Fetal Distress Prediction Based on Cardiotocographic (CTG) Data

ML PROJECT PRESENTATION

Group - 7

INDRAPRASTHA INSTITUTE *of*
INFORMATION TECHNOLOGY **DELHI**

Motivation

- ❖ The number of fetal and maternal deaths every year worldwide is staggering.
- ❖ Cardiotocography -

  ➔ monitoring technique used to determine a fetus' healthy being

  ➔ by simultaneously records the fetal heart rate and the mother's uterine contractions.

  ➔ provides obstetricians with crucial information about fetal state, used to detect abnormal fetal state and movements

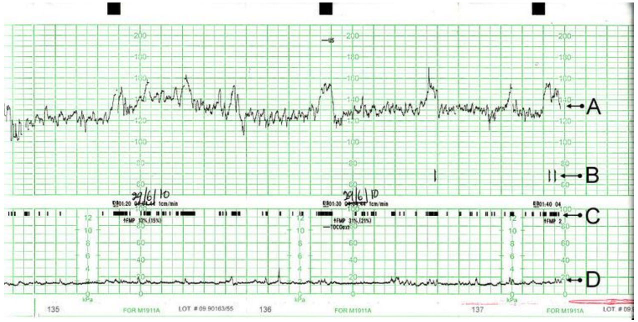
- ❖ Drawback - visual inspection of the data is often unreliable.
- ❖ Over 50% of fetal deaths are due to this inconsistency in pattern recognition and failure in receiving a timely intervention.

Motivation

What can be done to combat these challenges introduced due to inconsistent interpretations of CTG as a result of human error?

- ❖ Integrate computerised machine learning methods with obstetrician interpretations for better prediction of fetal state.
- ❖ develop a machine learning model that can identify high-risk fetuses accurately comparable to highly trained medical professionals.

We hope that this would play a significant role in reducing fetal mortality and congenital disabilities globally.



Example of CTG output. A: Fetal heartbeat; B: Indicator showing movements felt by mother (triggered by pressing a button); C: Fetal movement; D: Uterine contractions

Literature Review

Research Paper 1: *Classification and Feature Selection Approaches for Cardiotocography by Machine Learning Techniques*

In this paper[2], both R and Python machine learning techniques are used for performance analysis. Four different types of feature selection based on feature correlations and various models are employed for this study.

Research Paper 2: *Fetal state assessment based on cardiotocography parameters using PCA and AdaBoost*

Classification labels used in this paper are normal and pathological. *Principal Component Analysis (PCA)* performs dimensionality reduction and feature selection. It struggles with the problem of outliers as the dataset used has no outlier values, which is not justifiable in the real world. [3]

Literature Review

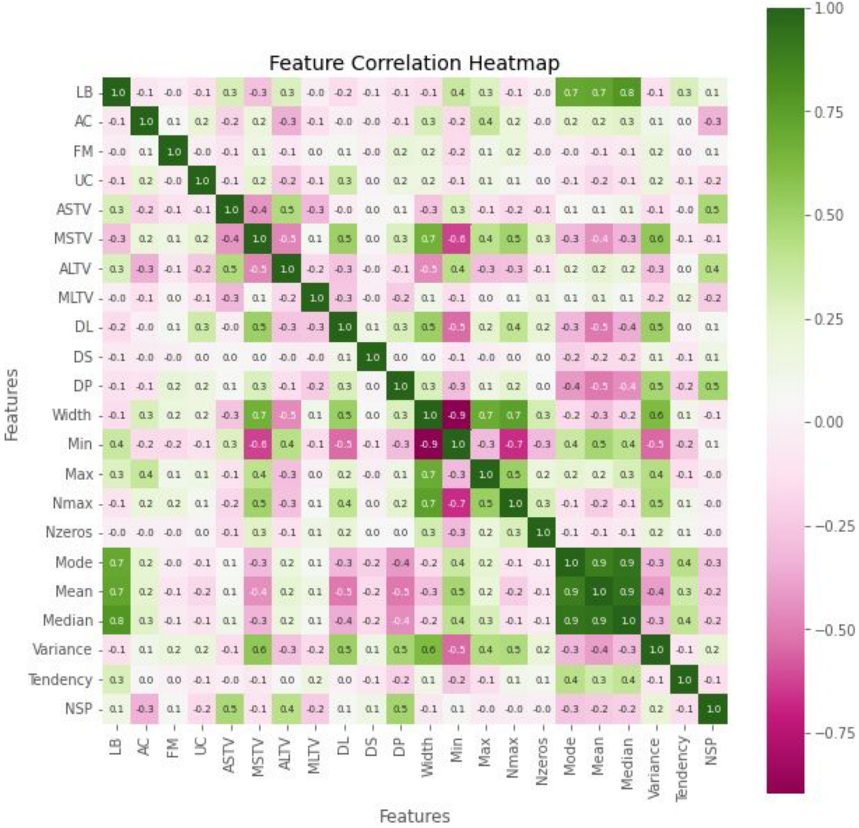
Research Paper 3: Comparison of Machine Learning Techniques for Fetal Heart Rate Classification

Unlike other studies, the authors in paper [4] also evaluated extreme learning machines[ELM] algorithm with five different activation functions apart from Random Forest Classifier, Support Vector Machines, Artificial Neural Network, and Radial Based function network.

Dataset Description

- We have used the Cardiotocography raw data from the UCI Machine Learning Repository available at https://archive.ics.uci.edu/ml/datasets/cardiotocography#
- The data consists of 2126 data samples and 28 features.
- It gives two types of classifications -
  1) Morphologic pattern (10 classes) and
  2) Fetal state (3 classes - N, S, P)

We have plotted a correlation heatmap of various features.



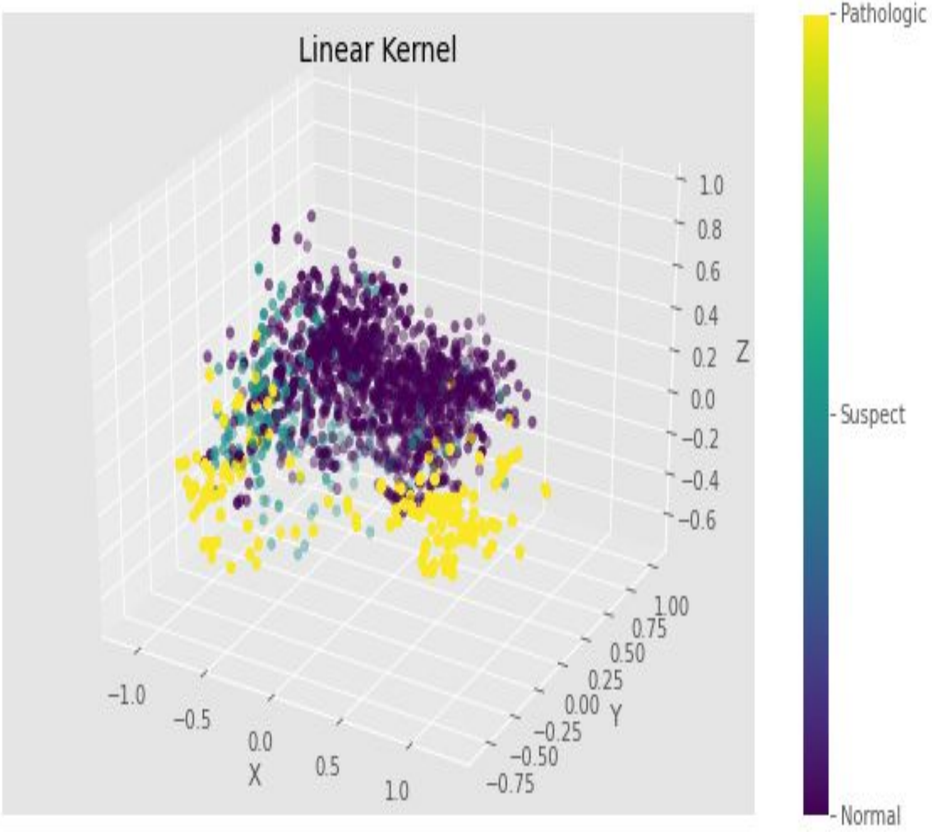
Dataset Description - PREPROCESSING

1) **Feature selection and cleaning -** We have dropped 7 irrelevant features as well as rows with null values and removed duplicate data samples. We thus obtained 2112 data samples and 21 features.
2) **Normalization -** In order to give equal weights to each feature in the dataset so that no single variable steers model performance in one direction, we performed data normalization using MinMax Scaler to fit between 0 and 1.
3) **Oversampling data -**

❖ The division of data into the three classes is imbalanced.

❖ We have used the random oversampling technique to avoid overfitting of the machine learning model on skewed classes by increasing the data samples of the classes with minority instances.

Dataset Description - Principal Component Analysis



- ❖ We performed dimensionality reduction using the Principal Component Analysis techniques on various kernels to reduce the dimension of the data.

- ❖ We noticed that our choice of 21 features was appropriate. Interestingly, we observed that each kernel created two distinct groups of "Pathologic" class : one is close to "Suspect" and one is far from it.

- ❖ However, the data is not linearly separable as observed from the plots of various kernels. We have added the plot for the "Linear" kernel (Figure 3) for reference. All other kernels give similar plots.

2 types of classification

Class Codes	Classes
N	Normal
S	Suspect
P	Pathologic

Table 1 : **3 - class description: Fetal State**

Class Codes	Classes
A	Calm Sleep
B	REM Sleep
C	Calm Vigilance
D	Active Vigilance
SH	Shift Pattern
AD	Accelerative/Decelerative pattern
DE	Decelerative pattern
LD	Largely Decelerative pattern
FS	Flat - Sinusoidal pattern
SUSP	Suspect Pattern

Table 2 : **10 - class description: Fetal Heart Rate Patterns**

Methodology

- The dataset was split into training and testing set using a 70:30 stratified split. After that we performed a 3 fold cross validation and achieved a train:validation:test split of 47:23:30.
- The following models were implemented for 3 class as well as 10 class classification on the dataset using the sklearn library-
  - Multinomial Logistic Regression
  - Gaussian Naive Bayes
  - Decision Trees
  - Random Forest
  - K-Nearest Neighbours
  - Boosting
  - Bagging
  - Support Vector Machine
  - Multi Layer Perceptron
- We also performed hyperparameter tuning using GridSearchCV and chose the best model for training and testing.
- We used matplotlib and seaborn plotting and visualization → histograms, ROC curves and matrices etc.

Methodology

❖ We used the following performance metrics to test our models ➔

- ➢ **Accuracy** - measures the overall efficiency of a classifier.
- ➢ **Precision** - ratio of true positives to the total of the true positives and false positives.
- ➢ **Recall** - ability of a classifier to categorize positively labeled data.
- ➢ **F1 score** - harmonic mean between precision and recall, gives good tradeoff between them

Classifier Algorithm	Accuracy	Precision	Recall	F1 score
Logistic Regression	91.95	91.47	91.54	91.50
Naive Bayes	73.14	88.67	73.14	77.45
KNN	95.26	94.63	95.26	95.01
Decision Tree	94.61	94.63	94.61	94.62
Random Forest	97.21	97.11	97.12	97.10
Boosting	95.26	95.42	95.26	95.33
Bagging	97.49	97.41	97.49	97.39
Support Vector Machine	96.37	96.25	96.37	96.29
Multi Layer Perceptron	94.88	94.71	94.88	94.78

Table 4 : **Evaluation Metrics for 3 class classification**

Classifier Algorithm	Accuracy	Precision	Recall	F1 score
Logistic Regression	80.91	81.02	80.91	80.66
Naive Bayes	13.09	5.68	13.24	7.12
KNN	76.02	76.00	76.02	75.46
Decision Tree	77.13	77.36	77.13	77.15
Random Forest	88.01	88.09	88.01	87.81
Boosting	64.83	66.36	64.82	61.82
Bagging	86.75	86.68	86.75	86.36
Support Vector Machine	82.33	82.32	82.33	82.23
Multi Layer Perceptron	84.54	84.70	84.54	84.53

Table 5 : **Evaluation Metrics for 10 class classification**

Methodology

Applied grid search technique on each model to find the optimal hyperparameters for it and got following model parameters :

Classifier Algorithm	Optimal Parameters
Logistic Regression	C=166.81, max_iter=5000, penalty='l1', solver='saga', multi_class = 'multinomial', random_state = 0
Naive Bayes	estimator = GaussianNB(), 'var_smoothing' = 0.139
KNN	n_neighbors= 1, metric = 'euclidean', weights= 'distance', algorithm = 'ball_tree'
Decision Tree	max_depth = 18, criterion = "entropy", max_features="auto" splitter = "best", random_state=0
Random Forest	n_estimators = 125, criterion = "entropy", max_features = 'auto'
Boosting (ADABOOST)	n_estimators = 135, algorithm='SAMME', learning_rate=1
Bagging	max_features=10, n_estimators=30, oob_score=True
Support Vector Machine	C=166.81005372000558, degree=2
Multi Layer Perceptron	activation='logistic', max_iter=300, random_state=1,solver='lbfgs'

Table 3 : **Tuned Hyperparameters for 3 class classification**

Results

For the 3 class classification, the highest accuracy has been observed both in Bagging (97.49%) and Random Forest (97.21%) Classifiers and Gaussian Naive Bayes gives the least accuracy (73.14%).

For the 10 class classification, the highest accuracy has been observed both in Random Forest (88.01%) and Bagging (86.75%) Classifiers and Gaussian Naive Bayes gives the least accuracy (13.09%).

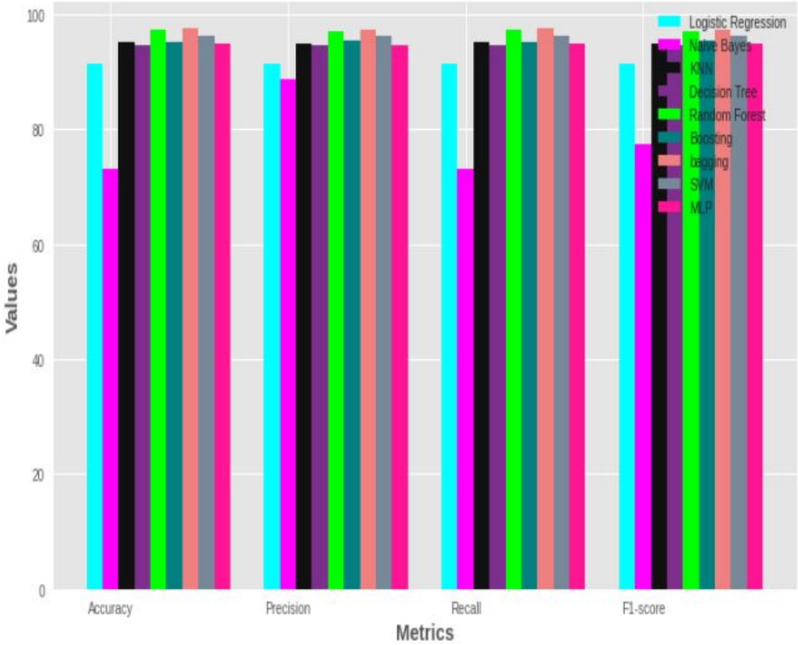
- Multinomial logistic regression ➜ perform multi-class classification, which offers better performance over the One-vs-All method
- Gaussian Naive Bayes ➜ least accuracy, assumes no dependency between attributes which is not true as the heat map depicts decent correlation between various features
- Decision tree ➜ handles high dimensional non parameterized data and works well with non-linearly separable patterns

Results

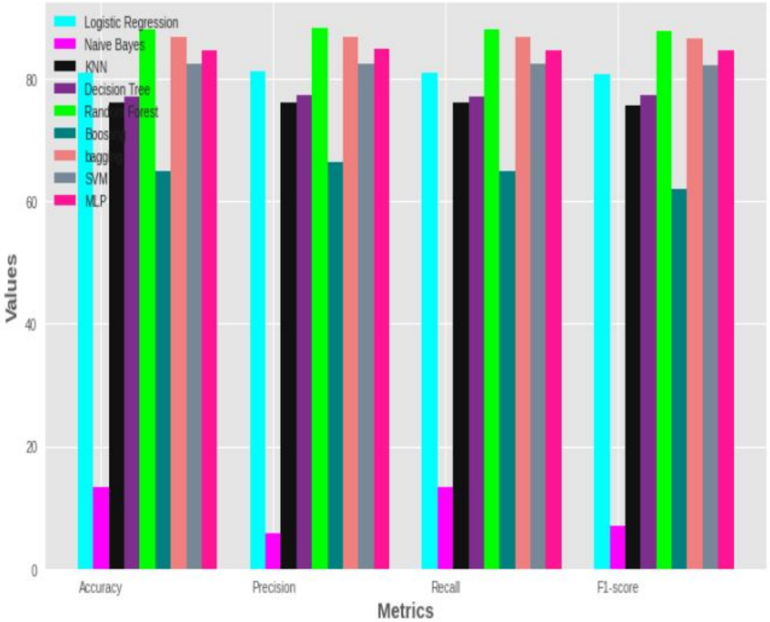
- <u>Random forest</u> ➔ ensemble method, combines the output of multiple unpruned decision trees and makes a prediction based on the majority vote
- <u>K-Nearest Neighbours model</u> ➔ works well because it operates on the correlation of features and favours noiseless data
- <u>Boosting</u> ➔ ensemble method which combines multiple weak learners to make a strong learner based on weighted averages
- <u>Bagging</u> ➔ ensemble learning technique, gives one of the best accuracies as it involves randomly selecting data with replacement in order to achieve reduced variance
- <u>Support Vector Machine</u> ➔ classifies the given labelled training data by creating an optimal hyperplane for classification
- <u>Multi Layer Perceptron</u> ➔ feedforward ANN consisting of input layer, output layer and hidden layers, uses the back propagation algorithm to classify instances of data

COMPARISON BETWEEN EVALUATION METRICS OF VARIOUS MODELS

➢ **3 class classification:**



➢ **10 class classification:**



Conclusion

- We learnt that preprocessing, extensive machine learning model techniques, and tools employed play a vital role in classification scores analysis
- Our methodology not just includes data cleaning and normalization but also incorporates advanced feature selection and engineering methods like Principal Component Analysis used for Dimensionality Reduction.
- For 3 class classification, we have achieved almost perfect accuracy, precision, recall, and F1 score metrics of around 95-98% in 7 out of 9 models.
- For 10 class classification, we have achieved an impressive accuracy, precision, recall, and F1 score metrics of around 80-88% in 5 out of 9 models.

Future Improvement - The dataset does not consider differences in sociodemographic characteristics of pregnant women and some other relevant features like age, nutritional status and so on. The results can be improved by obtaining a larger dataset with diverse sociodemographic characteristics.

Timeline

We stuck to the schedule in terms of the work done and have covered everything that we needed to do. Following is the work that we have done -

1	Pre-processing data, Data visualization
2	Feature Analysis and Selection, Plotting Maps, Dimensionality Reduction, Logistic Regression
3	Naive Bayes, Decision Trees(3 class classification)
4	Random Forests, K - Nearest Neighbours(3 class classification)
5	Analysis of Model Performance, Hyperparameter Tuning
6	LR, Naive Bayes, Decision Tree, Random Forest, KNN(10 class classification)
7	Bagging and Boosting(3 and 10 class classification)
8	Support Vector Machine, Multi Layer Perceptron(3 and 10 class classification)
9	Analysis of Model Performance, Hyperparameter Tuning Advanced Models, Drawing Final Conclusions
10	Report Writing and Presentation Making

Individual Contribution

- **Suyashi Singhal** → Data Preprocessing and Visualization, Result Analysis, Training Models and Hyper Parameter Tuning - [DT, KNN, LR, MLP, Bagging], Report Writing, Making Presentation
- **Ayush Mahant** → Dimensionality Reduction (Principal Component Analysis), Result Analysis, Training Models and Hyper Parameter Tuning - [RF, KNN, LR, MLP, Boosting] Report Writing, Making Presentation
- **Harshita Gupta** → Plotting Maps, Model selection, Training Models and Hyper Parameter Tuning -[RF, NB, LR, Bagging, SVM] Report Writing, Making Presentation
- **Rasagya Shokeen** → Plotting Maps, Model Selection, Training Models and Hyper Parameter Tuning - [DT, NB, LR, Boosting, SVM] Report Writing, Making Presentation

These were the assigned responsibilities for each team member. However, all the members equally contributed to all the work done.

THANK YOU