Evaluating the Effectiveness of LightGBM for Fetal Health Prediction

Claudio Januar Aristan
Computer Science Department
School of Computer Science
Bina Nusantara University
Jakarta, Indonesia,11530
claudio.aristan@binus.ac.id

Nicolas Yang
Computer Science Department
School of Computer Science
Bina Nusantara University
Jakarta, Indonesia,11530
nicolas.yang@binus.ac.id

Viery Octavianus Biliam
Computer Science Department
School of Computer Science
Bina Nusantara University
Jakarta, Indonesia,11530
viery.biliam@binus.ac.id

Felix Indra Kurniadi
Computer Science Department
School of Computer Science
Bina Nusantara University
Jakarta, Indonesia,11530
felix.indra@binus.ac.id

Riccosan
Computer Science Department
School of Computer Science
Bina Nusantara University
Jakarta, Indonesia,11530
riccosan@binus.ac.id

***Abstract*—The health of the fetus is essential for assuring that babies are born healthy. A mother's fetus must be in excellent health. The categorization of fetal health is essential for assessing the well-being and development of unborn infants. Despite this, numerous findings continue to indicate a high neonatal mortality rate. During pregnancy and labor, the Cardiotocography (CTG) is a common method for monitoring fetal health. It measures fetal heart rate and uterine contractions, allowing healthcare professionals to evaluate the baby's health and identify potential problems early on. However, interpretation of CTG data remains subjective, resulting in potential misdiagnoses of fetal health. This study seeks to reduce infant mortality rates caused by delivery complications by employing machine learning technology to assess a baby's health based on Cardiotocography test results. In this study, we evaluated the performance of three machine learning algorithms - XGBoost, Random Forest, and LightGBM - in predicting fetal health based on Cardiotocography data. We implemented various scenarios involving preprocessing techniques, feature selection methods, and hyperparameter tuning to enhance the model's prediction accuracy. The scenarios are without preprocessing, feature selection using Information Gain, feature selection using K-Best, Feature selection using both feature selection algorithm, without hyperparamter tuning and hyperparamter tuning with Grid Search. The result showed the use of all techniques combined proved beneficial, resulting in accuracies of 96% for XGBoost, 95% for Random Forest, and 97% for LightGBM.**

***Keywords*—cardiotocography, fetal health, machine learning, random forest, extreme gradient boosting, light gradient boosting machine**

I. INTRODUCTION

There are several factors that can have an impact on fetal health, including adequate nutrition, avoidance of alcohol and tobacco, management of illnesses, appropriate prenatal care, and a maternal lifestyle that is supportive of fetal growth and development. According to global statistics from 2015, the incidence of stillbirths was 18.4 per 1000 total births. Despite a downward trend in these figures, it is estimated that, based on the present pace of advancement, it will require over 160 years for a pregnant woman residing in Africa to attain a comparable probability of her newborn surviving childbirth as that of a woman in a high-income nation at present [1]. The majority of the 2.6 million stillbirths that occur annually are concentrated in low- and middle-income countries [2]. Approximately 50% of these fatalities, which equates to roughly 1.3 million deaths per year, transpire during the labor process. These fatalities generally pertain to full-term neonates who would have been anticipated to survive under normal circumstances.

Healthcare professionals utilize diverse diagnostic techniques to evaluate fetal health status in expectant mothers. Some of the diagnostic procedures utilized in obstetrics and gynecology are Ultrasound, Electronic Fetal Monitoring (EFM), Amniocentesis, Chorionic Villus Sampling (CVS), and Blood Tests, among other methods. Nonetheless, these processes typically require a considerable duration, varying from a few hours to several weeks. The objective of this investigation is to utilize Cardiotocography (CTG) technology for the evaluation of fetal well-being. The CTG test is a commonly used diagnostic tool for monitoring fetal conditions. It has a relatively short turnaround time of less than an hour and is often performed on a regular basis. Additionally, it is a cost-effective option, with comparable or even lower pricing than other tests available.

Improta et al. [3] emphasized that the interpretation of CTG signals is still subjective and error-prone. As a result, computerized systems for automated analysis have been developed to aid physicians in diagnosing fetal health or distress. Ramla et al. [4] utilized the Classification And Regression Tree (CART) Algorithm, which achieved an accuracy of

90.12%. On the other hand, Piri and Mohapatra [5] employed a Self-Organizing Neural Network (SOM) and Random Forest, with the latter yielding superior results. Peterek et al. [6] implemented eight distinct algorithms, including XGBoost and Random Forest, and attained a 94% accuracy rate. Sahin and Subasi [7] also employed eight distinct types of algorithms, with the Random Forest algorithm yielding the highest accuracy. Similarly, Subasi et al. [8] employed eight distinct varieties of algorithms, with the addition of a Bagging Ensemble Classifier. As with the previous two experiments, Random Forest produced the highest accuracy.

In this paper, some of the previously mentioned algorithms will be compared to other popular machine learning to classify the CTG data, paired with some other pre-processing methods in order to diagnose fetal health. This paper will use some of the new popular machine learning that, to the best knowledge of the authors, hasn't been used to classify CTG data. We specifically chose to utilize the CTG dataset due to CTG dataset is widely recognized and utilized in the field of obstetrics and gynecology for monitoring fetal well-being during pregnancy and labor. By using CTG dataset, we could benchmark our proposed method against existing methods, and potentially contribute to the advancement of clinical practices in obstetrics and gynecology.

II. Literature Review

The utilization of CTG as a diagnostic modality for detecting fetal distress during antepartum and intrapartum phases is extensively discussed by Ramla et al. The CTG data comprises four crucial elements, namely the baseline (BL) of the fetal heart rate, accelerations (ACC), decelerations (DCL), and variability. Medical professionals utilize these parameters to categorize the fetal state as either typical, questionable, or pathological. The researchers proposed employing the CART algorithm to forecast fetal health status based on Cardiotocography data, encompassing the Fetal Heart Rate (FHR) and Uterine Contraction (UC), as evaluated by proficient medical practitioners. The dataset comprised of 1655 instances of normal cases, 295 instances of suspicious cases, and 176 instances of pathologic cases, which were categorized based on their morphologic pattern and fetal state (Normal, Suspicious, Pathological). Furthermore, the scholars employed a 5-fold cross-validation technique and assessed the performance of the CART algorithm by utilizing precision, recall, and F-score metrics. The results indicate that the utilization of both the gini index and entropy computations resulted in an increased precision of the algorithm. Specifically, the gini index exhibited a more substantial enhancement in accuracy compared to entropy. This implies that the incorporation of the gini index has the potential to improve the precision of decision tree algorithms [4].

The intricacies of prenatal care are examined by Peterek et al. in light of the fact that a fetus lacks the ability to articulate or convey its own state, necessitating medical practitioners to depend on techniques such as ultrasonic, electrocardiography, or echocardiography for monitoring purposes. Since its inception, CTG has been recognized as a critical tool in prenatal care. The objective of the study conducted by the researchers was to evaluate and compare the efficacy of different algorithms, which have been previously proven successful in a range of classification tasks, in the classification of CTG data for the purpose of fetal health prediction. The corpus, which consists of 2126 entries, was generated at the Faculty of Medicine located at the University of Porto. The records were subject to manual evaluation by a panel of three obstetricians, who subsequently assigned them to one of three distinct categories: Normal, Suspicious, and Pathological. Two distinct methodologies were proposed for forecasting fetal well-being utilizing CTG data, utilizing a 10-fold cross-validation technique. The investigators employed the Kohonen Self-Organizing Neural Network (SOM) and Random Tree methodologies, and evaluated their outcomes by means of a Confusion Matrix. According to the results, it was observed that the Random Tree algorithm exhibited superior performance compared to the SOM algorithm, attaining a precision rate of 94.69%. The scholars proposed that enhancing the precision could be achieved by adapting the Random Forest algorithm from a broad classifier to a more specialized one that is specific to CTG [6].

The study conducted by Piri et al. aimed to investigate diverse techniques for forecasting fetal health status through the utilization of an Association Based Classification Approach employing multiple algorithms. The researchers opted to utilize CTG data that was publicly accessible in the UCI Machine Learning Repository to conduct an analysis of CTG recordings and identify fundamental characteristics. The dataset comprises of 21 distinct attributes derived from CTG signals, encompassing both fetal heart rate (FHR) and uterine contractions (UC). Additionally, three expert obstetricians were involved in the classification of the data into two categories: the class pattern (1-10) and the fetal state class (Normal = N, Suspect = S, Pathologic = P). This fetal state class served as the target feature for the study. The study employed a set of eight algorithms that were coupled with pruning and feature selection techniques to minimize complexity and enhance precision. The findings of the study revealed that XGBoost and Random Forest algorithms exhibited superior accuracy in comparison to the remaining six algorithms, irrespective of the presence or absence of feature selection. The results indicate that a 93% accuracy rate was attained in the absence of Feature Selection, whereas a 94% accuracy rate was achieved with Feature Selection. Although the improvement is marginal, the possibility of utilizing Feature Selection in our experiment cannot be dismissed [5].

According to Sahin and colleagues, the process of acquiring precise information about the fetus during pregnancy can be arduous, necessitating obstetricians to depend on indirect markers of the fetal condition. CTG has been found to be a valuable tool in clinical practice. However, it has been observed that CTG may generate inconsistencies that could result in unwarranted interventions. Therefore, the scholars suggested the utilization of ROC and F1-measure.The researchers conducted an experiment with the objective of evaluating the efficacy of eight distinct machine learning algorithms, namely ANN, SVM, k-NN, RF, CART, Logistic Regression, C4.5, and RBFN, in relation to the UCI CTG dataset. The present dataset, procured from the University of California, Irvine, encompasses CTG data that comprises discernible characteristics. The CTG data was evaluated by a group of three proficient obstetricians who classified it into two categories: normal or pathological, based on the condition of the fetus. The data pertaining to UCI cardiotocography was obtained through employment of the automated SISPORTO 2.0 software. The data was scrutinized to identify any dubious entries, and subsequently, the normal and pathologic classes were incorporated into the NP feature. The CTG dataset comprises 21 distinct features, of which 8 are continuous and 13 are discrete. The data was categorized according to the fetal state, specifically as either normal or pathological. The evaluation of the outcomes was conducted based on several metrics, including accuracy, specificity, sensitivity, F-measure, and ROC curve. According to the results, the Random Forest algorithm exhibited superior performance compared to the remaining seven algorithms, attaining a precision rate of 99.18%. The findings of this study suggest that Random Forest is a proficient classifier in discerning between normal and pathological categories of CTG data [7].

The objective of Subasi et al. was to utilize Bagging Ensemble Classifiers on the algorithms employed in their prior publication [7] with the intention of enhancing the precision of forecasting fetal health status. The CTG datasets sourced from UCI were subjected to evaluation by three obstetricians with expertise in the field. The purpose of the evaluation was to ascertain whether the datasets were indicative of normal or pathological conditions, with the embryo's status serving as the basis for this determination. The CTG dataset comprises a total of 21 features, out of which 13 are discrete and 8 are continuous. The assessment effectively evaluated approximately 1831 CTG data entries. The present study elucidates the utilization of Bagging Ensemble Classifiers in tandem with eight distinct algorithmic models, namely Support Vector Machine (SVM), K-Nearest Neighbors (KNN), Artificial Neural Network (ANN), Random Forest, Classification and Regression Tree (CART), C4.5, Reduced Error Pruning Tree (REP Tree), and Random Tree. The evaluation of the outcomes was conducted based on the metrics of accuracy, F-measure, and ROC Area.The results indicate that the implementation of the Bagging technique led to a slight enhancement in accuracy, albeit marginal, with a mere 0.6% increase observed. The implementation of the proposed modification has the potential to result in a reduction of accuracy within the REP Tree from 98.2% to 98.09%. Hence, although the Bagging Ensemble Classifier exhibits potential in the classification of normal and pathological cases of the CTG data, its efficacy in enhancing accuracy is limited [8].

Peterek et al. [6] discovered that utilizing ExtraTreesClassifier for feature selection during the pre-processing stage resulted in a noteworthy enhancement in accuracy, with an increase of up to 4%. In contrast to the results reported by Subasi et al. [8], the utilization of Bagging Ensemble Classifiers led to a marginal improvement in accuracy of merely 0.4%. The Random Forest algorithm exhibited superior performance in the majority of the studies. XGBoost exhibited a comparable level of accuracy to that of Random Forest. The present study's outcomes have led us to adopt a combination of XGBoost and Random Forest algorithms, while incorporating ExtraTreesClassifier for feature selection. Furthermore, we aim to investigate the potential of LightGBM as a novel classifier in our research.

III. Methodology

A. Dataset

Cardiotocography is a tool used by doctors or nurses to see the health condition of the fetus of a mother who is experiencing pregnancy or childbirth. The dataset of Cardiotocography contains data obtained from the results of pregnant women classified by experienced doctors. Cardiotocography data is obtained through the CEFM (Continuous Electronic Fetal Monitoring) test, where data could be recorded manually or automatically using a machine [7]. The dataset used for this experiment was obtained from the Kaggle website - https://www.kaggle.com/datasets/andrewmvd/fetal-health-classification. The dataset contains 2126 records obtained from the Cardiotocography test with 21 features, 13 of which are discrete and the other 8 are continuous. The records were then classified by three obstetricians into 3 types of uterine status, Normal, Suspect and Pathological. The class distribution of this dataset is not balanced, with 1655 of the records are Normal, 295 are Suspect and 176 are Pathological

B. Preprocessing

The first step in the preprocessing phase was the normalization of data. Data normalization is a method employed to standardize data values within a predetermined range, such as 0 to 1. The objective of this modification is to guarantee that each characteristic makes an equivalent contribution to the ultimate outcome, as stated in reference [9]. The Standard Scaler from the preprocessing package in the Sklearn library was employed in this experiment. The Standard Scaler was selected due to the rationale that datasets containing variables or features that span varying ranges may not have an equal impact on computational results, as stated in reference [9]. Moreover, the Standard Scaler confers an additional benefit of alleviating the constraints of diverse machine learning algorithms while handling variables that span a wide or heterogeneous range.

The subsequent stage of our study encompassed Feature Selection, a preprocessing technique in machine learning that identifies an optimal subset of features from a larger set. The process of feature selection has been shown to improve model performance, interpretability, and data dimensionality, while simultaneously decreasing computational expenses. The present study employed two distinct feature selection techniques, namely Information Gain and K-Best.

The Information Gain technique is a feature selection method that employs entropy to determine the Information Gain value of individual features. The utilization of Information Gain facilitates the establishment of a ranking. Subsequently, the system proceeds to choose the highest 'K' features by utilizing the aforementioned ranking. The mutual_info_classif method was utilized in this experiment to calculate the Information Gain.

The K-Best technique represents an alternative strategy for conducting feature selection. The feature selection process involves the identification of the optimal 'K' features through the application of statistical scores, which are computed using a designated score function such as chi-square, ANOVA F-value, and other relevant metrics. The choice of the score function is contingent upon the particular case under consideration. A statistical score is assigned to each feature, indicating its significance or pertinence. Analogous to the Information Gain technique, a feature ranking is conducted utilizing a statistical score, and the system selects the highest 'K' ranked features based on this score. The f_classif method, specifically the K-Best variant, was employed in the present study.

The two Feature Selection techniques delineated above are initially employed independently in the algorithm and subsequently merged via the Pipeline approach for algorithmic implementation.

C. Light Gradient Boosting Machine

LightGBM is a gradient boosting framework that is utilized for regression and classification tasks, and its name is an abbreviation for Light Gradient Boosting Machine. LightGBM and XGBoost share the approach of partitioning data into subgroups and utilizing boosting methodologies to create decision trees; however, their execution varies [10].

Similar to XGBoost, LightGBM works by constructing numerous decision tree and training them iteratively. The new trees are trained to correct the mistakes made by previous tree, therefore minimizing the difference between the predicted labels and the true labels. Once these trees are trained, LightGBM combined the prediction of all the trees and uses it to predict new data.

LightGBM introduced new techniques :

1) Gradient-based One-Side Sampling (GOSS): THe purpose of Gradient-based One-Side Sampling is to focuses on the instances that have the most impact on improving the model's performance while effectively discarding or downsampling less informative instances. LightGBM uses the gradients of each features to rank those features and use the features with higher gradient and drop or downsampling those with lower gradient.
2) Exclusive Feature Bundling Technique (EFB): Exclusive feature bundling technique aims to identify and group together features that have similar interactions with other features. It is designed to improve the efficiency and effectiveness of feature interactions during the training process. Exclusive feature bundling technique uses greedy algorithm to identify groups of features that have high interaction scores with each other but low interaction scores with other features.

To summarize, LightGBM is a powerful algorithm that employs gradient boosting methodologies to develop precise and effective models for both regression and classification assignments.

D. Hyperparamter Tuning

Hyperparameter tuning is a technique utilized to enhance the performance of machine learning models. The process entails identifying the optimal hyperparameter configuration for a specific model [11]. Hyperparameters are predetermined values that are not subject to learning during the training of a model. As such, they must be specified prior to the commencement of the model training process. Hyperparameters are a set of parameters that are not learned during the training process but are instead set prior to training. Examples of hyperparameters include the learning rate, which determines the step size at each iteration of the optimization algorithm, and the number of hidden layers in a neural network, which affects the model's capacity to learn complex representations. The primary aim of hyperparameter tuning is to achieve an optimal equilibrium where the resultant model does not exhibit either overfitting or underfitting of the data.

The purpose of Hyperparameter Tuning implementation is to achieve the best possible result. Hyperparameter Tuning is able to make LightGBM more adaptable toward specific properties of the dataset. By exploring different sets of parameter combination, we can identify optimal set of parameters to further increase our model performance.

The technique of Grid Search is frequently employed for the purpose of hyperparameter tuning. The approach entails training the model through exhaustive iterations of the provided hyperparameters, without any optimization or heuristic techniques. The optimal set of hyperparameters is selected based on its ability to achieve the highest performance on the validation set, as stated in reference [10]. Despite its simplicity, Grid Search can be computationally demanding and time-consuming due to its exhaustive exploration of a pre-defined set of hyperparameters.

The present investigation employed Grid Search technique with a 10-fold cross-validation approach. The hyperparameters that were investigated for XGBoost encompassed max_depth, learning_rate, n_estimators, subsample, and colsample_bytree. The Random Forest algorithm was configured with hyperparameters, namely n_estimators and max_depth. The LightGBM model was tuned using the hyperparameters of num_leaves and learning_rate.

IV. Experiment and Discussion

The present study was designed to assess the efficacy of our machine learning models by conducting experiments across six distinct scenarios. In the initial scenario, the models were utilized on the unprocessed data without any prior preprocessing procedures. In the second scenario, feature selection was implemented through the utilization of the Information Gain technique, whereas in the third scenario, feature selection was carried out via the K-Best method. The feature selection technique employed in the fourth scenario involved the integration of both Information Gain and K-Best. In the fifth experimental setting, we conducted hyperparameter tuning through Grid Search technique on the raw data, commonly known as Vanilla data. The sixth scenario involved a comprehensive approach that integrated various techniques, including preprocessing using Information Gain and K-Best feature selection methods, and subsequent hyperparameter tuning via Grid Search. In this research, data Normalization is not applied to all features, only applied to continuous features. The implementation of Data Normalization to discrete feature may leads to loss of interpretability. The diverse situations presented distinct perspectives on the efficacy of our models across varying circumstances, contributing to the thorough assessment of their prognostic aptitude. Table I, Table II, Table III, Table IV, Table V, and Table VI represent the scenarios respectively.

TABLE I
RESULT OF ALGORITHMS WITHOUT PREPROCESSING

Method	Precision	Recall	F1-Score	Accuracy
XGBoost	93	0.95	0.94	0.96
Random Forest	91	0.75	0.82	0.92
LightGBM	97	0.91	0.94	0,96

TABLE II
RESULT OF USING INFORMATION GAIN AS FEATURE SELECTION

Method	Precision	Recall	F1-Score	Accuracy
XGBoost	94	0.92	0.93	0.95
Random Forest	91	0.76	0.82	0.92
LightGBM	97	0.93	0.95	0.97

TABLE III
RESULT OF USING K-BEST AS FEATURE SELECTION

Method	Precision	Recall	F1-Score	Accuracy
XGBoost	92	0.93	0.93	0.95
Random Forest	89	0.76	0.81	0.91
LightGBM	97	0.94	0.95	0.97

TABLE IV
RESULT OF USING BOTH METHOD AS FEATURE SELECTION

Method	Precision	Recall	F1-Score	Accuracy
XGBoost	94	0.92	0.93	0.96
Random Forest	81	0.80	0.85	0.92
LightGBM	95	0.93	0.96	0.96

TABLE V
RESULT OF USING GRID SEARCH AS HYPERPARAMETER TUNING

Method	Precision	Recall	F1-Score	Accuracy
XGBoost	94	0.93	0.93	0.96
Random Forest	93	0.87	0.90	0.95
LightGBM	97	0.95	0.96	0.97

TABLE VI
RESULT OF USING ALL THE TECHNIQUES

	Precision	Recall	F1-Score	Accuracy
XGBoost	94	0.90	0.92	0.96
Random Forest	94	0.89	0.91	0.95
LightGBM	97	0.94	0.95	0.97

The tables depict the evaluation metrics of the three algorithms XGBoost, Random Forest, and LightGBM in a variety of circumstances.

The results of applying these algorithms to raw data are displayed in the first table. In terms of precision, F1-Score, and accuracy, LightGBM outperforms the other two, while XGBoost performs marginally better in recall. The second table displays the results of feature selection using Information Gain. The LightGBM algorithm once again demonstrated superior precision, F1-Score, and accuracy. However, XGBoost experienced a minor decrease in recall while maintaining the same level of accuracy. The third table depicts the outcomes of implementing the K-Best method for feature selection. The precision and accuracy of Random Forest decreased slightly in comparison to previous scenarios. LightGBM, meanwhile, maintained a solid performance across all metrics. The fourth table depicts the results when both the Information Gain and K-Best feature selection procedures were utilized. While XGBoost and LightGBM maintained relatively stable performance, Random Forest demonstrated a substantial decrease in precision but an increase in recall and F1-Score. The fifth table displays the results of tuning hyperparameters using Grid Search. All three algorithms demonstrated improved performance in comparison to their results without preprocessing, with LightGBM maintaining the lead. In the final table, the results of implementing all techniques, including preprocessing, feature selection, and hyperparameter optimization, are displayed. While all three algorithms maintained a high level of performance, LightGBM maintained the highest precision, F1-Score, and accuracy in every scenario. In general, these tables demonstrate the significance of preprocessing, feature selection, and hyperparameter tuning for improving the efficacy of machine learning algorithms.

By analyzing the outcomes of the various experimental scenarios, we can derive a number of essential insights. Without preprocessing, LightGBM achieved the highest precision and F1-Score, and tied with XGBoost for the highest accuracy. Nonetheless, when Information Gain was used for feature selection, all of the algorithms marginally improved, with LightGBM once again achieving the highest precision, F1-Score, and accuracy. When K-Best was used to select features, XGBoost and LightGBM had the same F1-Score and Accuracy, but XGBoost's precision decreased marginally. The performance of Random Forest remained comparatively stable across the first three scenarios, with only minor metric modifications. Interestingly, Random Forest's precision decreased when Information Gain and K-Best were used for feature selection, while XGBoost and LightGBM's performance remained consistent with their performance in the previous scenarios. Grid Search tuning of hyperparameters resulted in significant performance enhancements for all three algorithms, with Random Forest demonstrating the greatest improvement in precision and recall. In the final scenario, which combined all techniques, there was a minor decrease in recall for XGBoost and an increase in recall for Random Forest, resulting in comparable F1-scores for the two algorithms. LightGBM, on the other hand, consistently maintained the maximum performance across all metrics, indicating its adaptability to various preprocessing, feature selection, and hyperparameter tuning techniques.

Another result for this research is the purpose of Standard Scaler implementation is to ensure each features have the equal contributions toward the final outcome. It also decrease the time and resource to train the model, with the potential to affect model performance negatively.Data Normalization is not applied to all features, only applied to continuous feature. While the implementation of Feature Selection techniques is to improve the model performance by selecting the most relevant and informative feature. Similar to Standard Scaler, Feature Selection also decrease the time and resource to train the model as it reduces the dimensionality of the dataset, reducing the possibility of overfitting.

V. Conclusion

This study demonstrated the effectiveness of XGBoost, Random Forest, and LightGBM in predicting embryonic health status using CTG data. The implementation of various preprocessing techniques, feature selection strategies, and hyperparameter tuning techniques revealed their impact on the performance of these algorithms. Despite the fact that all three algorithms demonstrated potential for this endeavor, LightGBM consistently delivered the highest performance across all metrics and techniques. This result suggests that LightGBM may be the most robust and dependable model for this particular endeavor. Notable is the fact that the performance enhancement brought about by the use of Bagging Ensemble Classifiers and the feature selection methods Information Gain and K-Best was minimal. This observation may suggest that the initial data were already in a reasonably effective format for the task, and that additional preprocessing and feature selection provided only marginal enhancements. Last but not least, hyperparameter tuning using Grid Search yielded significant improvements, particularly for the Random Forest algorithm, highlighting the significance of this step in optimizing machine learning models.

References

[1] J. E. Lawn, H. Blencowe, P. Waiswa, A. Amouzou, C. Mathers, D. Hogan, V. Flenady, J. F. Frøen, Z. U. Qureshi, C. Calderwood, S. Shiekh, F. B. Jassir, D. You, E. M. McClure, M. Mathai, S. Cousens, Lancet Ending Preventable Stillbirths Series study group, and Lancet Stillbirth Epidemiology investigator group, “Stillbirths: rates, risk factors, and acceleration towards 2030,” *Lancet (London, England)*, vol. 387, no. 10018, pp. 587–603, Feb. 2016.

[2] W. H. Organization, “Newborn Mortality,” 2021. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/levels-and-trends-in-child-mortality-report-2021

[3] G. Improta, C. Ricciardi, F. Amato, G. D’Addio, M. Cesarelli, and M. Romano, “Efficacy of Machine Learning in Predicting the Kind of Delivery by Cardiotocography,” in *XV Mediterranean Conference on Medical and Biological Engineering and Computing – MEDICON 2019*, ser. IFMBE Proceedings, J. Henriques, N. Neves, and P. de Carvalho, Eds. Cham: Springer International Publishing, 2020, pp. 793–799.

[4] M. Ramla, S. Sangeetha, and S. Nickolas, “Fetal Health State Monitoring Using Decision Tree Classifier from Cardiotocography Measurements,” in *2018 Second International Conference on Intelligent Computing and Control Systems (ICICCS)*, Jun. 2018, pp. 1799–1803.

[5] J. Piri and P. Mohapatra, “Exploring Fetal Health Status Using an Association Based Classification Approach,” in *2019 International Conference on Information Technology (ICIT)*, Dec. 2019, pp. 166–171.

[6] T. Peterek, P. Gajdoš, P. Dohnálek, and J. Krohová, “Human Fetus Health Classification on Cardiotocographic Data Using Random Forests,” in *Intelligent Data analysis and its Applications, Volume II*, ser. Advances in Intelligent Systems and Computing, J.-S. Pan, V. Snasel, E. S. Corchado, A. Abraham, and S.-L. Wang, Eds. Cham: Springer International Publishing, 2014, pp. 189–198.

[7] H. Sahin and A. Subasi, “Classification of the cardiotocogram data for anticipation of fetal risks using machine learning techniques,” *Applied Soft Computing*, vol. 33, pp. 231–238, Aug. 2015. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1568494615002653

[8] A. Subasi, B. Kadasa, and E. Kremic, “Classification of the Cardiotocogram Data for Anticipation of Fetal Risks using Bagging Ensemble Classifier,” *Procedia Computer Science*, vol. 168, pp. 34–39, Jan. 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1877050920303872

[9] V. N. G. Raju, K. P. Lakshmi, V. M. Jain, A. Kalidindi, and V. Padma, “Study the Influence of Normalization/Transformation process on the Accuracy of Supervised Classification,” in *2020 Third International Conference on Smart Systems and Inventive Technology (ICSSIT)*, Aug. 2020, pp. 729–735.

[10] G. Ke, Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, Q. Ye, and T.-Y. Liu, “Lightgbm: A highly efficient gradient boosting decision tree,” in *Proceedings of the 31st International Conference on Neural Information Processing Systems*, ser. NIPS’17. Red Hook, NY, USA: Curran Associates Inc., 2017, p. 3149–3157.

[11] B. H. Shekar and G. Dagnew, “Grid Search-Based Hyperparameter Tuning and Classification of Microarray Cancer Data,” in *2019 Second International Conference on Advanced Computational and Communication Paradigms (ICACCP)*, Feb. 2019, pp. 1–8.