An Empirical and Statistical Analysis of Fetal Health Classification Using Different Machine Learning Algorithm

Conference Paper · April 2024

DOI: 10.1109/ICDECS59733.2023.10503087

4 authors, including:

Nagavibha R.
Christ University, Bangalore

Mishbah Pinto Chandy
Christ University, Bangalore

Prathap Rudra Boppuru
Christ University, Bangalore

4th IEEE International Conference on Data Engineering and Communication Systems (ICDECS)

An Empirical and Statistical Analysis of Fetal Health Classification Using Different Machine Learning Algorithm

Nagavibha R
Department of Computer Science and Engineering
Christ (Deemed to be University)
Bengaluru, India
nagavibha.r@btech.christuniversity.in

Mohith Manoharan Satheesh
Department of Computer Science and Engineering
Christ (Deemed to be University)
Bengaluru, India
mohith.ms@btech.christuniversity.in

Mishbah Pinto Chandy
Department of Computer Science and Engineering
Christ (Deemed to be University)
Bengaluru, India
mishbah.pinto@btech.christuniversity.in

Boppuru Rudra Prathap
Department of Computer Science and Engineering
CHRIST (Deemed to be University)
Bengaluru, India
boppuru.prathap@christuniversity.in

2024 4th International Conference on Data Engineering and Communication Systems (ICDECS) | 979-8-3503-9335-4/24/$31.00 ©2024 IEEE | DOI: 10.1109/ICDECS59733.2023.10503087

***Abstract*— The health of both the mother and the baby is affected by how well the fetus is doing during pregnancy, making it a matter of utmost importance. To achieve the best results possible, it is essential to regularly monitor and intervene when needed. While there are many ways to observe the well-being of the fetus in the mother's womb, using artificial intelligence (AI) has the potential to enhance accuracy, efficiency, and speed when it comes to diagnosing any issues. This study focuses on developing a machine learning-driven system for accurate fetal health classification. The dataset comprises detailed information on the signs and symptoms of pregnant individuals, particularly those at risk or with emerging fetal health issues. Employing a set of ten machine learning models namely Naïve Bayes, Logistic Regression, Decision Tree, Random Forest, KNN, SVM, Gradient Boosting, Linear Discriminant Analysis, Quadratic Discriminant Analysis Light Gradient Boosting Machine (LGBM) along with ensemble-based processes, the Light Gradient Boosting Machine (LGBM) has been identified as a standout performer, accomplishing an accuracy of 96.9%. Furthermore, our exploration demonstrates overall performance like character fashions, signaling promising prospects for sturdy and correct fetal fitness class systems. This study highlights the power of machine learning that could revolutionize prenatal care by identifying fetal health problems early.**

I. INTRODUCTION

Maternal health and fetal mortality have become a growing concern in the world today, 7 deaths have been observed per 1000 live births as of 2018 for the low middle-income countries, and on the other hand for high-income and upper-middle-income countries this value was noted to be 3 and 7 deaths per 1000 live births according to UNICEF. The constant monitoring of fetal health helps in early diagnosis of diseases helps in decision making with respect to both maternal and fetal health, it has also been found that it helps in prevention of perinatal mortality i.e. deaths occurring between the 22nd week of pregnancy and the first 7 days after birth, it on a whole also helps you to be prepared with respect to the future challenges that may occur which are new to the healthcare professionals also.

The utilization of ML algorithms is of utmost importance in the classification of Fetal Health. It helps in easily identifying and controlling any possible risks, thus enabling accurate and continued health monitoring. Simultaneously, it plays a significant role in decision-making processes and ensures efficient allocation of resources. This study aims to compare different models to identify which one is the most appropriate for Fetal Health Classification.

Several conditions affect a baby's development while in the womb. Their severity ranges from one condition to another and they may be caused by genetic factors, maternal illnesses, environmental influences, or problems during pregnancy. A few common fetal health problems Down syndrome, growth restrictions, and infections transmitted from mothers result in fetal distress. Therefore, on-time prenatal care, regular check-up visits to the doctor and diagnostic tests are necessary to identify and manage such issues. On most occasions, early detection enables parents to undertake medical interventions like specialized care or treatment options that can help alleviate risks while maximizing the infant's health outcomes as much as possible.

Analysis of fetal health gives valuable insights into development progressions, potential risks, and general well-being throughout gestation. This involves studying key indicators such as heart rate patterns and movements that provide a comprehensive overview of the state of the fetus. Consequently, this examination helps healthcare providers the possibilities of complications, predict outcomes, and decide whether specialized attention is needed or not. Knowing that this analysis offers a complete picture of fetal health conditions in healthcare teams allows for prophylactic measures to be taken that will make pregnancy and childbirth safer and healthier for both mother and baby alike.

We have considered some of the most prominent classification methods for the proposed work such as, including Naïve Bayesian, K-nearest neighbour, SVM, Decision Tree, Logistic Regression, Random Forest, Gradient Boosting, LDA, QDA, and LGBM over a dataset to obtain a comprehensive understanding of the algorithms 'performance and choosing the most optimum one.

Section 2 outlines previous projects done in this regard, and Section 3 describes the various classification algorithms and methodology used. Section 4 discusses the result obtained, and Section 5 discusses the conclusions from this piece of work.

II. Literature Review

In practice, Cardiotocography is the preferred oblique method for assessing fetal health. The fetal heart rate monitor (CTG) has two different signals, continuous recording of heart rate (FHR) and uterine activity (UC). However, the predictive ability of CTG remains controversial [5]. This has been reviewed by Devoe et al. Based on his review of 45 studies, he found that the sensitivity of using CTG ranged from 2% to 100%, while the specificity ranged from 37% to 100% [2]. According to the analysis of several studies by Devoe et al [5], this is an opportunity to use machine learning to overcome the lack of desired and specific results and eliminate problems with the predictive power of CTG use. Abderrazak Rafie, Salma Chenouni and others. Rafie et al. [1] prepared this study in which data were extracted from UCI [11] and 2126 CTGs were recorded and classified by some obstetricians to calculate uterine contraction and fetal heart rate. Tags represent each document. The data includes normal fetuses, abnormal fetuses, and the remaining fetuses classified as pathological, and results obtained using random forest, support vector technology, and network design (ANN).Şahin et al. [3] used the same UCI dataset [11] as the authors [1] and used 8 different machine learning models, namely Support Vector Machine (SVM), Radial Basis Function Network (RBFN, classification and regression trees, Artificial Neural Network (ANN), C4.5 decision tree classifier, random forests, k nearest. The results of this study show that random forest can be considered as a good classification system for normal and diseased CTG data. Rahmayanti, N., Pradani, H., Pahlawan, M. and Vinarti, R. et al. [4] conducted a study in which they employed various approaches to achieve their results. The first approach involved the removal of outliers and the utilization of the same data from the upsampling process, albeit with variable multicollinearity left unaddressed. The second approach, on the other hand, entailed the retention of outliers and the utilization of normal data without pre-synchronization. Lastly, the third approach entailed the utilization of pre-processed data, which incorporated the exclusion of anomalies, the removal of multiple variables, and the implementation of a top-down model to ensure data equilibrium. RF, XGB, SVM, LGBM, and KNN were the models that were being experimented upon and they had an accuracy than ranged from 89%-99%. Hoodbhoy et al. [6] have obtained the similar dataset to the [4] and have used XGBoost, Random Forest and Decision Tree and have concluded that these models have a high precision [>96] and that the XGBoost model has the highest precision among all the other implementations. Cömert and Kocamaz [9] utilized a similar dataset[11] and conducted various machine learning methods, including Artificial Neural Networks (ANN), Support Vector Machines (SVM), Extreme Learning Machines (ELM), Radial Basis Function Networks (RBFN), and Random Forest. Among these techniques, ANN showcased superior efficacy, achieving a sensitivity of 99.73% and specificity of 97.94%. The ELM approach was thoroughly investigated, with a focus on tuning parameters such as activation functions and the number of nodes in hidden layers, spanning a wide range of values. The performance metrics were utilized to compare the results of each classifier to one another.Subasi et al. [7] decided to evaluate student performance using UCI's CTG dataset [8]. The CTG data were analyzed by three expert gynaecologists to determine whether the embryo was normal or pathological. CTG data has 21 features, 13 of which are discrete and 8 are continuous. Upon completion, approximately 1,831 pieces of CTG data were tested correctly. They use the Bagging Ensemble Classifier, an ensemble of meta-predictors that fits each base classifier to the variance of the original data and then aggregates their predictions （via voting or averaging） to make the final prediction. Such meta-estimators can often be used to reduce the variance of black box predictions （such as decision trees） by introducing randomness into their construction process and then combining with it. If this situation creates imbalance, it can be easily used in a design. This may include stopping overfitting protection used in some algorithms. Models that fit the bootstrap paradigm are more likely to change. Creating specific data models based on several models without integrating them into a whole and choosing the best one does not provide satisfactory solutions. Maximum 99.02% distribution has been implemented using different machine learning techniques such as SVM, k-NN, ANN, Random Forest, CART, C4.5, REP tree and packing of mixed objects by Random Forest.

Dixit[10] also evaluated the literature containing data on cardiotocography (CTG) and various indicators of fetal health. It contains 2,130 cases with 21 features, providing a general framework for assessing and predicting fetal health [11]. This product includes important measurements such as abdominal heart rate, acceleration, deceleration, and various indices obtained from the CTG signal. Additionally, the dataset includes fetal health classification, with three categories indicating whether normal, abnormal, or fetal conditions are present. The authors analyzed the data using algorithms such as logistic regression, k-nearest neighbours, SVM, decision trees, Extra Trees, Random Forest, Gradient Boosting and NN MLP and concluded that their tree model plus achieved the highest value among all standard exposures. The standard is 93.66%. It also has high recall and accuracy of 93.66% and 93.82% respectively. This shows that the Extra Trees model performs very well in determining the correct values at very low cost. The F1 score is 93.71%, indicating the balance between precision and recall. An e-health service and prediction system for pregnant women was developed by Akbulut et al. [12]. Data to forecast the unborn child based on maternal data and medical availability was created using the performance benchmark, which was trained using clinical data of 96 pregnant women. At Istanbul RadyoEmar Diagnostic Radiology Center, information was gathered from parent surveys and the thorough reports of three physicians. This study has a prediction accuracy of more than 89.5% in the decision forest model's development assessment. A proposed methodology, put forth by Attallah et al. [13] entails a four-step process. These steps encompass feature extraction, classification, segmentation, and optimization. Initially, the developing fetus is partitioned into regions of interest, with the remaining part being addressed during the segmentation procedure. Subsequently, abnormal regions are extracted from comprehensive brain scans. In an effort to tell apart normal and diseased fetal cells, a series of statistical attributes are computed during the feature extraction stage after the image has undergone a Discrete Wavelet Transform (DWT). Subsequent to this, the model is trained and tested using a diverse range of classification models, such as SVM, LDA, K-NN, and subspace classification, in the classification phase. Georgulas et al. [14] have presented a novel approach for the extraction and categorization of fetal heart rate signals, which enables the correlation between fetal heart rate and umbilical artery pH during labor. The proposed methodology involves the extraction of time scale-related features from FHR signals utilizing Discrete Wavelet Transform (DWT), followed by their classification using Support Vector Machine (SVM). This method has been evaluated against various sets of gastrointestinal organism's data pertaining to blood pH, and the results have demonstrated its efficacy, achieving complete classification.

In order to enhance the efficacy of diagnosing pregnant women with Diminished Fetal Movement (DFM), Piri and Mohapatra [15] have put forward a proposition to establish a system for devising a model employing associative classification techniques in order to assess fetal movement. The authors have successfully executed a variety of algorithms, which encompass the CBA-M1 and CBA-M2 algorithms. By conducting a comparison with a select few standard classifiers, it has been ascertained that Random Forest and XGBoost exhibit tremendous performance for the intended objective. Jie and Zhizhong [16] collaborated with hospitals and EFM systems to obtain 4473 data files, including 3012 normal data, 1024 abnormal data, and 437 abnormal records. To analyze the products, the authors use CNN, MLP and SVM models. CNN processes the data in parallel after splitting the multivariate FHR into ten d-window segments. The test results of our experiments show that the accuracy of CNN is higher than MLP and SVM.Kuzu and Santur [17] proposed a new and effective method to diagnose fetal health and distinguish fetal status between binary and multiclass based on FHR and UC data. Summary data can be found here [11]. Based on the data used, binary classification (Normal (N), Pathological (P)) and multi-classification (Normal (N), Disease (P), Suspicious (S)) were created. The data uses Random Forest (RF), Horizontal Linear (LR), Decision Tree (DTC),, Random Sequence (XGB), Support Vector Machine (SVM) and AdaBoost (AB) methods. 23 characters are used to enter the network. Models created using XGBoost technology predict fetal development with 100% accuracy. Muhammad Hussain et al. [18] implemented and evaluated each deep neural procedure [11] on the CTG dataset. These include AlexNet, Random Forest, GoogleNet, DesnseNet, NiftyNet , Recurrent Neural Networks and the planned SVM AlexNet.Encompassing Process includes Includes Involving The authors used the recently created AlexNet-SVM hybrid architecture, which comprises of an input layer, an overlay layer, an upgraded SVM full connection layer, and an output layer. The AlexNet algorithm learns from filters via a convolution method. Extracted features are moved to a layer that contains two or more maps. This methodology outperformed existing architectures, having achieved 99.72% accuracy. Yin and Bingi's [19] aim was to devise a highly efficient technique for categorizing fetuses according to their risk of mortality using cardiotocogram data. Comprehensive information regarding the dataset can be found in the source [11]. After careful adjustments, the suggested models accomplished 99.19% accuracy when the light gradient boost was used, 98.79% accuracy when the extreme gradient boost was applied and an astounding 99.59% accuracy when the support vector machine was used in relation to our test data.. These three models are not only exceedingly precise but also expeditious, rendering them suitable for implementation worldwide, particularly in scenarios where it is impractical for an obstetrician to assess fetal well-being on an individual basis. The authors have effectively elucidated their findings through the utilization of methodologies such as SHAP (SHapley Additive exPlanations), LIME (Local Interpretable Model-Agnostic Explanations), and FAB （Feature Altering for Explanations of Black Box Models).Bhowmik et al. [20] reported the analysis of prenatal CTG data [11] and established a useful tree-based learning (EL) benchmark for predicting fetal health. In this study, the Stacking process in EL is introduced and a brief summary of the method is given. Stacking EL technology is proposed in this paper using four tree learning machines as base learners: random forest classifier, deep forest classifier, decision tree classifier and more tree classifiers.Shruthi and Poornima [21] employ the dataset provided by Kaggle in order to make this prediction. Despite the fact that the dataset [11]. The variable of interest in this dataset is the fetal health status, which can be categorized as either normal, suspect, or pathological. The authors delve into a comprehensive discussion regarding the utilization of the Random Forest Classifier and assert that the ensemble technique involves the utilization of two distinct approaches, namely bagging and boosting. Abiyev et al. [22] recommend using a type-2 fuzzy neural network (T2-FNN) to assess the health of the fetus. The design techniques and architecture of the T2-FNN system are presented. A gradient algorithm and cross-validation technique are used in the system's construction. The created T2-FNN is tested with fetal datasets. It is found that the system operates more efficiently the more rules there are. The authors compare the outcomes of various systems used to evaluate the fetus's health with the T2-FNN approach.

III. METHODOLOGY

A conventional workflow for training and evaluating a machine learning classification model is shown in Figure(1).

Input data: The input data can be understood as the raw data which is taken as an input for the simulation that may contain features that can contribute to predicting outcomes.

Split data: The input data is then split into two separate sets namely test and train. This is a crucial step in machine learning to ensure that the model can be trained on 70% of the data and tested on a separate 30% data to evaluate its performance.

The 70/30 split is a common heuristic in machine learning for partitioning datasets. It aims to provide a balanced approach to training a model with enough data while keeping a significant portion for testing its performance. This split is justified by the need for sufficient training data, which is typically 70%, and adequate testing data, which is 30%, to ensure a substantial amount of unseen data for evaluation. The 70/30 split helps balance overfitting and underfitting by providing enough data for learning while reserving a significant portion for unbiased evaluation. It has been empirically successful for a wide range of problems, especially when the dataset size is moderate. It also offers computational efficiency, as training complex models on large datasets can be computationally expensive and time-consuming.

Classification models: Classification algorithm is chosen to create a model for the given data, the training data is then used to train the model subsequently during which the model will have learned to associate input features with the target variable as per the patterns found in training data.

Trained model: The classification model can be trained using the training set, with parameters adjusted to better fit the data.

Test model: The trained model is then tested with the help of the test data which is a key indicator as the model will come across data that it has never come across before.

Performance evaluation: Test model and the test data goes through the performance evaluation phase that helps us tell how well the model has performed with the help of indicators like accuracy, F1 score, precision etc.

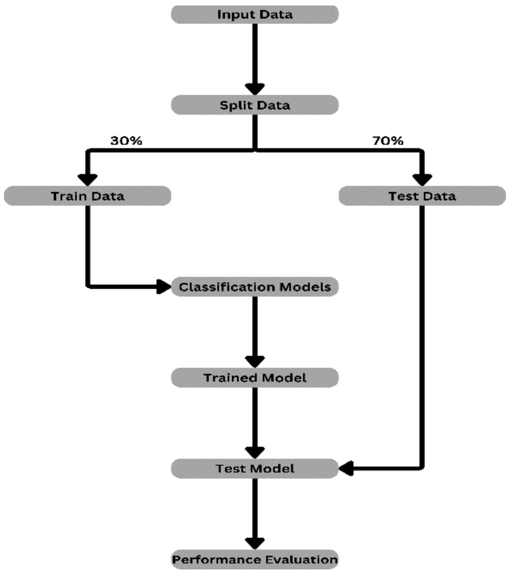

Figure (1): Workflow of ML classification models

IV. Result and discussion

The research experimented on the ten different Machine Learning algorithms were applied on the CTG records of 2126 pregnant women to predict an adverse fetal outcome [11]. The data is loaded into the csv file and then the data preprocessing takes place.

4.1 Data Pre-processing

4.1.1 Data Balancing

Preventing the model from developing bias towards a single class is one of the benefits of balancing a dataset during model training. To put it another way, simply because the model has more data, it will no longer favor the majority class. To create balanced datasets, we modified the unequal data classes in this paper using the up-sampling technique. The oversampling method attempts to achieve equilibrium by increasing the size of rare samples when the amount of data is not sufficient. Synthetic Minority Over-Sampling Technique (SMOTE) is the concept that we have employed. When using SMOTE, examples that are close to each other in the feature space are chosen, a line is drawn between them, and a new sample is drawn at a point along the line. Initially the dataset had 77.1% for normal, 13.8% for suspicious and 8.28% for pathological. After SMOTE technique of balancing the data normal is 33.3%, suspicious is 33.2% and pathological is 33.4%.

4.1.2 Scaling

Data scaling involves adjusting feature values within a predefined range, enhancing algorithmic performance by adjusting the minimum and maximum values such as 0 to 1 or -1 to 1.

4.2 Hyper parameter Tuning

The optimization of performance for a machine learning algorithm often involves the systematic adjustment of configuration settings, known as hyper parameters. This process, referred to as hyper parameter tuning, is commonly achieved through the utilization of techniques such as grid search or random search. Table (1) depicts about results comparison of ML models before hyperparameter tuning and Table (2) depicts about results comparison of ML models after hyperparameter tuning.

	Model	Accuracy	Precision	Recall	F1_Score
0	Naive Bayes	0.761	0.793	0.761	0.765
1	Logistic Regression	0.889	0.894	0.889	0.890
2	Decision Tree	0.918	0.920	0.918	0.918
3	Random Forest	0.960	0.961	0.960	0.960
4	KNN	0.937	0.941	0.937	0.937
5	SVM	0.907	0.912	0.907	0.908
6	Gradient Boosting	0.947	0.949	0.947	0.947
7	LDA	0.883	0.895	0.883	0.884
8	QDA	0.334	0.111	0.334	0.167
9	LGBM	0.964	0.964	0.964	0.964

Table (1): Result of the ten models before hyper parameter tuning.

	Model	Accuracy	Precision	Recall	F1_Score
0	Naive Bayes	0.761	0.793	0.761	0.765
1	Logistic Regression	0.889	0.894	0.889	0.890
2	Decision Tree	0.918	0.919	0.918	0.918
3	Random Forest	0.958	0.958	0.958	0.958
4	KNN	0.948	0.952	0.948	0.947
5	SVM	0.933	0.940	0.933	0.934
6	Gradient Boosting	0.962	0.963	0.962	0.962
7	LDA	0.880	0.891	0.880	0.881
8	QDA	0.782	0.840	0.782	0.786
9	LGBM	0.969	0.969	0.969	0.969

Table (2): Result of the ten models after hyper parameter tuning.

4.3 Data Visualization:

1) Box-Wisker plot: Figure (2) shows about box-wisker plot. The highest to lowest mean accuracy in a dataset is determined by various factors, including mean values and monitoring information dissemination. The chart displays boxes sorted from highest to lowest mean accuracy, with outliers found outside whiskers. The left

is 25%, the right's 75% with a median of 50%.

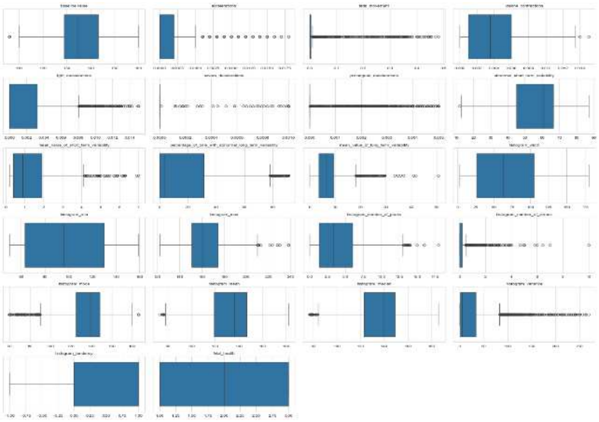

Figure (2). Box-Wisker Plot

2) Histogram: Figure (3) shows about histogram visualization. A histogram is a bar graph displaying data as buckets of classes, with the x-axis representing data classes and the y-axis representing data occurrences. In this figure, it represents about baseline value plotted against the number of occurrences.

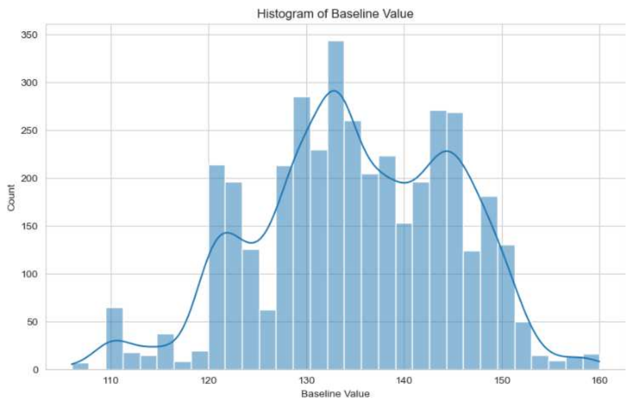

Figure (3). Histogram of base values

3) Confusion matrix:
The confusion matrix, a tool used to evaluate the effectiveness of a classification model, includes four distinct predicted and actual class combinations. The control group is normal, while the case group is suspicious or abnormal. It is a very useful tool for measuring precision, recall, accuracy, F1 score. The confusion matrix for the Light Gradient Boosting Model, which provides the prediction values, is shown in Table (3). There were 43 inaccurate forecasts out of a total of 1150 correct ones. The model's classification report is shown in Table (4). The total F1 score attained in this case is 96.33%. The individual F1-score is 96% for normal, 95% for suspect and 98% for pathological.

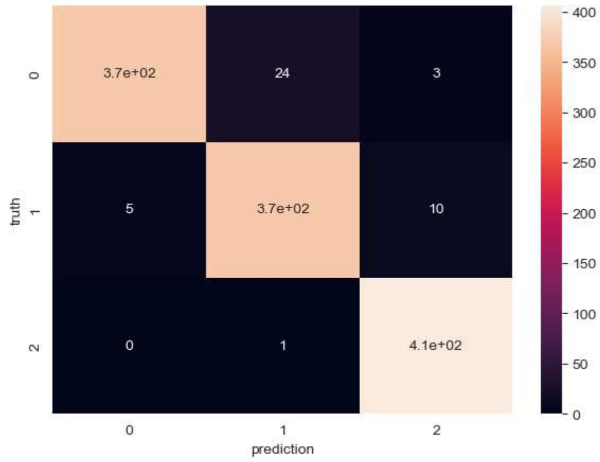

Table (3): Result of the confusion matrix for LGBM

Classification report:

	precision	recall	f1-score	support
1	0.99	0.93	0.96	396
2	0.94	0.96	0.95	384
3	0.97	1.00	0.98	407
accuracy			0.96	1187
macro avg	0.96	0.96	0.96	1187
weighted avg	0.96	0.96	0.96	1187

Table (4): Result of the classification report for LGBM

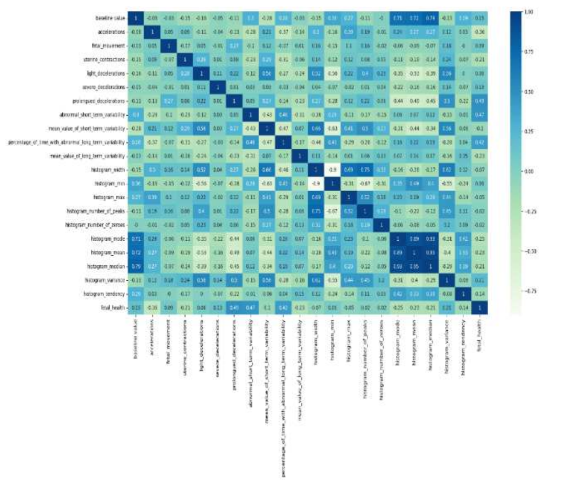

Figure (4). Heat map representing the features

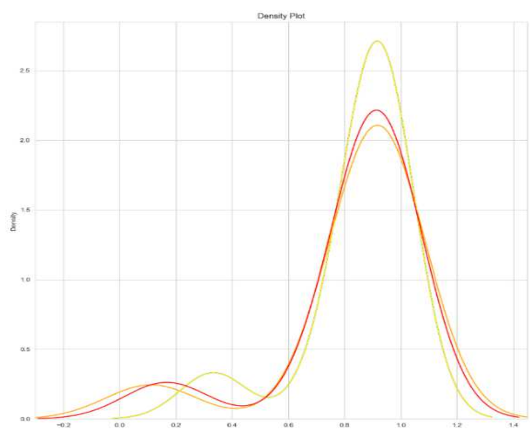

Figure (5): Density plot representing the accuracy, precision, recall and f1-score of all the models.

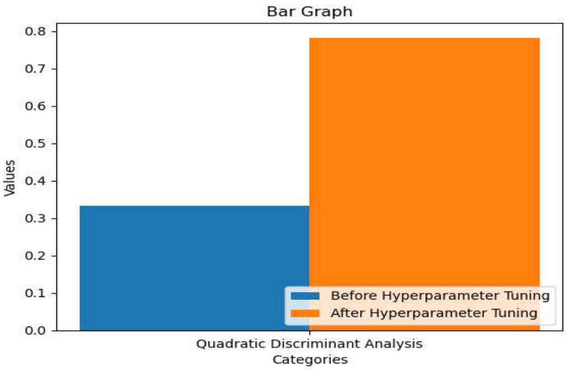

Figure (6): Accuracy of QDA model before and after hyper-parameter tuning.

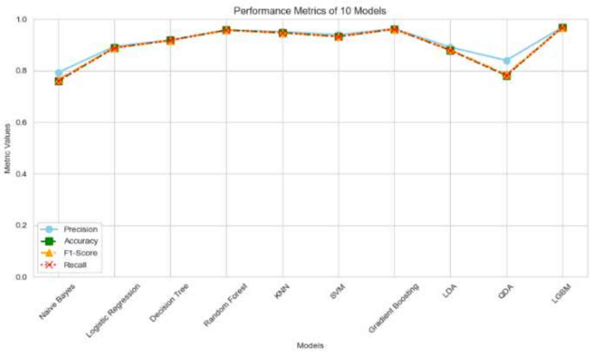

Figure (7): Accuracy, precision, recall and f1-score of all ML models.

Figure (4) depicts about the heat map representing all the features in the data set. Figure (5) shows the density plot based on the recall, f1-score, accuracy and precision of all the models. Figure (7) visualizes a line chart about the performance of all the machine learning models. Table (1) shows the results of the ML models before hyperparameter tuning and Table (2) shows the results of ML models after hyperparameter tuning. The study compares machine learning models before and after hyperparameter tuning, revealing the strengths and weaknesses of each approach. LGBM and Random Forest show superior performance with high scores across accuracy, precision, recall, and F1-score. QDA underperforms with scores not exceeding 0.334. Post-tuning, LGBM maintains its lead, slightly improving slightly to near-perfect performance and the best hyperparameters used to improve its performance are learning rate 0.164, max depth 16, min child samples 17, n-estimators 162. However, QDA sees the most dramatic improvement, with all scores more than doubling, highlighting the potential of hyperparameter tuning in enhancing model performance with reg param 0.1 as its best hyperparameter. Figure (6) shows the change in accuracy pre-tuning and post-tuning. LGBM consistently outshines other models due to its robustness, reliability, ability to handle large datasets efficiently and importance to feature selection.

V. Conclusion

The dataset from the publicly available University of California, Irvine Machine Learning Repository [11] was the subject of a benchmark study that we conducted for this publication. The dataset contains data of 2,126 pregnant women in the last phase of their pregnancies which were included in this study. Twenty-one characteristics from this dataset are utilized in the CTG FHR and UC measurements. We classified the fetus into three classes—normal, suspicious, and pathological—using the UCI dataset to validate this experiment. The dataset undergoes pre-processing data entry, whereby data is first smoothed via up-sampling and then scaled to fit inside a specific range of values.

This data is then evaluated using a confusion matrix, where we analyze precision, recall, F1-score, and finally support, which gives us precision. We experiment with CTG database on ten different machine learning models including Naive Bayes, SVM, Decision Tree, Random Forest, in, Gradient Boosting, Linear Discriminant Analysis, Quadratic Discriminant Analysis, Light Gradient Boosting Machine (LGBM), Logistic Regression. We determined the most accurate learning model, which is LGBM, which gives an accuracy of 96.9%, followed by random forest with 96.2%. The future scope of the comparison is to identify more accurate and accurate models that will affect society and help detect fetal abnormalities during pregnancy.

References

[1] Rafie, A., Chenouni, S., Alami, N., El Mallahi, M., Blali, A., & Tahiri, A. (2022). Classification of fetal state using machine learning models. In *E3S Web of Conferences* (Vol. 351, p. 01027). EDP Sciences.

[2] Devoe, L.D, R. A. Castillo, and D. M. Sherline. (1985) “The nonstress test as a diagnostic test: A critical reappraisal.” Am. J. Obstet. Gynecol 152 (8): 1047–1053

[3] Sahin, H., & Subasi, A. (2015). Classification of the cardiotocogram data for anticipation of fetal risks using machine learning techniques. *Applied Soft Computing*, *33*, 231-238.

[4] Rahmayanti, N., Pradani, H., Pahlawan, M., & Vinarti, R. (2022). Comparison of machine learning algorithms to classify fetal health using cardiotocogram data. Procedia Computer Science, 197, 162-171.

[5] Ayres-De-Campos, D., C. Costa-Santos, and J. Bernardes. (2005) “Prediction of neonatal state by computer analysis of fetal heart rate tracings: The antepartum arm of the SisPorto® multicentre validation study.” Eur. J. Obstet. Gynecol. Reprod. Biol 118 (1): 52-60.

[6] Hoodbhoy, Z., Noman, M., Shafique, A., Nasim, A., Chowdhury, D., & Hasan, B. (2019). Use of machine learning algorithms for prediction of fetal risk using cardiotocographic data. *International Journal of Applied and Basic Medical Research*, *9*(4), 226.

[7] Subasi, A., Kadasa, B., & Kremic, E. (2020). Classification of the cardiotocogram data for anticipation of fetal risks using bagging ensemble classifier. *Procedia Computer Science*, *168*, 34-39.

[8] K. Bache and M. Lichman, “UCI Machine Learning Repository [http://archive. ics. uci. edu/ml]. University of California, School of Information and Computer Science,” Irvine CA, 2013.

[9] Cömert, Z., & Kocamaz, A. D. N. A. N. (2017). Comparison of machine learning techniques for fetal heart rate classification. *Acta Physica Polonica A*, *132*(3), 451-454.

[10] Dixit, R. R. (2022). Predicting Fetal Health using Cardiotocograms: A Machine Learning Approach. *Journal of Advanced Analytics in Healthcare Management*, *6*(1), 43-57.

[11] Ayres-de Campos D, Bernardes J, Garrido A, Marques-de-Sá J, Pereira-Leite L. SisPorto 2.0: A program for automated analysis of cardiotocograms. J Matern Fetal Med 2000;9:311-8.

[12] Akbulut, A., Ertugrul, E., & Topcu, V. (2018). Fetal health status prediction based on maternal clinical history using machine learning techniques. *Computer methods and programs in biomedicine*, *163*, 87-100.

[13] Attallah, O., Gadelkarim, H., & Sharkas, M. A. (2018, December). Detecting and classifying fetal brain abnormalities using machine learning techniques. In *2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA)* (pp. 1371-1376). IEEE.

[14] Georgoulas, G., Stylios, C., & Groumpos, P. (2006). Feature extraction and classification of fetal heart rate using wavelet analysis and support vector machines. *International Journal on Artificial Intelligence Tools*, *15*(03), 411-432.

[15] Piri, J., & Mohapatra, P. (2019, December). Exploring fetal health status using an association based classification approach. In *2019 International Conference on Information Technology (ICIT)* (pp. 166-171). IEEE.

[16] Li, J., Chen, Z. Z., Huang, L., Fang, M., Li, B., Fu, X., ... & Zhao, Q. (2018). Automatic classification of fetal heart rate based on convolutional neural network. *IEEE Internet of Things Journal*, *6*(2), 1394-1401.

[17] KUZU, A., & SANTUR, Y. (2022, November). Fetal Health Pattern Classification Using Ensemble Learning. In *Proceedings of the 3rd International Conference on Applied Engineering and Natural Sciences, Konya, Turkey* (pp. 10-13).

[18] Muhammad Hussain, N., Rehman, A. U., Othman, M. T. B., Zafar, J., Zafar, H., & Hamam, H. (2022). Accessing artificial intelligence for fetus health status using hybrid deep learning algorithm (AlexNet-SVM) on cardiotocographic data. *Sensors*, *22*(14), 5103.

[19] Yin, Y., & Bingi, Y. (2023). Using Machine Learning to Classify Human Fetal Health and Analyze Feature Importance. *BioMedInformatics*, *3*(2), 280-298.

[20] Bhowmik, P., Bhowmik, P. C., Ali, U. M. E., & Sohrawordi, M. (2021). Cardiotocography data analysis to predict fetal health risks with tree-based ensemble learning. *Inf. Technol. Comput. Sci*, *5*, 30-40.

[21] Shruthi, K., & Poornima, A. S. (2023). A Method for Predicting and Classifying Fetus Health Using Machine Learning. *International Journal of Intelligent Systems and Applications in Engineering*, *11*(2), 752-762.

[22] Abiyev, R., Idoko, J. B., Altıparmak, H., & Tüzünkan, M. (2023). Fetal Health State Detection Using Interval Type-2 Fuzzy Neural Networks. *Diagnostics*, *13*(10), 1690.