International Journal of Scientific Research in Computer Science, Engineering and Information Technology
ISSN : 2456-3307 (www.ijsrcseit.com)
doi : https://doi.org/10.32628/CSEIT228628

Comparative Analysis of Ensemble Learning Methods for Enhancing Fetal Health Prediction Using Cardiotocography

Meena Chidambaram, Jens Joy
School of Computing & Information Technology, REVA University, Bengaluru, Karnataka, India

Article Info

Publication Issue :
Volume 8, Issue 6
November-December-2022

Page Number : 217-224

Article History
Accepted: 09 Nov 2022
Published: 22 Nov 2022

ABSTRACT

Nearly half of the world's stillbirths occur during labour and delivery. Early detection of any fetal distress can prompt the doctors to take appropriate measures. Cardiotocography (CTG) is one such technique that continuously records the fetal heart rate and uterine contractions during childbirth. Along with indicating signs of fetal hypoxia, CTG can also be interpreted to detect fetal abnormalities. Using the cardiotocography dataset from the UCI Machine Learning Repository, our paper displays a comparative analysis of different classifiers and ensemble learning methods such as max voting, weighted average, blending, bagging and boosting to enhance the fetal state prediction. Of all the ensemble methods used in our analysis, it was found that the Light Gradient Boosting Machine (LightGBM) gave the highest accuracy of 95.90%, which exceeded similar existing models. This increase in accuracy can prove to be potentially life saving, aid doctors in a more accurate detection of fetal abnormalities, minimize human error and decrease infant mortality rates.

Keywords : Cardiotocography, Ensemble Learning, Light Gradient Boosting, Perinatology, Fetal Health

I. INTRODUCTION

The moments leading up to the birth of a child can be equally overwhelming, chaotic and anxiety inducing for doctors as well as mothers. Doctors must keep an eye on every possible parameter indicating the fetal state. Perinatology is a sub-specialization in obstetrics which focuses on the care of a fetus experiencing distress during and/or immediately after childbirth. Cardiotocography (CTG), as the name suggests [1] ('Cardio-' - indicating heart rate, '-toco-' - indicating uterine contractions, and '-graphy' - indicating graphical output) is a graphical recording tool that measure fetal heart rates and uterine contractions.

Ideally, perinatologists and related medical examiners would analyze the CTG test recordings manually and determine appropriate further actions required, if any. As the rate of machine errors is less than that of human errors, there may be certain circumstances where machine learning would have great use in this area of healthcare so as to point out valuable information a doctor may have otherwise missed at such a critical time [2].

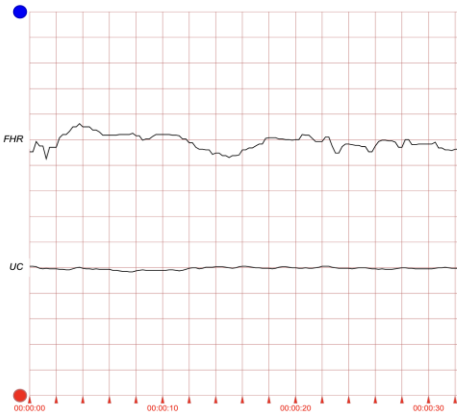

Figure 1. A sample cardiotocogram depicting the fetal heart rate and uterine contractions .

Figure 1 is an example of a cardiotocogram from the publicly available Physionet/CTU-CHB Intrapartum Cardiotocography Database [3]. A sample was chosen from the database and the waveform was displayed using Physionet's built-in waveform visualization tool, *Lightwave*. It shows the continuous recording of the fetal heart rates (FHR) and uterine contractions (UC) necessary for categorization of fetal health states in a given period of time.

This paper aims to improve the accuracy of existing decision support systems for predicting the status of fetal health [4]. Initially, we used basic classification algorithms to find out the highest possible accuracy without any means of optimization and set this as our base accuracy. Following which, we analyzed the dataset with 11 basic and advanced ensemble learning methods so as to increase and optimize our base accuracy value. Through this analysis, we were also able to show the optimization power of different ensemble learning methods [5].

Followed by this introduction we have included sequential sections consisting of the description of the dataset, merits and demerits in comparison to related studies and a process-oriented methodology of our analysis.

The final sections of this paper include the results of our analysis, followed by conclusion statements, future enhancements and references.

II. DATA DESCRIPTION

The cardiotocographic data used in this analysis has been acquired from the UCI Machine Learning Repository [6]. It was donated in 2010, by medical faculty members of the Biomedical Engineering Institute and University of Porto, Portugal.

The dataset contains 2126 samples of fetal CTG readings, automatically processed from raw cardiotocograms, which were further classified by expert obstetricians using the respective diagnostic features. It has 21 features (all with numerical values) including fetal heart rate signals, uterine contractions and other relevant measurements. The samples are classified into 3 fetal health state categories namely Normal=1, Suspected=2 and Pathological=3.

Table I shows the 4 different assessments generally used as diagnostic parameters for the classification label assigned [7]. Based on the diagnostic assessment results from Table 1, a fetus is classified as Normal, Suspicious or Pathological by using the criteria listed in **Table II [7].**

As for data preprocessing, the data was automatically processed from the source due to which, there are no missing, noisy or null values in the dataset.

TABLE I

DIAGNOSTIC ASSESSMENTS USED FOR CATEGORIZATION OF FETAL HEALTH STATES

Category	Baseline (bpm)	Range (bpm)	Decelerations	Accelerations
Normal	110-160	≥ 5	none[1]	- present - sporadic[2]
Suspicious	100-109 161-180	< 5 ≥ 40 minutes> 25	- early/variable decelerations - individual prolonged decelerations up to 3 minutes	- present - periodical occurrence (with every contraction)
Pathological	< 100 > 180 sinusoidal[3]	< 5 > 90 minutes	- atypical variable decelerations - late decelerations - isolated prolonged decelerations > 3 minutes	- absent > 40 minutes (significance still unclear, evaluation questionable)

[1] FHR deceleration amplitude ≥ 15 bpm, duration ≥ 15 seconds

[2] FHR acceleration amplitude ≥ 15 bpm, duration ≥ 15 seconds

[3] sinusoidal FHR: ≥ 10 bpm, duration ≥ 10 minutes

TABLE II

CRITERIA FOR CATEGORIZATION OF FETAL HEALTH STATES BASED ON DIAGNOSTIC RESULTS

Category	Definition	Type of Action Required
Normal	All four assessment criteria are normal	No action required
Suspicious	At least one assessment criterion is suspicious and all others are normal	Need for conservative action
Pathological	At least one assessment criterion is pathological* or two or more are suspicious	Need for conservative and invasive action

Figure 2 shows that, among the 2126 samples, 1655 samples belong to Class-1, 295 samples belong to Class-2 and 176 belong to Class-3.

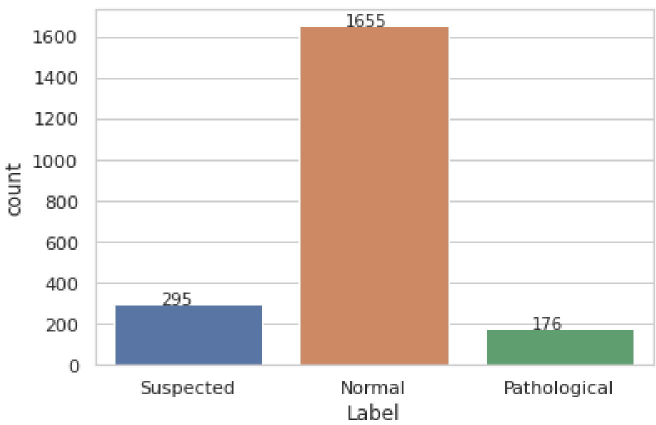

Figure 2. Barplot depicting the number of samples in each fetal health state category.

III. RELATED WORKS

In his study [8], Jassem Alhaj Tamer trained different machine learning classification algorithms such as C5.0 Decision Tree, K-Neighbors, Random Forest, Support Vector Machine and Naive Bayes on the same dataset. Among these, C5.0 Decision Tree gave the highest accuracy of 91.76%. He also trained a deep learning multilayer perceptron model which performed with an accuracy of 84%.

It is useful to note that using deep learning methods on small datasets may result in a high prediction variance. On the contrary, using ensemble methods can decrease prediction variance. Our paper has outperformed these accuracies significantly by using combinations of different traditional machine learning algorithms and ensemble methods exclusively.

In another similar study [9], Afridi et. al used correlation based feature selection to eliminate unnecessary features and trained the resulting data on 6 different classification algorithms. Among these, the Naive Bayes algorithm gave the highest accuracy of 83.06%. Through their study, we can conclude that feature selection may not give the best results all the time. Hence, our paper achieves a higher accuracy using all the features.

A marginal increase in accuracy can have a strong positive impact on classification predictions and function approximations when dealing with medical records. This ensures a better notion of reliability. Islam et al [10], conducted a similar analysis by training 3 classification algorithms, out of which Random Forest Classifier gave the highest accuracy of 95.11%. Hence, by using ensemble learning techniques our paper displays an enhanced accuracy of 95.90%.

IV. METHODOLOGY

Figure 3 represents the technical flow of our analysis. As discussed earlier, the dataset has been obtained from the UCI Machine Learning Repository and with regards to data pre-processing, the data was automatically processed by the source and did not contain any missing, noisy or null values.

Initially, we shuffled and batched the data by using either train-test-split (30% test size) or cross validation (10-folds) methods for final validation and evaluation. These batches were then fed into five individual base classification algorithms namely, K-Neighbors, Decision Tree, Support Vector Machine, Gaussian Naive Bayes and Linear Discriminant Analysis. Following evaluation, these models were compared to find the highest accuracy among them. This was set as the maximum base accuracy.

The next steps involved performing the same analysis using the different ensemble learning methods. Post evaluation, the accuracy of each ensemble learning model was compared with the previously obtained maximum base accuracy. Ensemble learning models resulting in a higher accuracy were kept as positive results supporting the fact that ensemble models can enhance the base classifiers accuracy, while the others were discarded.

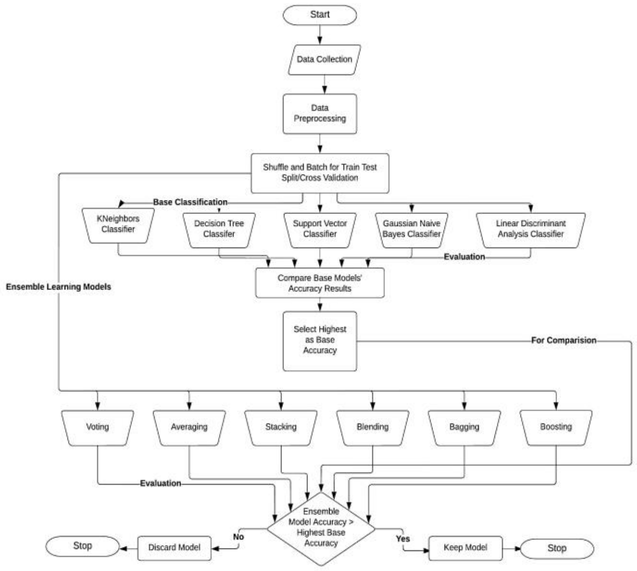

Figure 3. Flowchart depicts filtering of ensemble learning methods that enhance maximum base accuracy

V. RESULTS

Ensemble learning is the process of combining multiple machine learning methods to optimize and enhance prediction accuracy. The following results of our analysis support this fact. **Table III** represents comparison of the performances between base classifiers. Among these, the decision tree classifier is the model with the maximum base accuracy of 92.63%, while the Gaussian Naive Bayes classifier has the lowest accuracy of 79.00%

Table IV represents the performance of basic ensemble learning techniques such as voting and averaging. **Tables V and VI** represent performances of advanced ensemble learning techniques including stacking, blending, bagging and boosting.

After evaluation, as highlighted in bold, we noticed that averaging, max voting, bagging meta-estimator, random forest, gradient boosting machine and light gradient boosting machine methods significantly optimize the base accuracy of the decision tree classifier. Among these, the light gradient boosting Machine outperforms the other models with an accuracy of 95.90% as shown in **Table VII**. Gradient boosting comes at a close second with an accuracy of 95.60%.

TABLE III

PERFORMANCE ANALYSIS OF BASE CLASSIFIERS

Classifier	Accuracy
Gaussian Naive Bayes Classifier	79.00%
Linear Discriminant Analysis Classifier	87.00%
Support Vector Classifier	88.00%
K Neighbors Classifier (4 Neighbors)	89.00%
Decision Tree Classifier	**92.63%**

TABLE IV

PERFORMANCE ANALYSIS OF AVERAGING AND VOTING ENSEMBLE METHODS

Ensemble Technique	Validation Method	Classifiers Used	Accuracy
Weighted Average	Train Test Split Test Size = 30%	K Neighbors Classifier (4 Neighbors) Decision Tree Classifier Random Forest Classifier	92.47%
Averaging	**Train Test Split Test Size = 30%**	**K Neighbors Classifier (4 Neighbors) Decision Tree Classifier Random Forest Classifier**	**92.78%**
Max Voting	**Cross Validation 10 folds**	**K Neighbors Classifier (4 Neighbors) Decision Tree Classifier Random Forest Classifier**	**93.70%**

TABLE V

PERFORMANCE ANALYSIS OF STACKING AND BLENDING ENSEMBLE METHODS

Ensemble Technique	Validation Method	Classifiers Used	Accuracy
Stacking	Cross Validation 10 folds	**Level 0 Classifiers:** - K Neighbors Classifier (4 Neighbors) - Decision Tree Classifier -Random Forest Classifier **Level 1 Classifier**: Decision Tree Classifier	91.20%
Blending	Train Test Split Test Size = 0.3	**Base Models:** - K Neighbors Classifier (4 Neighbors) -Decision Tree Classifier - Random Forest Classifier **Blender:** Decision Tree Classifier	92.16%

Machine			

TABLE VI

PERFORMANCE ANALYSIS OF BAGGING AND BOOSTING ENSEMBLE METHODS

Ensemble Technique	Validation Method	Specifications	Accuracy
AdaBoost	Cross Validation 10 folds	Learning Rate = 0.1 N_estimators = 300	90.5%
Bagging Meta-Estimator	Cross validation 10 folds	N_estimators = 300	94.3%
Random Forest	Cross validation 10 folds	N_estimators = 300	94.6%
Gradient Boosting Machine	Cross validation 10 folds	Learning Rate = 0.1 N_estimators = 300	95.6%
Light Gradient Boosting	Cross validation 10 folds	Learning Rate = 0.1 N_estimators = 300	95.9%

TABLE VII

PERFORMANCE ANALYSIS OF ENSEMBLE METHODS WITH HIGHEST PERFORMANCE RATES

Ensemble Method	Accuracy
Averaging	92.78%
Max Voting	93.70%
Bagging Meta-Estimator	94.30%
Random Forest	94.60%
Gradient Boosting Method	95.60%
Light Gradient Boosting Method	**95.90%**

VI. CONCLUSION & FUTURE SCOPE

Not all birth defects can be found prenatally, before the baby is born. But tools and techniques such as cardiotocography and high resolution ultrasounds are huge steps toward early detection and diagnosis.

These also aid medical examiners to take timely actions depending on the state of fetal health.

As mentioned in **Table II**, signs and symptoms resulting in a suspected state fetus require conservative actions while those of a pathological state require invasive actions. While manually analyzing CTG readings, the line that differentiates these two states can be a blur. An invasive procedure may be performed on a fetus that does not require it or vice versa. This is where machine learning plays a role by accurately differentiating between the individual states, thereby reducing unintentional errors and misinterpretation.

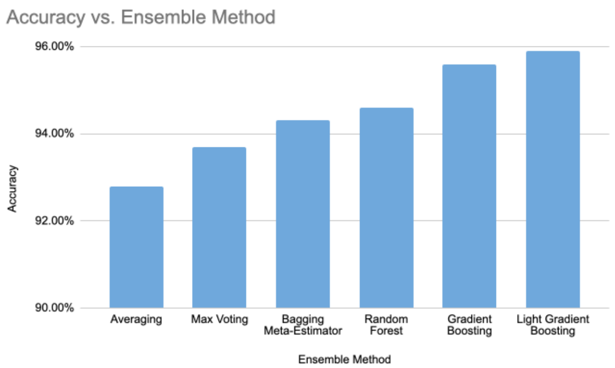

Figure 4. Comparison of top performing ensemble methods.

After analyzing all the results, we concluded that the light gradient boosting machine gave the highest accuracy, which implied that the model was 95.90% **(Figure 4)** accurate while performing predictions on new or unseen data samples.

While testing multiple models, it was apparent that by acquiring a larger dataset of similar attributes, the samples could be trained on deep learning models in the future. This would potentially increase the accuracy rates and perform more efficiently with lesser human intervention. The readings obtained from a CTG could also be used to detect or diagnose other fetal abnormalities such as hypoxia. Further research could be conducted to find other uses of CTG readings like APGAR Scores [11], thereby conducting analysis on these as well.

VII. REFERENCES

[1]. "Meaning of Cardio-toco-graphy", Patient Resources, Dr. Emeil Kamel. Link: https://www.dremeilkamel.com.au/patient-resources/obstetrics/cardiotocography/

[2]. Whittingham, R. (2003). The Blame Machine: Why Human Error Causes Accidents (1st ed.). Routledge. https://doi.org/10.4324/9780080472126

[3]. Chudácek ˇ et al.: Open access intrapartum CTG database. BMC Pregnancy and Childbirth 2014 14:16. https://doi.org/10.1186/147123931416 [Online] Available: https://physionet.org/content/ctu-uhb-ctgdb/1.0.0/

[4]. Pankaj Bhowmik, Pulak Chandra Bhowmik, U. A. Md. Ehsan Ali, Md. Sohrawordi, "Cardiotocography Data Analysis to Predict Fetal Health Risks with Tree-Based Ensemble Learning", International Journal of Information Technology and Computer Science(IJITCS), Vol.13, No.5, pp.30-40, 2021. DOI: 10.5815/ijitcs.2021.05.03

[5]. M. Manikandan & P. Vijayakumar (2021): Improving the Performance of Classifiers by Ensemble Techniques for the Premature Finding of Unusual Birth Outcomes fromCardiotocography, IETE Journal of Research, DOI: 10.1080/03772063.2021.1910579

[6]. Dua, D. and Graff, C., "UCI machine learning repository," [Online]. Available: archive.ics.uci.edu/ml/datasets/Cardiotocography

[7]. German Society of Gynecology and Obstetrics (DGGG); Maternal Fetal Medicine Study Group (AGMFM); German Society of Prenatal Medicine and Obstetrics (DGPGM); German Society of Perinatal Medicine (DGPM). S1-Guideline on the

Use of CTG During Pregnancy and Labor: Long version - AWMF Registry No. 015/036. Geburtshilfe Frauenheilkd. 2014 Aug;74(8):721-732. doi: 10.1055/s-0034-1382874. PMID: 27065483; PMCID: PMC4812878.

[8]. Tamer, Jassem Alhaj (2020) “Abnormal Foetuses Classification Based on Cardiotocography Recordings Using Machine Learning and Deep Learning Algorithms.” Masters thesis, Dublin, National College of Ireland. Link: https://norma.ncirl.ie/4360/

[9]. Afridi, R., Iqbal, Z., Khan, M., Ahmad, A., & Naseem, R. (2019). Fetal Heart Rate Classification and Comparative Analysis Using Cardiotocography Data and Known Classifiers. International Journal of Grid and Distributed Computing (IJGDC), ISSN: 2005-4262 (Print); 2207-6379 (Online), NADIA, 12(1), 31-42. doi: 10.33832/ijgdc.2019.12.1.03.

[10]. Islam, S.M., & Yulita, I.N. (2020). Predicting Fetal Condition from Cardiotocography Results Using the Random Forest Method.

[11]. Ogasawara, J., Ikenoue, S., Yamamoto, H. et al. Deep neural network-based classification of cardiotocograms outperformed conventional algorithms. Sci Rep 11, 13367 (2021). https://doi.org/10.1038/s41598-021-92805-9

Cite this article as :

Meena Chidambaram, Jens Joy, "Comparative Analysis of Ensemble Learning Methods for Enhancing Fetal Health Prediction Using Cardiotocography", International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT), ISSN : 2456-3307, Volume 8 Issue 6, pp. 217-224, November-December 2022. Available at doi : https://doi.org/10.32628/CSEIT228628

Journal URL : https://ijsrcseit.com/CSEIT228628