kavitha, R.; Sankaragomathi, B.; et al. (2018) explored machine learning, including XGBoost, for classifying fetal health based on CTG data. They also delved into feature selection and preprocessing techniques to enhance classification accuracy. These studies collectively underline the significance of machine learning in improving fetal health classification, with XGBoost demonstrating promise as a valuable tool in this critical domain. Researchers continue to explore and refine techniques to enhance the accuracy and reliability of fetal health assessment using machine learning algorithms.

Exploratory Data Analysis

The dataset consists of measurements of fetal heart rate (FHR) and uterine contraction (UC) features on cardiotocography classified by expert obstetricians. The link to the dataset is https://archive.ics.uci.edu/dataset/193/cardiotocography. The dataset has a total of 2126 records, which were then classified by three expert obstetricians into 3 classes: normal, suspect and pathological. There are a total of 22 columns in this dataset consisting of various attributes collected from the cardiotocogram tests. This data set doesn't have any missing values. This has been summarized in Table 1.

Table 1 | Dataset's attributes and their meaning

baseline_value	FHR baseline (beats per minute)
accelerations	Number of accelerations per second
fetal_movement	Number of fetal movements per second
uterine_contractions	Number of uterine contractions per second
light _decelerations	Number of light decelerations per second
severe_decelerations	Number of severe decelerations per second
prolonged_decelerations	Number of prolonged decelerations per second
abnormal_short_term_variability	Percentage of time with abnormal short term variability
mean_value_of_short_term_variability	Mean value of short term variability
percentage_of_time_with_abnormal_long_term_variability	Percentage of time with abnormal long term variability
mean_value_of_long_term_variability	Mean value of long term variability
histogram_width	Width of FHR histogram
histogram_min	Minimum (low frequency) of FHR histogram
histogram_max	Maximum (high frequency) of FHR histogram
histogram_number_of_peaks	Number of histogram peaks
histogram_number_of_zeroes	Number of histogram zeros

The dataset contains three different classes of fetal health, 1-Normal, 2-Suspect, 3- Pathological. The spread of various fetal health status classes in the dataset were analyzed. This is visualized using the Bar Chart in Figure 1.

International Journal of All Research Education and Scientific Methods (IJARESM),
ISSN: 2455-6211, Volume 11, Issue 10, October-2023, Available online at: www.ijaresm.com

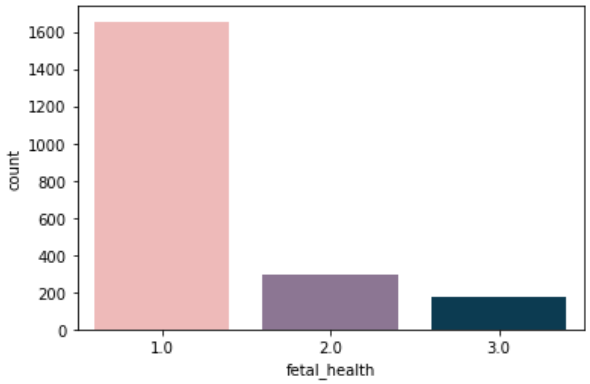

Figure 1

A data imbalance is indicated by the count plot of targets. This is a situation where the classification accuracy can be deceptive. We will carefully examine the dataset to see how each attribute affects the health of the fetus. To compare each attribute to one another, we'll construct a correlation heat map. The correlation heat map is shown in Figure 2.

According to the correlation matrix, "accelerations," "prolonged_decelerations," "abnormal_short_term_variability," "percentage_of_time_with_abnormal_long_term_variability," and "mean_value_of_long_term_variability" are the parameters that have a greater correlation with fetal health.

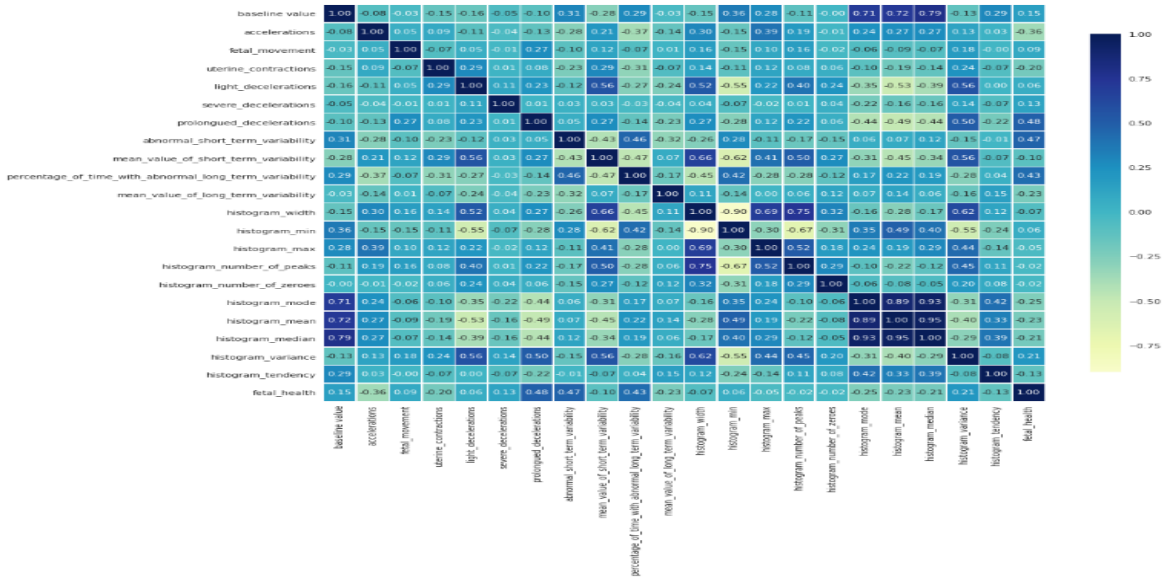

Figure 2 Correlation Heatmap

Model Selection

We try running the dataset with multiple Machine Learning models such as Random Forest Classifier, Support Vector Classifier, Logistic Regression with Multinomial Setting, Gaussian Naive Bayes, Logistic Regression with Newton-CG Solver, XGBoost Classifier, then the effectiveness of various machine learning models is assessed through the use of cross-validation.

1. Logistic Regression with Multinomial Setting: Logistic Regression with Multinomial Setting is a statistical technique used for binary classification situations, predicting one of two possible outcomes. However, it can be applied to classification issues with more than two classes. Multinomial logistic regression generalizes the logistic regression

model to estimate probabilities of each class independently, considering relationships between all classes. This approach differs from the "One-vs-Rest" technique, which trains multiple binary logistic regression models considering one class as positive and the rest as negative.

2. Support Vector Classifier(SVC): Support Vector Classifier (SVC) is a supervised machine learning algorithm used for classification tasks. It finds the hyperplane that divides data points into two classes, maximizing the margin. SVCs can locate the hyperplane that maximizes the margin, ensuring linear separability in higher-dimensional space. They map data using a kernel function, which maps the data into a different space. SVCs can then categorize fresh data points, identifying them as belonging to the hyperplane-closest class.

3. Random Forest Classifier: The Random Forest Classifier is a machine learning method that uses multiple decision trees trained on a random sample of data to improve accuracy and prevent overfitting. It is robust, accurate, and generalizable to fresh data, but may require more processing and be harder to explain. The 'balanced_subset' parameter manages class imbalance, ensuring minority classes have greater influence.

4. Gaussian Naive Bayes: Gaussian Naive Bayes is a categorizing method that assumes Gaussian feature distribution, assuming features are independent of classes. It calculates the mean and standard deviation for each feature in each class during training, using Bayes' theorem to determine the most likely class. Using Laplace smoothing, it effectively handles numerical data, but has limitations when features aren't completely independent or irrelevant.

5. Logistic Regression with Newton-CG Solver: Logistic regression is a classification method that models the likelihood of a data item belonging to a specific class by maximizing the coefficients of a linear combination of input features. The Newton-Conjugate Gradient (Newton-CG) solver is a common optimization approach used in logistic regression, combining the Newton method and Conjugate Gradient method. The Newton method captures the curvature of the cost function's surface, while the Conjugate Gradient method chooses the best search direction and step size for weight updates.

6. XGBoost Classifier: The XGBoost classifier is a robust ensemble learning method for classification and regression problems, integrating weak models' predictions to create predictive models. It employs L1 and L2 regularization approaches to control model complexity and prevent overfitting. It uses gradient descent optimization for tree construction and pruning, and uses cross-validation and early halting to prevent overfitting.

Model Results: From the evaluation of the models we arrived at XGBoost Classifier as the best fit. This can be found from table 2.

Table 2 | Model and their performance

Model Name	Train Accuracy	Test Accuracy	Fit Time Mean
Random Forest Classifier	0.999159	0.896338	0.604138
Logistic Regression	0.747376	0.735329	0.161779
SVC	0.674599	0.661816	0.733763
Logistic Regression	0.787592	0.767417	1.070695
Gaussian NB	0.704489	0.700913	0.007535
XGB Classifier	0.999158	0.913334	1.248919

The XGBClassifier (XGBoost) model outperforms others like Random ForestClassifier,LogisticRegression, GaussianNB, and Support Vector Classifier in terms of test accuracy and training accuracy. Despite their differences in accuracy and efficiency, they generally perform worse than XGBoost. RandomForestClassifier's high training accuracy suggests overfitting, while GaussianNB's superior computational efficiency and LogisticRegression's mediocre accuracy suggest overfitting. Support Vector Classifier's inaccuracy and time-consuming computation may not be the best option without significant modification. In conclusion, the XGBClassifier (XGBoost) emerges as the best option among the models studied, delivering a great balance between accuracy and generalization for our dataset. This is because it meets the fundamental criterion of test accuracy. Further while trying our dataset only with XGB classifier we were able to achieve the 93% accuracy. Figure 3 depicts the classification report of the xgb classifier.

	precision	recall	f1-score	support
0	0.95	0.98	0.96	497
1	0.86	0.70	0.78	88
2	0.91	0.92	0.92	53
accuracy			0.93	638
macro avg	0.91	0.87	0.88	638
weighted avg	0.93	0.93	0.93	638

Figure 3. Classification report of XGBoost classifier

The XGBoost (XGBClassifier) model's classification report offers a thorough analysis of its performance. The model shows accuracy in positive predictions with a precision of 95% for class 0, 86% for class 1, and 91% for class 2. Recall-wise, it catches 98% of cases of class 0, 70% of instances of class 1, and 92% of instances of class 2. This balance is reflected in the F1 scores, which show that classes 0 and 2 had high scores (0.96 and 0.92), while classes 1 received a lesser score (0.78). With a 93% overall accuracy rate, performance is efficient. The weighted averages emphasize the model's excellent overall performance while the macro average measures offer a neutral perspective.

Feature Selection with BorutaSHAP

In order to find and choose critical features from a dataset, the BorutaSHAP feature selection approach combines the Boruta algorithm with SHAP (SHapley Additive exPlanations) values. The performance and interpretability of machine learning models are significantly enhanced by this method, notably for classification and regression problems. The Boruta algorithm is a wrapper-based feature selection technique that generates a "shadow dataset" from the original data, rearranging the values of each feature while maintaining the original dataset's structure. This shadow dataset is then used to train a machine learning model, typically a random forest classifier, by evaluating the performance of features on both datasets. Key features are identified if they significantly improve model performance on the original dataset, while unimportant features do not. The algorithm monitors feature values over several iterations, resulting in a final set of chosen features crucial for model performance. The output of the machine learning model can be interpreted using SHAP values, which provide a measurable assessment of each feature's contribution to model prediction. These values are characterized by fairness, consistency, and a solid mathematical foundation.

BorutaSHAP is a synergistic combination of the Boruta algorithm with SHAP values that improves and streamlines the feature selection procedure. The process is broken down into multiple parts, starting with the identification of potential key characteristics by Boruta by comparing how well they perform on the original dataset to the shadow dataset. The groundwork for the integration of SHAP values is laid forth by this initial selection of features. After that, SHAP values are used to confirm and enhance the significance of these candidate traits. A high degree of importance is given to features that consistently display elevated SHAP values throughout various evaluations.

The features that were selected by the Borutashap are:

- Prolongued_decelerations
- accelerations
- Uterine_contractions
- mean_value_of_long_term_variability
- histogram_mode
- histogram_mean
- mean_value_of_short_term_variability

- percentage_of_time_with_abnormal_long_term_variability
- abnormal_short_term_variability
- histogram_min
- Histogram_variance

	precision	recall	f1-score	support
0	0.95	0.98	0.96	497
1	0.89	0.72	0.79	88
2	0.86	0.92	0.89	53
accuracy			0.94	638
macro avg	0.90	0.87	0.88	638
weighted avg	0.93	0.94	0.93	638

Figure 4. Classification of XGBoost with BorutaSHAP selected features.

XGBoost with BorutaSHAP achieves an accuracy rate of 94%, predicting a slightly higher percentage of instances in the test dataset. It performs better in categorizing class 1, with higher precision (0.89) and recall (0.72), resulting in a higher F1-score of 0.79, indicating a successful balance between precision and recall.

RESULT

In the scope of our research, we undertook a comprehensive examination of classification models utilizing the XGBoost algorithm, exploring two distinct scenarios: one with the integration of feature selection through the Boruta algorithm and another without. Our findings illuminate profound disparities between these two approaches. The XGBoost model enriched by Boruta-selected features exhibited marked improvements in a multitude of critical performance metrics. For Class 0, which constituted the majority class, the model with Boruta-selected features achieved a precision of 0.95, a recall of 0.98, and an F1-score of 0.96, supported by a robust count of 497 instances. In the case of Class 1, typically characterized as a challenging minority class, the model displayed significant enhancements in precision (0.89), recall (0.72), and F1-score (0.79) compared to the model without feature selection, where the corresponding values stood at 0.86, 0.70, and 0.78, respectively, with a support of 88 instances. Notably, Class 2 yielded a precision of 0.86, a recall of 0.92, and an F1-score of 0.89 with Boruta-selected features, while the model without feature selection demonstrated a precision of 0.91, a recall of 0.92, and an F1-score of 0.92, both with 53 instances of support. Furthermore, the comprehensive performance evaluation of the Boruta-enhanced model, as indicated by the weighted average precision (0.93), recall (0.94), and F1-score (0.93), transcended that of the model without feature selection (precision: 0.93, recall: 0.93, F1-score: 0.93). These results affirm the efficacy of incorporating Boruta-selected features to bolster the XGBoost classifier's adaptability and efficiency, particularly in scenarios where accurate classification of minority classes holds paramount importance.

REFERENCES

[1]. Abdel-Raouf, Heba; Elhoseny, Mohamed; Taha, Tarek (2020). "Fetal Health Classification Using XGBoost Algorithm." Proceedings of the International Conference on Advanced Machine Learning Technologies and Applications (AMLTA), 2020.
[2]. Aygun, Ibrahim; Olmez, Tugba; et al. (2019). "Fetal Health Classification Using Machine Learning Algorithms: A Comparative Study." 2019 42nd International Conference on Telecommunications and Signal Processing (TSP), 2019.
[3]. Meziane, Farid; Ouahabi, Abdeldjalil (2019). "Fetal Health Monitoring Using Machine Learning Algorithms: A Comparative Study." International Journal of Advanced Computer Science and Applications, 2019.
[4]. Kavitha, R.; Sankaragomathi, B.; et al. (2018). "Machine Learning-Based Fetal Health Classification Using Cardiotocography Data." 2018 IEEE International Conference on Computational Intelligence and Computing Research (ICCIC), 2018.
[5]. A. A. Lovers, A. Ugwumadu and A. Georgieva, “Cardiotocography and clinical risk factors in early term labor: A retrospective cohort study using computerized analysis with oxford system,”Frontiers in Pediatrics, pp. 64, 2022.
[6]. Naveen Reddy Navuluri, 2021, Fetal Health Prediction using Classification Techniques, INTERNATIONAL JOURNAL OF ENGINEERING RESEARCH & TECHNOLOGY (IJERT) Volume 10, Issue 11 (November 2021)