Fetal Arc: Predicting Fetal Health, and birth-weight of fetus using Machine Learning

Simran
Indraprastha Institute of Information Technology Delhi (IIIT-Delhi)
Okhla Industrial Estate, Phase III
New Delhi, India - 110020
simran21146@iiitd.ac.in

Abstract

The moment a child is born, the mother is also born. She never existed before. The woman existed, but the mother, never. A mother is something absolutely new. The lines talk about pregnancy, which is one of the most beautiful phases in women's life. To make it nicer and easier, it is utmost important to take care of fetal health. This project focuses on machine learning techniques used for predicting Fetal Health as Normal, Suspect or Pathological using cardiotocography (CTG) data and predicting birth weight of baby using gestational age and mother's features. For birth weight prediction, Random Forest Regressor and AdaBoostRegressor are used in a weighted fashion to give final result of birth weight in kgs, with root mean squared error (rmse) being 0.42 on train set and 0.44 on test set. Going for the second problem of predicting fetal health as normal, suspect or pathological, Support Vector Classifier, Decision Tree Classifier, Adaboost Classifier, Random Forest Classifier are used and through majority voting the final label is assigned. This technique gave macro recall score of 0.95 on train set and 0.92 on test set. The report would discuss about approach for solving problem and results with analysis.

1 Introduction

Pregnancy is the most delightful period. Healthy pregnancy leads to healthy baby. So fetal care becomes utmost important. According to WHO, one million babies die within 24 hours of birth due to premature birth and complications during birth. Also, around 810 women die each day during delivery or soon after delivery. This really causes the need to take care of fetus with utmost priority.

Cardiotocography (CTG) is well-known and most widely used method to know about fetal health which records (graph) the fetal heartbeat (cardio) and uterine contractions (toco) of mother during pregnancy. It is carried out during third trimester or sometimes even during final trimester and during delivery so as to know if fetal heartbeat is not hampered by uterine contractions. For mother, the uterine contractions are measured, when in no pain, the baseline is recorded, when in pain, pain intensity, time duration of contraction and gap between two contractions from tocograph is recorded. For fetal heart, the baseline, variability (denoting happy state of baby), acceleration, deceleration are measured. Normal condition occurs when every parameter is within desired range, fetal health is suspectible when one of parameters is abnormal and pathological when more than one parameters are not normal. In case of suspectible, call is made for more tests while for pathological state, there are emergency actions taken by the doctor. The machine learning algorithms provide a quick support and equip the doctors to take actions immediately in case the fetus is in abnormal condition.

Birth weight stands most crucial for fetus, defining the risk during delivery, mortality rate within one year and somewhat related to diseases that occur in adulthood. Birth weight is difficult to measure directly but a rough estimate can be made by experienced doctors. Low birth weight can potentially causes major issues and Over weight can lead to serious injuries to mother and fetus during delivery, hence getting even a rough estimate in this direction would also serve to be of great help. Various machine learning techniques can be employed to predict fetal birth weight by exploiting the features of mother.

2 Literature Survey

This section discusses the research carried out in fetal health classification and birth weight prediction.

2.1 Fetal Health Classification

The paper (1) takes the help of CTG data and classify fetal health as physiological, suspect and pathological on the basis of fetal heart beat and uterine contractions of mother. It uses random forest method for classification and achieved accuracy of 94.69%. The paper (2) uses Association Based Classification Approach and was 84% accurate. (3) worked on Artificial Neural Network to classify fetal health and attained F-Score of 0.9784, 0.4514, 0.9724 for Normal, Suspect and Pathological class respectively. Along the same lines, (4) experimented with ANN and Logistic Regression to attain accuracy of 98.5% and 98.7% respectively. But this paper only considers two fetal states: normal and pathological.The paper (5) uses blending ensemble for classification using soft voting to attain accuracy rate of 0.959, an AUC of 0.988, a recall rate of 0.916, a precision rate of 0.959, a F1 of 0.958.

2.2 Birth Weight Prediction

In order to get birth weight (6) uses Linear-Regression, Support Vector Regression (SVR), Back Propagation Neural Network(BPNN) which yield accuracy of 0.670, 0.680, 0.705 respectively. The parameters used were: Height (cm), Uterine Height (cm), Head Circumference (cm), Abdominal Circumference of fetal (cm), Abdominal Circumference of pregnant women (cm) etc. (7) article tries out Hadlock (52.3% accuracy), GA-BP (63.1% accuracy), Random forest(60.0% accuracy), XGBoost (62.1% accuracy), LightGBM (59.4% accuracy). But this article tweaked some parameters and created an ensemble model using all the above stated algorithms and attained accuracy of 64.3%. In all the above research papers, there was continuous valued output for weight, whereas, (8) solved it as classification problem and classified birth-weight as low weight or high weight. The paper used self-created dataset containing 445 instances and 18 features. There was 86% accuracy for Gaussian Naive Bayes and 100% accuracy for Random Forest. For paper (9), the bagged tree, achieved pretty good results concerning accuracy and area under the ROC curve, which are 0.849 and 0.636, respectively. (10) also viewed problem as classification problem and had output classes as: low weight, normal weight which were identified through Artificial Neural Network with accuracy score of 100%.

3 Dataset and Analysis/ Preprocessing

Dataset for predicting Fetal Health is taken from kaggle. The link to dataset is:

`https://www.kaggle.com/andrewmvd/fetal-health-classification`

The dataset has 2126 training instances, with 21 attributes.The output label is Fetal health: 1 - Normal 2 - Suspect 3 - Pathological. The independent attributes are given in Table 1.

Attributes
baseline value
accelerations
fetal_movement
uterine_contractions
light_decelerations
severe_decelerations
prolongued_deceleration
abnormal_short_term_variability
mean_value_of_short_term_variability
percentage_of_time_with_abnormal _long_term_variability
mean_value_of_long_term_variability
histogram_width
histogram_min
histogram_max
histogram_number_of_peaks
histogram_number_of_zeroes
histogram_mode
histogram_mean
histogram_median
histogram_variance
histogram_tendency

Table 1: Attributes for fetal health classification.

"baseline value" is Baseline Fetal Heart Rate which helps in monitoring acceleration and deceleration which means that both attributes are computed with respect to baseline. "accelerations" means number of accelerations per second (speed up in fetal heart beat at least 15 beats above the baseline within duration of 15 seconds to come back, more the acceleration, happier is the baby). "fetal_movement" is having obvious meaning which is captured when mother presses button of machine when she feels movement of baby. This activity is recorded per 20 minutes. "uterine_contractions" is when there is pressure on stomach. Deceleration means decrease in fetal heart beat due to contraction and then coming back as pressure is released. Deceleration is of three types: Light, Severe or Prolongued as are the attributes: "light_decelerations", "severe_decelerations", "prolongued_deceleration" respectively. Light deceleration occurs when fetal heart beat decreases 15 beats below baseline with uterine contraction, this is done because fetus is under pressure i.e. head of baby is getting compressed, so autonomous nervous system instructs heart to decrease heart beat. As contraction comes down, the heart beat should increase along-with. Light deceleration is normal scenario. Severe deceleration occurs when heart beat increases with some delay i.e. it is not increasing as contraction moves down. Here, some fetal tests needed to be done. Prolonged deceleration is when the heartbeat does not come back to normal for more than 15 seconds. This is an alarming situation when doctors response immediately. Variability is variation in lines of tocograph. More the variations, happier is the child. Rest are the histogram parameters which assist in predicting fetal health. On performing EDA, it was discovered that some features had very high correlation (See Figure 1), so some features can be removed.

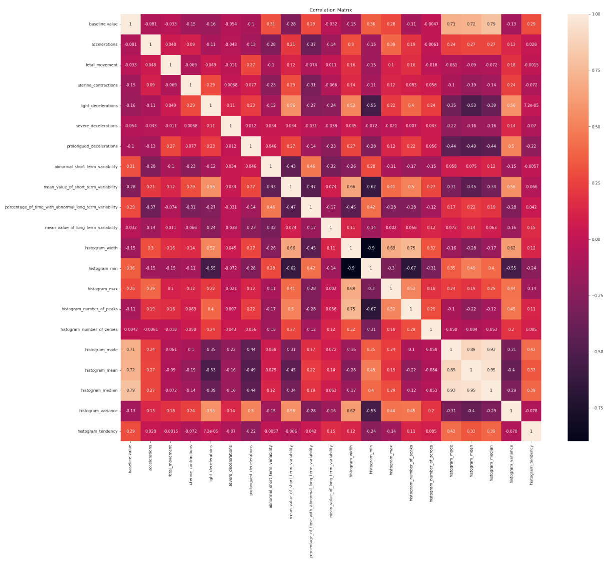

Figure 1: Correlation matrix

Now, the final set of features are given in Table 2. Almost all features are present, only histogram mode and histogram median are removed. In total there are 19 features now in the dataset.

This dataset needs to be preprocessed using Standard Scaler, as values for all the features vary a lot. Next thing to focus on is, the distribution of classes in dataset (See Figure 2), which shows that "NORMAL" class takes up the most portion and then "SUSPECT" and "PATHOLOGICAL" are really less in numbers, so there is class imbalance.

Attributes
baseline value
accelerations
fetal_movement
uterine_contractions
light_decelerations
severe_decelerations
prolongued_deceleration
abnormal_short_term_variability
mean_value_of_short_term_variability
percentage_of_time_with_abnormal _long_term_variability
mean_value_of_long_term_variability
histogram_width
histogram_min
histogram_max
histogram_number_of_peaks
histogram_number_of_zeroes
histogram_mean
histogram_variance
histogram_tendency

Table 2: Final Attributes for fetal health classification.

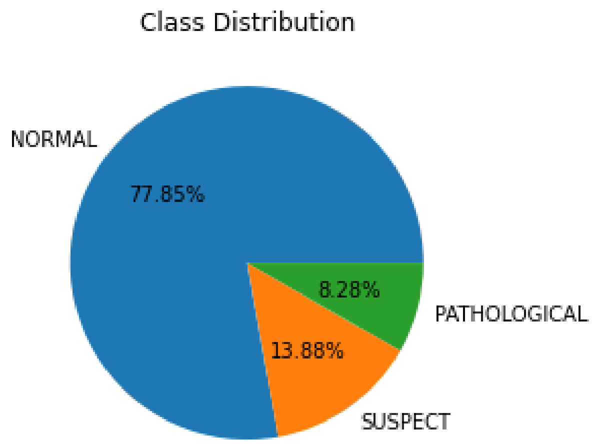

Figure 2: Class Distribution

Overall the conclusions from the dataset are:

- The no. of features are now reduced to 19.
- There is class imbalance.
- Standard Scaling is applied on dataset.

For the birth weight prediction, there are 6 attributes and 1174 instances. The dataset can be found from:

http://people.reed.edu/~jones/141/Bwt.dat

The dataset attributes shows that health of mother is most important attribute for baby's health and various features of mother are tested so as to predict the weight of baby. The dataset description is shown in table below.

Attributes	Attribute Description
bwt	baby's weight in ounces at birth
gestation	duration of pregnancy in days
parity	first born = 1, later birth = 0
age	mother's age in years
height	mother's height in inches
weight	mother's weight in pounds
smoke	mother smokes (1=yes, 0=no)

Table 3: Attributes for birth weight prediction.

Now, usually in India, kilograms is used more often, so all the parameters are changed as per that and final attributes are shown in Table 4.

Attributes	Attribute Description
bwt	baby's weight in kgs at birth
gestation	duration of pregnancy in days
parity	first born = 1, later birth = 0
age	mother's age in years
height	mother's height in inches
weight	mother's weight in kgs
smoke	mother smokes (1=yes, 0=no)

Table 4: Final Attributes for birth weight prediction.

Let's see some EDA results which can be exploited later or help us to understand data. See Figures 3-8 to get to know about data.

Note the following insights about data.

- First born child has a little lower birth weight as compared to child born after first child.
- It is quite obvious that as gestational age increases then birth weight also increases.
- Baby after 42 age and before 20 years of age is risky.

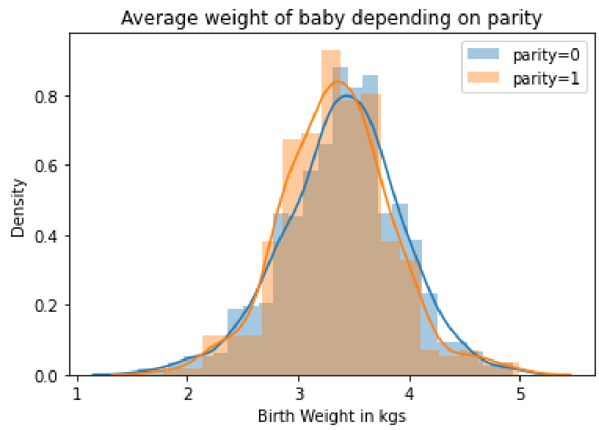

Figure 3: Exploring relation between birth weight and parity

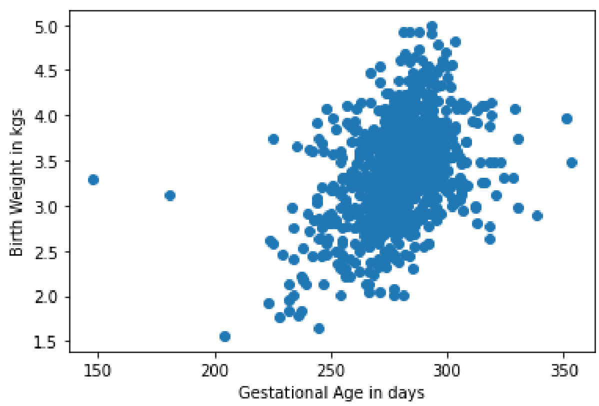

Figure 4: Exploring relation between birth weight and gestational age

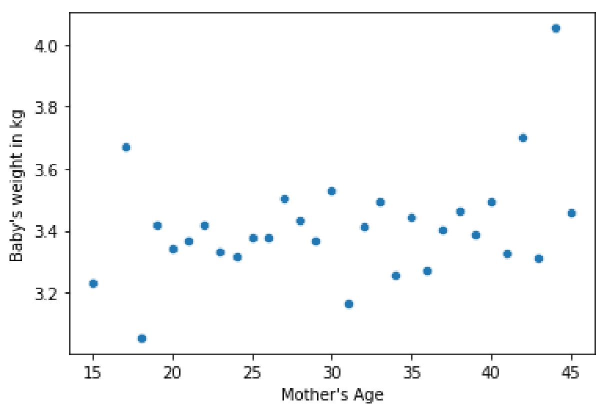

Figure 5: Exploring relation between birth weight and mother's age

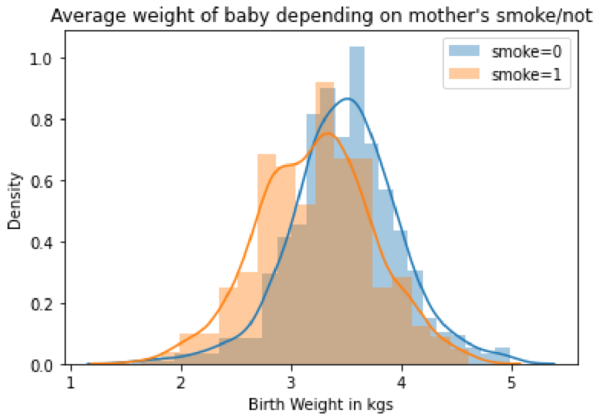

Figure 6: Exploring relation between birth weight and mother's smoking habit

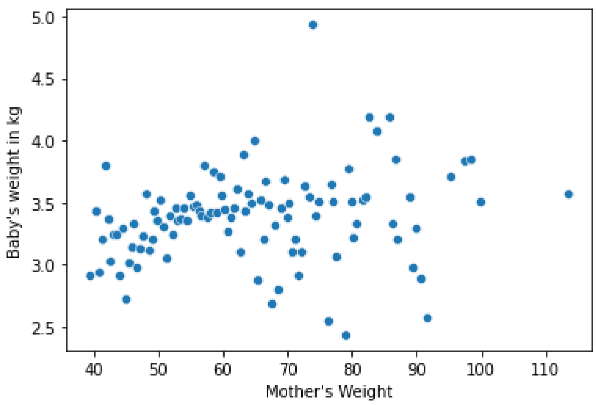

Figure 7: Exploring relation between birth weight and mother's weight

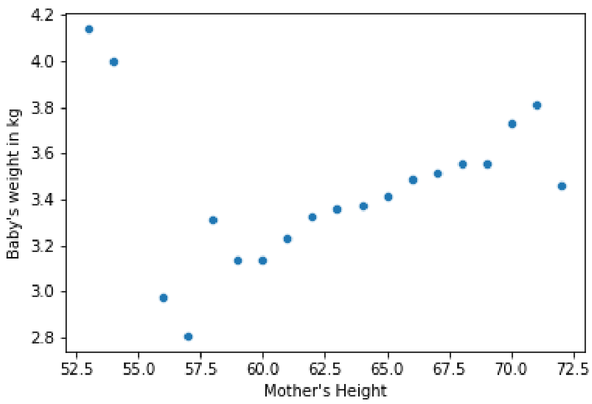

Figure 8: Exploring relation between birth weight and mother's height

- Usually smoker mothers have lower weight babies.

- Tall mothers give birth to babies with weight on little higher side.

4 Baseline Systems: Results and Analysis

For the fetal health classification, there can be potentially two baselines. One is by using mode which captures the class with highest frequency and other baseline model can be Logistic Regression. Before that it is important to decide the metric to evaluate the models. Here, "SUSPECT" and "PATHOLOGICAL" are the states which need to be focused on and they must not be missed i.e. we can't let go False Negatives, so recall seems to be good option, now for final metric, macro recall can be used keeping in mind the class imbalance. More the recall, better is model. See the evaluation metrics for both the baseline models on train and test set in Figure 9. Here, logistic regression model performs better.

```
-----------------------------------------------
M1: Mode
M2: Logistic Regression
-----------------------------------------------
            TRAINING SET RESULTS
-----------------------------------------------
       Class  Recall Score(M1)  Recall Score(M2)
      NORMAL       1.000000          0.959909
     SUSPECT       0.000000          0.816327
PATHOLOGICAL       0.000000          0.683983
MACRO RECALL       0.333333          0.820073
-----------------------------------------------
            TESTING SET RESULTS
-----------------------------------------------
       Class   Recall Score(M1)   Recall Score(M2)
      NORMAL       1.000000          0.927928
     SUSPECT       0.000000          0.689655
PATHOLOGICAL       0.000000          0.656250
MACRO RECALL       0.333333          0.757944
```

Figure 9: Evaluation for baseline models (Fetal Health Classification)

For birth weight prediction, mean and kNN seem to be good choice for baseline models. To evaluate them, root mean square error can be used and one with lesser RMSE value will be better. RMSE is not robust to outliers and that's what we want, we want that it should capture errors like that, so that overweight or underweight babies can be predicted. RMSE has same units as of original metric, so here RMSE will be in kgs. Here, KNN is better baseline model (Figure 10).

```
---------------------------
        Train Set  Test Set
---------------------------
MEAN     0.517      0.532
KNN      0.399      0.506
```

Figure 10: Evaluation (RMSE) for baseline models (Birth Weight Prediction)

5 Final Models

Fetal Health prediction is a classification problem with class imbalance. Various models can be used, but after trying various models, the final model is majority voting for SVC, Decision Tree Classifier, Random Forest Classifier and Adaboost Classifier. All the models give their predictions and final result depends on majority voting. The parameters for the base models are shown in Table 5. The base estimator in Adaboost is the Decision Tree shown in Table 5.

Base Models	**Parameters**
SVC	class_weight="balanced" OVO kernel="rbf"
RandomForest	n_estimators=65 class_weight="balanced" criterion="gini" min_samples_leaf=15
DecisionTree	max_depth=7 min_samples_split=10
Adaboost	n_estimators=40 learning_rate=1.5 base_estimator=DTC

Table 5: Base Models used for Majority Voting (Fetal Health Classification).

Birth Weight prediction is a regression problem. Various models can be used, but after trying various models, the final model is weighted model with RandomForestRegressor and AdaboostRegressor as base models, where Adaboost contributes to final result 5/6 to it's final prediction and RandomForest contributes 1/6 of it's final prediction. The parameters for the base models are shown in Table 6. Base Estimator in Adaboost Regressor is DecisionTreeRegressor(criterion="absolute_error", max_depth=3, random_state=1)

Base Models	**Parameters**
RandomForest	n_estimators=4 max_depth=4 criterion="squared_error" random_state=8
Adaboost	n_estimators=11 learning_rate=1.2 loss="exponential"

Table 6: Base Models used for Weighted Prediction (Birth Weight Estimation).

6 Result, Analysis and comparison with baseline

Following are results for both modules: fetal health classification and birth weight prediction.

6.1 Fetal Health Classification

The main task was to get better recall score, especially for classes "SUSPECT" and "PATHOLOGICAL", and it was achieved to some extent. Figure 11 shows how the final model performs on train and test set, while Figure 13 is full classification report. To have the comparison with baseline, see Figure 12. Clearly, the final models outperforms both the baselines with high macro recall score of 0.95 for training data and 0.92 for test data.

```
---------------------------------------
                TRAINING SET  TESTING SET
---------------------------------------
       Class  Recall Score  Recall Score
      NORMAL      0.931921      0.912913
     SUSPECT      0.965986      0.931034
PATHOLOGICAL      0.956710      0.921875
MACRO RECALL      0.951539      0.921941
```

Figure 11: Evaluation for final model

6.2 Birth Weight Prediction

After a lot of tries, reducing RMSE to lower value was a painstaking task. After all those efforts, the results are shown below (Figure 14). Well, overall it performed better than both baseline models. It's test RMSE is least which is 0.441, while it is a little higher for other two. It may seem like not much difference from baseline, but it was really hard to achieve this gap as well and here every decimal point matters, as there is short range of possible weights.

```
---------------------------------------------
M1: Mode (Baseline 1)
M2: Logistic Regression (Baseline 2)
M3: Final Model
---------------------------------------------
             TRAINING SET RESULTS
---------------------------------------------
        Class Recall(M1)  Recall(M2)  Recall(M3)
       NORMAL 1.000000    0.959909    0.931921
      SUSPECT 0.000000    0.816327    0.965986
 PATHOLOGICAL 0.000000    0.683983    0.956710
 MACRO RECALL 0.333333    0.820073    0.951539
---------------------------------------------
             TESTING SET RESULTS
---------------------------------------------
        Class Recall(M1)  Recall(M2)  Recall(M3)
       NORMAL 1.000000    0.927928    0.912913
      SUSPECT 0.000000    0.689655    0.931034
 PATHOLOGICAL 0.000000    0.656250    0.921875
 MACRO RECALL 0.333333    0.757944    0.921941
```

Figure 12: Comparison of final model with baseline

```
TRAIN CLASSIFICATION REPORT
--------------------------------------------------
              precision    recall  f1-score   support

      NORMAL       0.99      0.93      0.96      1322
PATHOLOGICAL       0.96      0.97      0.96       147
     SUSPECT       0.72      0.96      0.82       231

    accuracy                           0.94      1700
   macro avg       0.89      0.95      0.91      1700
weighted avg       0.95      0.94      0.94      1700

TEST CLASSIFICATION REPORT
--------------------------------------------------
              precision    recall  f1-score   support

      NORMAL       0.99      0.91      0.95       333
PATHOLOGICAL       0.84      0.93      0.89        29
     SUSPECT       0.69      0.92      0.79        64

    accuracy                           0.92       426
   macro avg       0.84      0.92      0.87       426
weighted avg       0.93      0.92      0.92       426
```

Figure 13: Classification report for final model

```
----------------------------------
                  Train Set  Test Set
----------------------------------
MEAN (Baseline 1)  0.517     0.532
KNN  (Baseline 2)  0.399     0.506
Final Model        0.421     0.441
```

Figure 14: Evaluation (RMSE) for final model and comparison with baseline

7 Conclusion

By all counts, this project can be one stop solution to fetal care. Fetal Arc, it says, where arc means from point of good news till delivery, taking care of your baby. The machine learning techniques have got really good results and in few cases are better than previous work. The birth weight prediction module is better than previous work, so it can be deployed end to end, but for fetal health classification though the recall score is good but still, in such cases having high recall score is still not safe, afterall it is cost of life of both baby and mother, so more work needs to be done in this field so as to make a better system to predict fetal health as normal, suspect or pathological. This paper can provide direction to research in this area.

8 Contribution of each group member

The complete project is done single handedly by the author but she would like to thank god, family, professors, TAs and everyone who helped her to make this project possible.

9 Links

Link for code and dataset:
https://drive.google.com/file/d/1AAjE9xKJTebOV6Rgg90GMe3DTpP8EEcL/view?usp=sharing

Link for deployed website:
https://fetalhealth-simran.herokuapp.com/

References

[1] Peterek, J. (2014). Human Fetus Health Classification on Cardiotocographic Data Using Random Forests. In Intelligent Data analysis and its Applications, Volume II (pp. 189-198). Springer International Publishing.

[2] Piri, J., Mohapatra, P. (2019). Exploring Fetal Health Status Using an Association Based Classification Approach. In 2019 International Conference on Information Technology (ICIT) (pp. 166-171).

[3] Chinnasamy, S., Chitradevi, M., Geetharamani, G. (2012). Classification of Cardiotocogram Data using Neural Network based Machine Learning Technique. International Journal of Computer Applications, 47, 19-25.

[4] Şahin, Hakan Subasi, Abdulhamit. (2012). Classification of Fetal State from the Cardiotocogram Recordings using ANN and Simple Logistic.

[5] Li, J., Liu, X. (2021). Fetal Health Classification Based on Machine Learning. In 2021 IEEE 2nd International Conference on Big Data, Artificial Intelligence and Internet of Things Engineering (ICBAIE) (pp. 899-902).

[6] Tao, J., Yuan, Z., Sun, L. et al. Fetal birthweight prediction with measured data by a temporal machine learning method. BMC Med Inform Decis Mak 21, 26 (2021). https://doi.org/10.1186/s12911-021-01388-y

[7] Lu, Y., Zhang, X., Fu, X., Chen, F., Wong, K. (2019). Ensemble Machine Learning for Estimating Fetal Weight at Varying Gestational Age. Proceedings of the AAAI Conference on Artificial Intelligence, 33(01), 9522-9527.

[8] Zakir Hussain, Malaya Dutta Borah (2020). Birth weight prediction of new born baby with application of machine learning techniques on features of mother. Journal of Statistics and Management Systems, 23(6), 1079-1091.

[9] Moreira, M., Rodrigues, J., Furtado, V., Mavromoustakis, C., Kumar, N., Woungang, I. (2019). Fetal Birth Weight Estimation in High-Risk Pregnancies Through Machine Learning Techniques. In ICC 2019 - 2019 IEEE International Conference on Communications (ICC) (pp. 1-6).

[10] Al-Shawwa, Mohammed Abu-Naser, Samy Nasser, Ibrahim. (2019). Developing Artificial Neural Network for Predicting Mobile Phone Price Range. 3. 1-6.