Volume 9, Issue 10, October– 2024 International Journal of Innovative Science and Research Technology
ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24OCT1526

Fetal Health Classification and Birth Weight Estimation Using Machine Learning

Shreeya R Hegde[1]
CSE, JSSATE-Bangalore

Sinchana S[2]
CSE, JSSATE-Bangalore

Supriya P Nadgir[3]
CSE, JSSATE-Bangalore

Vinamratha R Jagirdar[4]
CSE, JSSATE-Bangalore

Abstract:- This paper addresses fetal health prediction via Cardiotocography (CTG) data analysis, utilizing models like Logistic Regression, SVM, and boosting algorithms. Feature selection methods such as PCA and LDA are employed, along with SMOTE for dataset balancing. CatBoost model emerges as superior with 99% accuracy. Fetal weight prediction remains challenging, tackled through machine learning algorithms incorporating parameters like gestational period and maternal factors. Models like Random Forest and Adaboost are employed, with RMSE analysis guiding their combination for improved prediction accuracy. Overall, the paper emphasizes leveraging ML for fetal health classification and birth weight prediction.

Keywords:- *Fetal Health, Cardiotocography, Logistic Regression, SVM, XGBoost, CatBoost, SMOTE Technique, Fetal Birth Weight.*

I. INTRODUCTION

Monitoring fetal growth during pregnancy presents a formidable challenge within the medical domain, characterized by its complexity and critical importance. The primary objective of fetal monitoring is to assess the health status of the unborn baby, a task heavily contingent upon the mother's own well-being.

Cardiotocography yields a plethora of invaluable data, encompassing crucial parameters such as uterine contractions, fetal heart rate, and occurrences of accelerations and decelerations. These metrics provide clinicians with profound insights into the intricate dynamics of fetal health, facilitating timely interventions when necessary. The prediction of fetal birth weight shortly before delivery enables obstetricians to make informed decisions regarding the most appropriate delivery mode for pregnant women, as birth weight stands as a pivotal determinant of neonatal outcomes and survival rates.

The application of machine learning (ML) in fetal health classification and weight estimation emerges as a promising frontier. ML algorithms, including Decision Trees, Random Forest, AdaBoost, and Support Vector Machines, exhibit remarkable capabilities in discerning intricate patterns within diverse and voluminous datasets. Their capacity to learn from historical data, adapt to new information, and furnish nuanced insights positions them as invaluable tools for addressing the complexities inherent in fetal health assessment. By harnessing the power of ML algorithms, the objective extends beyond mere enhancement of assessment accuracy to empowering healthcare professionals with tools conducive to early detection and proactive management of potential pregnancy complications.

II. LITERATURE SURVEY

This article employs the LightGBM model for fetal health classification, boasting high predicted accuracy. However, limitations include its complex nature, making interpretation challenging, and substantial computational resource demands, potentially hindering its use in resource-limited settings, despite its potential to enhance fetal health assessment and management[1]. The paper proposes a predictive method based on ensemble learning for the classification of fetal health using a cardiotocography dataset obtained from nonstress tests (NST) and use of XGBoost method. It also discusses the decrease in accuracy for pathological data due to downsampling of unbalanced dataset[2]. The paper discusses the use ML algorithms such as SVM, random forests, and deep learning for fetal health classification. The limitations addressed in the paper include the need for large amounts of labeled data for training machine learning models, the potential for overfitting, and the interpretability of complex models. The paper also highlights the challenges of selecting appropriate features and optimizing hyperparameters in machine learning models[3]. Models like deep learning neural networks, k-nearest neighbor, and blended ones achieved an average accuracy of 99%. However, lacking explainability raised concerns. This paper proposes a high-performing classifier with added explainability. Trained on 2126 data points, its reliability in real-world scenarios may be affected despite its high classification rate[4]. ML has been widely applied in various subjects related to maternal and fetal health. The most common use of ML is the prediction of perinatal disorders. Other ML applications include biomarker discovery, risk estimation, and drug screening. The review was not performed using a systematic protocol, which may introduce bias. The review was limited to seven selected pregnancy diseases and complications [5]. The paper provides

Volume 9, Issue 10, October– 2024 International Journal of Innovative Science and Research Technology
ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24OCT1526

a comprehensive review of deep-learning algorithm like CNN for fetal ultrasound-image analysis. The limitation is the lack of quantitative comparison among different approaches due to the use of different datasets, some of which are small in size[6]. This research covers the findings and analyses of multiple machine learning models for fetal health classification. For classification, random forest (RF), logistic regression, decision tree (DT), support vector classifier, voting classifier, and K-nearest neighbor were utilized. When the results are compared, it is discovered that the random forest model produces the best results. It achieves 97.51% accuracy, which is better than the previous method reported[7]. A study comparing clinical and ultrasonographic fetal weight estimations found Hadlock's equation estimated mean birth weight at 2.90 kg, Dare's at 3.07 kg, close to the actual average of 3.01 kg. Ultrasound tended to overestimate lower and underestimate higher birth weights. No significant difference was found between ultrasound and clinical estimates[8].

III. METHODOLOGY AND IMPLEMENTATION

A. Fetal Health Classification:

➢ *Dataset description:*

Each row in the dataset represents a fetal cardiotocogram examination, which is a non-invasive method used to monitor fetal health during pregnancy. The attributes include important measurements such as baseline fetal heart rate, accelerations, decelerations, and various indices derived from the CTG signals. Additionally, the dataset includes the corresponding fetal health classification, with three classes indicating the presence of normal, suspicious, or pathological fetal conditions.

CTG is a highly imbalanced dataset, as the total of 2126 samples includes 1655 normal, 295 suspect and 176 pathological ents.

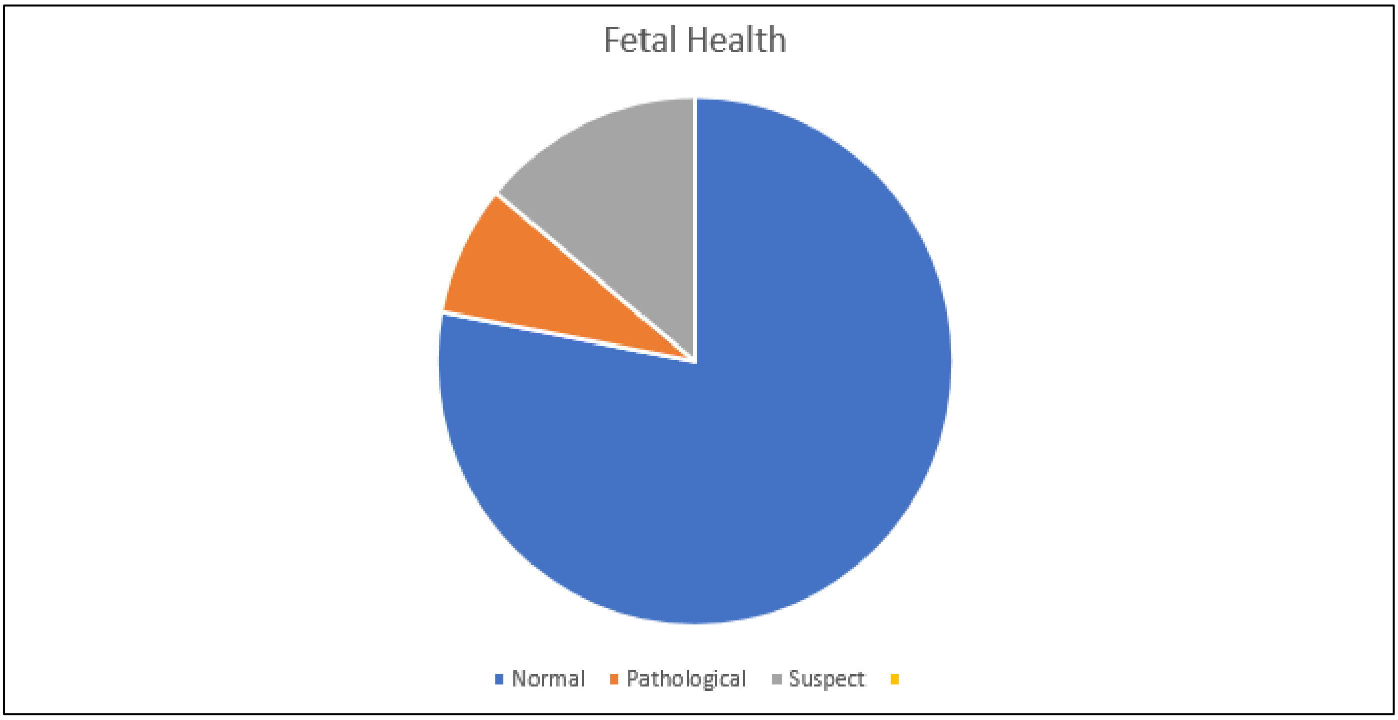

Fig 1 Fetal Health Classification Pie Chart

➢ *Exploratory Data Analysis:*

- **Data Preprocessing:** Data was checked to ensure no missing values were present, and values were standardised.
- **Data Visualisation:** Boxen plots were created for every attribute to split along fetal health classification target, analyse their relationships according to fetal health.

Through Data Visualisation it was evident that there were more decelerations among suspect and pathological records.

➢ *Outlier Detection:*

Inter Quartile Range Capping method is used to remove outliers. Interquartile range is calculated, then upper and lower limits are defined. Outliers beyond the limits are trimmed and replaced with upper and lower limit values.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24OCT1526

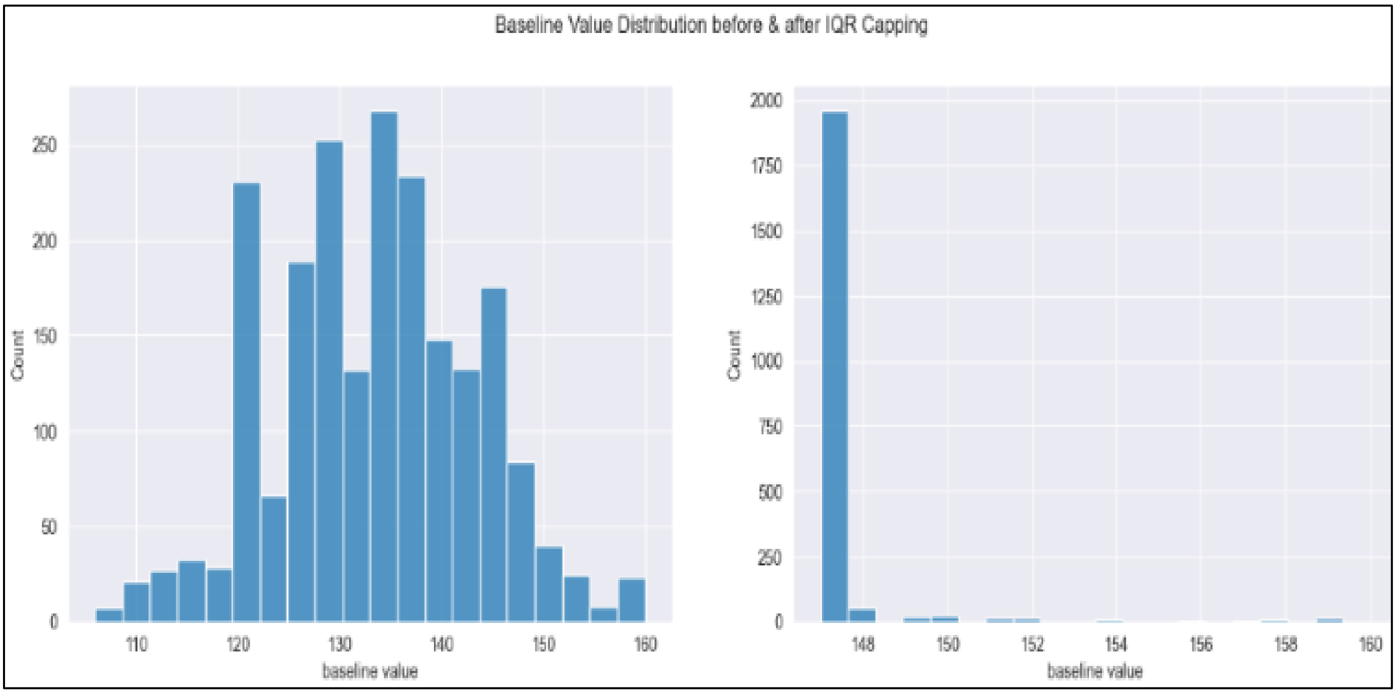

Fig 2 Baseline Value Distribution Before and After IQR Capping

➢ *Feature Selection:*

PCA and LDA are demenionality reduction techniques here:

Principle Component Analysis (PCA) is used to create a dataframe containing principal components. Two subplots are plotted, one showing the percentage of explained variance by each principal component, and the other showing the cumulative percentage of explained variance.

Unlike PCA, which is unsupervised and focuses on maximizing variance, LDA is supervised and aims to find the linear combinations of features that best separate the classes in the dataset. LDA seeks to find the linear combinations of features that best separate the classes in the data, rather than maximizing the variance as in PCA. The features selected by PCA and LDA are the principal components and linear discriminants, respectively. The number of components selected for each technique (**n_components**) determines the dimensionality of the reduced feature space.

➢ *Data Resampling:*

As the number of normal records are more compared to suspected and pathological, this leads to class imbalance and hence decreases the model performance. to address this class imbalance issue we have implemented SMOTE data augmentation technique. SMOTE (Synthetic Minority Over-sampling Technique) is a technique used to address class imbalance in classification tasks.

SMOTE works by generating synthetic samples for the minority class based on the existing minority samples. It creates synthetic samples by interpolating between existing minority samples in the feature space.

➢ *Data Modelling:*

We have used several models like SVM, Random Forest, Decision Trees Logistic Regression, Boosting Algorithms like Adaboost, XGBoost, & CatBoost.

- *Support Vector Machine:*

Support Vector Machine (SVM) is a powerful supervised learning algorithm used for classification and regression tasks. SVM aims to find the optimal hyperplane that separates different classes in the feature space with the maximum margin. In classification, this hyperplane is used to classify new data points into one of the classesParameters like the choice of kernel (e.g., radial basis function, sigmoid), regularization parameter (C), and kernel-specific parameters are optimized to improve model performance to a value of 0.91.

- *K Nearest Neighbors :*

KNN is employed for classification tasks, where the goal is to predict the class or category of a given sample based on the class labels of its neighboring samples in the feature space. The hyperparameters of the KNN algorithm, including the number of neighbors **(n neighbors)**, the weight function used in prediction **(weights),** and the algorithm used to compute nearest neighbors **(algorithm)**, are tuned using grid search. This process helps identify the optimal combination of hyperparameters that maximizes the model's performance with accuracy value of 0.91.

- *Random Forest:*

Random Forest is an ensemble learning method that constructs a multitude of decision trees during training and outputs the mode of the classes or the average prediction of the individual trees. Random Forest Classifier within a stratified K -Fold cross-validation framework to evaluate the model's performance is employed.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24OCT1526

- *Logistic Regression:*

Logistic Regression is a statistical method used for binary classification problems. Grid search cross-validation is performed to search for the best combination of hyperparameters from the provided parameter grid. It provides accuracy value of 0.90.

- *AdaBoost:*

AdaBoost (Adaptive Boosting) is a powerful ensemble learning algorithm used mainly for classification tasks. It works by combining multiple weak learners (often decision trees) to create a strong classifier. Here, AdaBoostClassifier using Stratified K-Fold Cross-Validation is implemented and evaluated for its performance.

- *XGBoost :*

XGBoost belongs to the family of gradient boosting algorithms, which are ensemble learning methods that build a strong predictive model by combining multiple weak models sequentially. Here, hyperparameters are optimised using Optuna and correspondingly to find the best hyperparameters for the XGBoost classifier, considering the F1 score as the optimization metric. It provides 0.97 as accuracy metric after data resampling with SMOTE.

- *CatBoost :*

CatBoost is based on the gradient boosting algorithm, which is an ensemble learning technique that builds a strong predictive model by combining multiple weak models (typically decision trees). It optimizes a loss function using gradient descent at each iteration to improve the model's performance. Here CatBoost model provides highest accuracy of 0.99 with SMOTE augmentation.

➢ *Model Evaluation:*

Model evaluation is performed using various metrics such as F1-score, precision, recall, and accuracy. After fitting the model on the training data and making predictions on the test data, these metrics are calculated to assess the performance of the model. The evaluation process is carried out by using Confusion Matrix. different performance metrics used here are:

B. Fetal Birth Weight Estimation:

➢ *Dataset Description:*

Dataset is taken from Stat Labs by Nolan and Speed, originally from the Child Health and Development Studies conducted at the Oakland, CA, Kaiser Foundation Hospital.

➢ *The Variables in the Dataset are*

- Bwt: baby's weight in ounces at birth .
- Gestation: duration of pregnancy in days
- Parity: parity indicator (first born = 1, later birth = 0)
- Age: mother's age in years
- Height: mother's height in inches
- Weight: mother's weight in pounds (during pregnancy)
- Smoke: indicator for whether mother smokes (1=yes, 0=no)

Here "bwt" is the target variable where we predict the birth weight for the foetus beforehand using other parameters. Then data preprocessing is done and checked for any null values which are absent in this case.

➢ *Data Models:*

Various models like mean regressor, kNN, Linear Regression, SVR , Decision tree Regressor, Ridge regression , Random Forest Regressor, Adaboost regressor are used. These models are evalusted using various hyperparameters. Training and testing RMSE values are evaluated and printed.

Root Mean Squared Error (RMSE) is a widely used metric for evaluating the performance of regression models. It measures the average magnitude of the errors between the predicted values and the actual values in a regression problem. RMSE is calculated by taking the square root of the mean of the squared differences between the predicted and actual values.

RMSE is useful because it provides a measure of the spread of the errors in the same units as the target variable, making it interpretable and easy to compare across different models and datasets. Lower values of RMSE indicate better model performance, as they represent smaller average errors between predicted and actual values.

We have built a final model by combining Randon Forest and Adaboost Regressor Models. Prediction of both models are combined using weighted averaging.

RMSE value for this final model is also printed and this final model is used for fetal birth weight prediction.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24OCT1526

* *Flowchart for Fetal Birth Weight Prediction:*

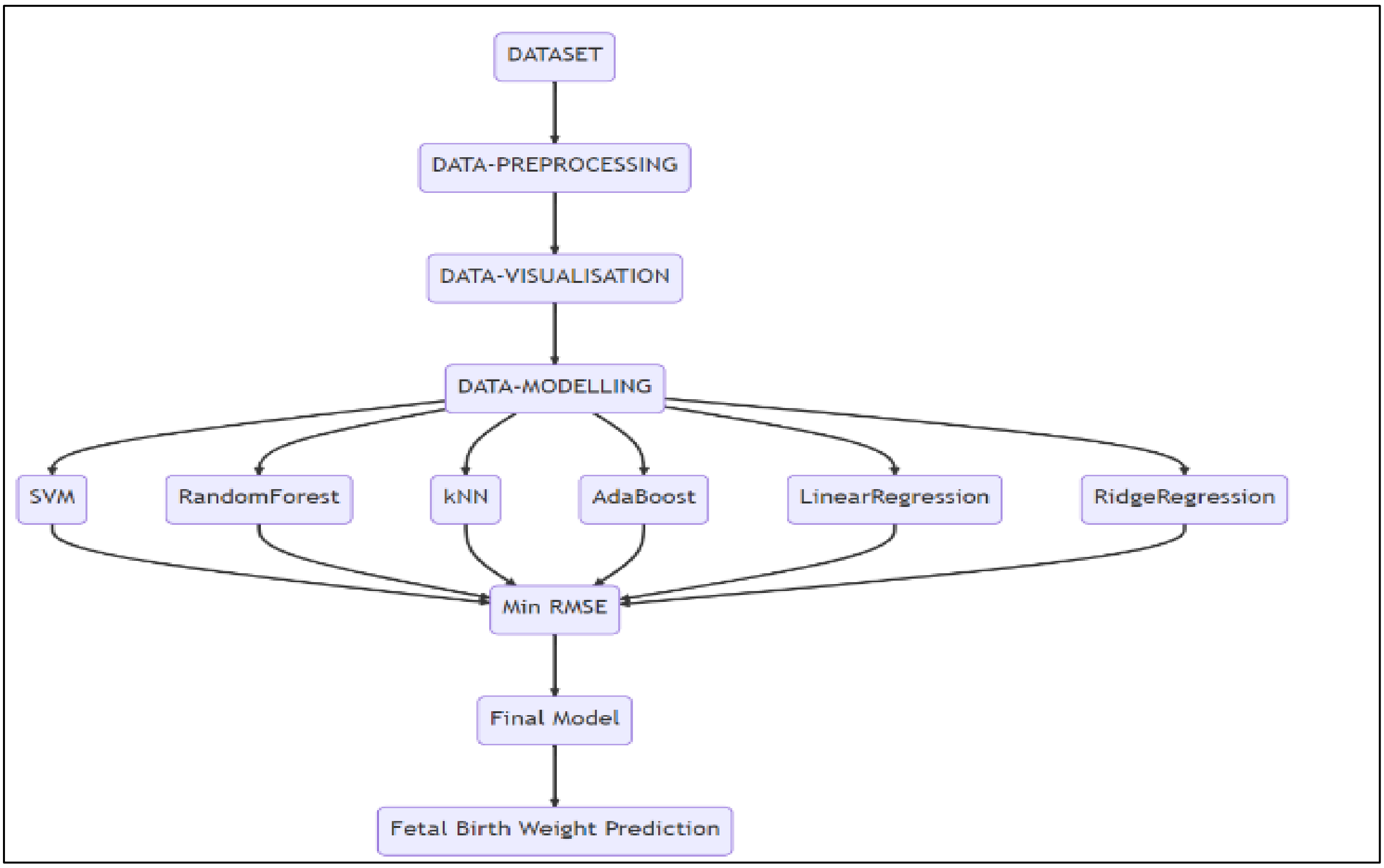

Fig 3 Flowchart for Fetal Birth Weight Prediction

* *Flowchart for Fetal Health Classification:*

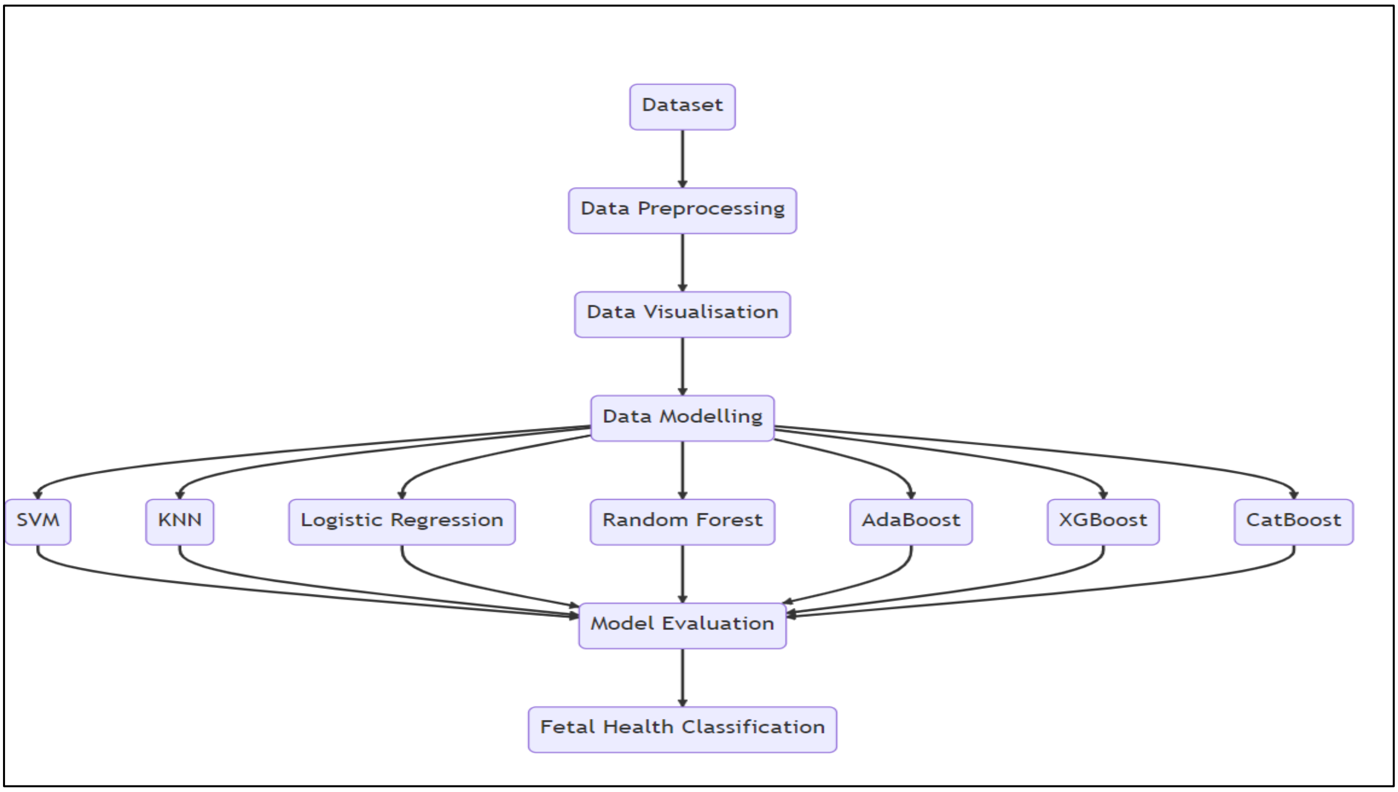

Fig 4 Flowchart for Fetal Health Classification

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24OCT1526

IV. RESULTS

➢ *Comparative Study of Models with and Without SMOTE: -*

Table 1 Comparative Study of Models with and Without SMOTE

Model	Accuracy (without SMOTE)	Accuracy (with SMOTE)
SVM	0.91	0.92
KNN	0.91	0.92
Logistic Regression	0.90	0.90
RandomForest	0.94	0.97
AdaBoost	0.84	0.85
XGBoost	0.92	0.97
CatBoost	0.95	0.99

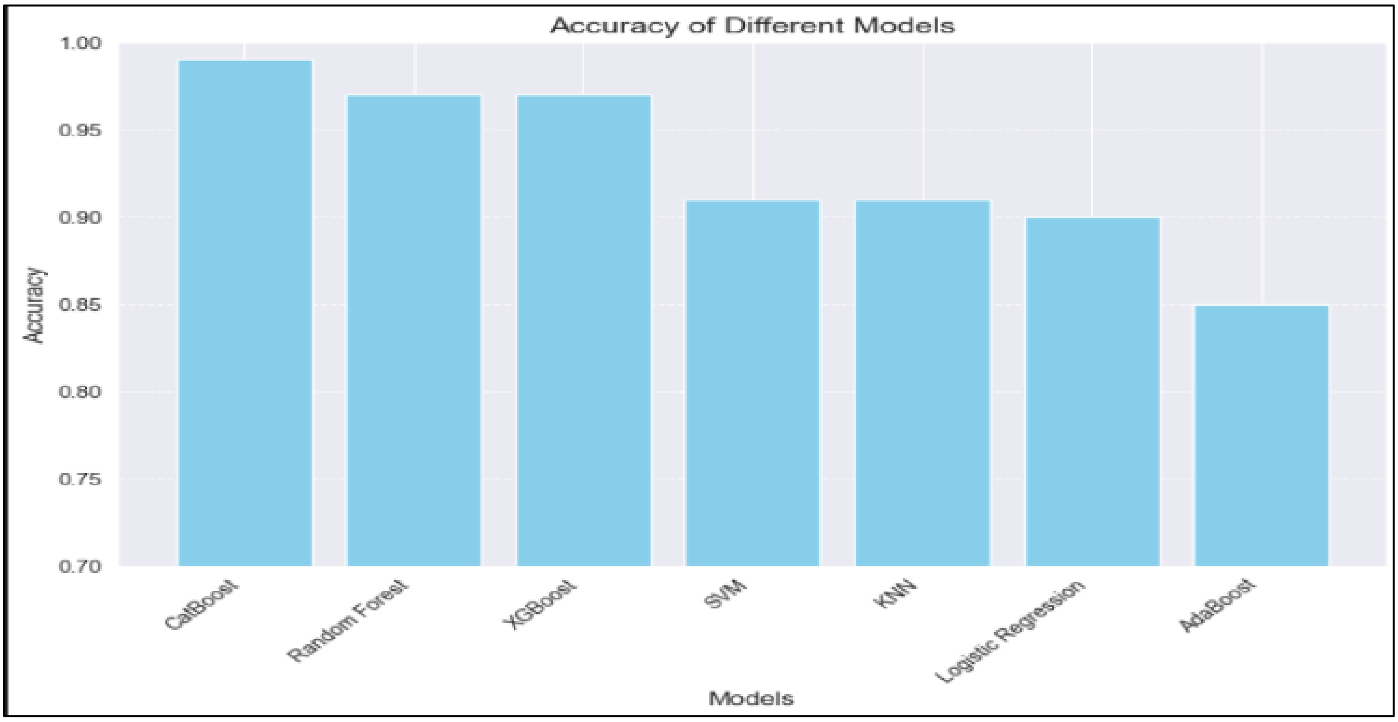

Fig 5 Accuracy Bar Graph

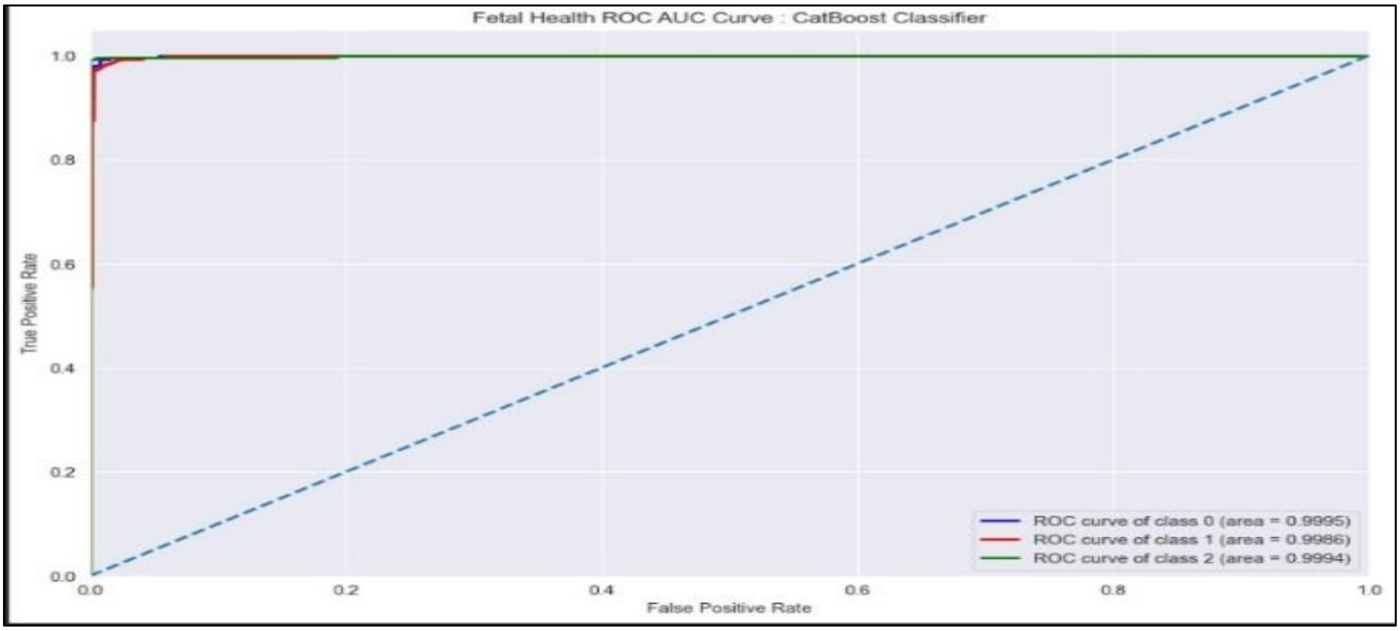

Fig 6 ROC AUC Curve for Catboost model

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24OCT1526

With the results of the SMOTE implementation with CatBoost being nearly perfect, we will include the ROC AUC curve for some of the models.

V. CONCLUSION

In conclusion, the results of this study suggest that, even though several boosting algorithms are used for comparison, the CatBoost model is a promising tool for fetal health classification. The model's high accuracy, generalizability, and sensitivity to potential abnormalities suggest that it could be used to improve the early detection and management of fetal health problems. Evaluation metrics such as accuracy, precision, recall, and F1-score provided insights into the models' capabilities in correctly identifying instances of normal, suspect, and pathological fetal health. Furthermore, techniques such as cross-validation and hyperparameter tuning were employed to optimize model performance. Regarding the fetal birth weight estimation, our findings imply that clinical assessment of birth weight can be used as a diagnostic tool, and that clinical estimation is sufficient for managing labour and delivery in a term pregnancy. This study found that clinical birth weight estimation can help manage labor and delivery in a term pregnancy, even in developing country like India. Our findings are significant because ultrasound is not widely available in many health-care delivery systems in developing countries like ours, particularly in rural areas. Our user friendly Interface makes it easy for common man to use these models more effectively.

FUTURE ENHANCEMENT

Further we can enhance this model by adding different image processing techniques for the classification of fetal health. We can also try to explore and engineer new features from the existing dataset or incorporate additional relevant data sources to enhance the discriminative power of the models. The birth weight estimation in case where twins will be born can be another area of future research. Establish mechanisms for continuous monitoring, evaluation, and updating of the fetal health classification and birth weight estimation models based on feedback from healthcare providers, evolving clinical guidelines, and advancements in predictive modeling techniques. Collect data from wearable sensors worn by pregnant individuals to capture real-time physiological parameters such as heart rate variability, physical activity levels, sleep patterns, and stress levels. Incorporating these novel data streams into birth weight estimation models could offer dynamic and personalized insights into maternal-fetal health dynamics. By pursuing these avenues for enhancement, the fetal health classification and birth weight estimation models can evolve into more reliable, interpretable, and impactful tools for prenatal care, ultimately contributing to improved maternal and neonatal health outcomes.

REFERENCES

[1]. Unveiling the Unborn: Advancing Fetal Health Classification through Machine Learning(2023): -Sujith K Mandala
[2]. Early Diagnosis and Classification of Fetal Health Status from a Fetal Cardiotocography Dataset Using Ensemble Learning(2023):- Adem Kuzu and Yunus Santur
[3]. Fetal Health Classification from Cardiotocograph for Both Stages of Labor—A Soft-Computing Based Approach(2023):-Sahana Das, Himadri Mukherjee, Kaushik Roy and Chanchal Kumar Saha
[4]. Using Machine Learning to Classify Human Fetal Health and Analyze Feature Importance(2023):-Yiqiao Yin and Yash Bingi
[5]. Machine learning applied in maternal and fetal health: a narrative review(2023):
[6]. -Daniela Mennickent,André Rodríguez, Ma Opazo, Claudia Riedel,Erica Castro, Alma Eriz-Salinas, Javiera Appel-Rubio, Claudio Aguayo, Alicia Damiano, Enrique Guzmán, Gutiérrez
[7]. A Review on Deep-Learning Algorithms for Fetal Ultrasound-Image Analysis(2022):-Maria Chiara Fiorentino, Francesca Pia Villani, Mariachiara Di Cosmo, Emanuele Frontoni, and Sara Moccia
[8]. Comparative Analysis of Different Efficient Machine Learning Methods for Fetal Health Classification(2022):-Md Takbir Alam,1 Md Ashibul Islam Khan,Nahian Nakiba Dola, Tahia Tazin , Mohammad Monirujjaman Khan
[9]. Estimation of fetal weight by clinical method and its comparison with ultrasonography and its correlation with actual birth weight in term singleton pregnancy(2022):-Kantamani Pavithra, Kanaparthy Priyadarshini, Y Annapoorna