Fetal risk prediction through different supervised machine learning techniques

<u>Submitted by</u>

Meherab Hassan

ID: 181-35-2375

Department of Software Engineering

Daffodil International University

<u>Supervised by</u>

Ms. Syeda Sumbul Hossain
Lecturer

Department of Software Engineering

Daffodil International University

This Project report has been submitted in fulfillment of the requirements for the Degree of Bachelor of Science in Software Engineering.

APPROVAL

This thesis/project/internship titled on "**Fetal risk prediction through different supervised machine learning techniques**", submitted by Meherab Hassan, ID: 181-35-2375 to the Department of Software Engineering, Daffodil International University has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of Bachelor of Science in Software Engineering and approval as to its style and contents.

BOARD OF EXAMINERS

-- Chairman

Dr. Imran Mahmud
Associate Professor and Head
Department of Software Engineering
Daffodil International University

-- Internal Examiner 1

Nusrat Jahan
Assistant Professor
Department of Software Engineering
Daffodil International University

-- Internal Examiner 2

Khalid Been BadruzzamanBiplob
Senior Lecturer
Department of Software Engineering
Daffodil International University

-- External Examiner

Professor Dr M Shamim Kaiser,
Professor
Institute of Information Technology
Jahangirnagar University

DECLARATION

I hereby declare that this thesis has been done by us under the supervision of **Ms. Syeda Sumbul Hossain**, Lecturer, Department of Software Engineering, Daffodil International University. It is also declared that neither this thesis nor any part of this has been submitted elsewhere for award of any degree.

Meherab

Meherab Hassan

Student ID: 181-35-2375
Batch**:** 25th
Department of Software Engineering
Faculty of Science & Information Technology
Daffodil International University

Certified by:

Ms. Syeda Sumbul Hossain
Lecturer
Department of Software Engineering
Faculty of Science & Information Technology
Daffodil International University

ACKNOWLEDGEMENT

First of all, we are grateful to the Almighty Allah for giving us the ability to complete the final thesis.

We would like to express our gratitude to our supervisor **Ms. Syeda Sumbul Hossain** for the consistent help of my thesis and research work, through his understanding, inspiration, energy, and knowledge sharing. Her direction helped us to finding the solutions of research work and reach to our final theory.

We would like to express my extreme sincere gratitude and appreciation to all of our teachers of **Software Engineering** department for their kind help, generous advice and support during the study.

We are also express our gratitude to all of our friend's, senior, junior who, directly or indirectly, have lent their helping hand in this venture.

Last but not the least, we would like to thank our family for giving birth to us at the first place and supporting me spiritually throughout my life.

Meherab Hassan

181-35-2375

Contents

Chapter 1
Introduction

Background:

Childbirth is a natural phenomenon through which every living being on the planet saves their extinction. Doctors and health experts made it clear that this is yet a sensitive and necessary occurrence and this should never be treated as an illness. But due to some internal or external features of the mother or the fetus abnormalities are often shown during childbirth and it becomes somewhat fetal for either or both. According to a WHO report substantial global progress has been made in reducing child deaths since 1990. The total number of under-5 deaths worldwide has declined from 12.6 million in 1990 to 5.2 million in 2019. Since 1990, the global under-5 mortality rate has dropped by 59%, from 93 deaths per 1,000 live births in 1990 to 38 in 2019. (World Health Organization, Children: improving survival and well-being, link https://www.who.int/news-room/fact-sheets/detail/children-reducing-mortality)

The crucial age for a child to have disease by birth is from the age from birth to 5. During this time the body remains sensitive to the external environment most of the fetal outcome occurs during childbirth. Therefore CTG is a necessary reading which can show the overall condition of the heart of the fetus. According to the obstetricians the heart of the fetus in the second and third trimester of the pregnancy is very sensitive, as it is much different than a normal human heart, and it should be monitored. Up to 50% of FHR patterns classified as pathological reflect physiological changes and can therefore be classified as false positives (false pathological). This can lead antepartum to increased numbers of induced births and higher numbers of operative deliveries. [1]

Machine learning and Artificial Intelligence will play a vital role in stabilizing the efforts and improving the monitoring of FHR through Cardiotocogram data. Through different research people have progressed on this topic but a different approach is yet still necessary for enhancing the accuracy and perfection.

Motivation:

The information obtained from CTG is used for early detection of pathological conditions to help obstetricians and gynecologists make predictions. Prevents future problems and permanent harm to the fetus before they occur. Through the birth of a baby, Hypoxia can cause temporary disability and death. Due to misdiagnosis of FHR pattern recording and Inappropriate treatment of the fetal can lead to more than half of these deaths . Convenient, but there is Successful CTG surveillance can be inconsistent, especially in low-risk pregnancies. Inaccurate case Evaluated fetal pain, which can lead to useless treatment or if there is an inadequate assessment of fetal well-being Then essential treatments may be ruled out.

There are numerous researches in CTG fetal risk prediction but the main motivation towards this research is to avoid biased decision making by the applied model and to organise the necessity of the features of the dataset before fitting it in a model.
According to the past couple of researches on this particular topic it was acknowledged that the visual analysis of the CTG data done by the obstetricians could not have been objective and correct. Therefore, in this research we are putting up effort in making the dataset as unbiased as possible before fitting into the model.

Problem Statement:

According to the current research the major limitation of the Research are [2] The visual analysis of the CTG data done by obstetrician could not have been objective
and correct. [3] The used dataset is an imbalanced one so in future the dataset
is to be balanced first and then the associative classification
model is to be modified with some evolutionary flavor to get
better accuracy results. [4] work is that this dataset was obtained from a repository in the developed world. Due to the differences in sociodemographic characteristics of pregnant women in LMICs, the machine learning algorithm may report a different accuracy. Further, this dataset did not include any information on participants' sociodemographic data or other relevant clinical characteristics, such as primiparity, maternal nutritional status, and anemia, gestational age, fetal well-being, etc. which may affect the intrapartum course of events and could potentially contribute toward
further refinement of the AI model.

Research Questions:

RQ1: How can we get the unbiased accuracy of the CTG readings?
RQ2: What will be the accuracy after implementing the selected features?

Research Objectives:

RO1: To get unbiased results from the categorical data.
RO2: To gain a more unbiased and improved result from machine learning models.

Research Scope:

The main scope of this kind of research is to lessen the amount of time and labour given to analyze a CTG reading that is directly derived from the machine. But, the process is basic and it was done before therefore, the scope lies on the discussion that, how accurately and how just is the result of the research. With the success of the research the process can be sent to a productional level which any health clinic or any hospital can use to lessen the amount of their stress about finding the complication in a fetus in its early stage with just the help of plain numbers in a CTG scan.

The dataset which is used in this research is being balanced through pre-processing tools, as a matter of fact it can be used to justify the modeling process of further machine learning research done through it.

Thesis Organization:

This thesis report consists of five sections. The first section is just the introductory part. In this section a brief introduction about the topic of the research, it's socio economic viewpoints and overall summary is shown. In the next section the Literature review will be summarized to give a detail about the works findings and problems or limitations of the previous researches will be shown. Then the report will show the modeling part of the research, that will be the core part of the report because details about the data set on which we are working on and the model fitting analytical description and even the description and orientation of different machine learning models used in this topic will be discussed. Later on the results and the outcome of this research will be discussed in the result and conclusion section. It will be the portion of this report where most of the numbers, statistical diagram and even the values will be shown. The outcome will be described and the last part will consist of the discussion and limitation of my outcome and future contribution of the research i.e in the conclusion and recommendation portion.

After these five chapters reference of this report will be given at the end of the report.

Chapter 2
Literature Review

The method of monitoring cardiotocogram is very common now a days during the time of pregnancy, it is a method to read the heart beat variation of the fetus and has a temporal relation to the uterine contraction to the mother, thus the fetal health condition can be determined by the reading. In this research some constructive reviews about the models that were used before and their failings as well as their success were performed.

1st point of research

The most recent findings were based on an ensemble classifier model. [2] The whole analysis was meant for the improvement of the previous result of calculating fetal risk of child and parent before and after the Birthing procedure. By training the classifiers' data with k-fold cross validation and using the the algorithm of Bagging with Random
Forest. They achieved an accuracy of 99.02%. The experimental results of this study reveal that Bagging ensemble with Random Forest can be utilized to classify the normal and pathological cases of the CTG data. According to the research they seem to claim that after visualization the obstritians could be wrong and unobjective. It is a matter of fact that they could have overlooked the correlation between the features and their data or the imbalancement of the dataset [3]. Therefore having such a perfect accuracy in a data that is classified into multiple classes made a major contribution. Due to this a study was made [5] for the differentiation of ensemble technique and deep learning for the detection of hypoxia. To analyze hypoxia detection performance in each model, evaluation focuses on precision, recall, and f1-score for class 0 (normal) and class 1 (hypoxia) where the model can recognize the normal and hypoxia label appropriately. In the first scenario among the ensemble learning models used were Bagging Tree, AdaBoost, and Voting Classifier and Deep learning models. So here a using bagging multiclass dataset should seem to error in accuracy.

2nd point of research

In this particular portion the research was based on the heart rate analysis and metabolic movement of the fetus. [6][7] The contraction progressed in the last few minutes. Normal and academic fetus Significant increase in linear frequency domain index one two Three On the other hand, the nonlinear index has decreased significantly. Moderate to severe fetal anemia, Significant reduction in nonlinear index, both Recording start and end logs. acceptable Discrimination against moderate to severely acidic foets It was possible only in combination with linear Non-linear index. As the work requires a new integrated methodology Identify FHR signals from a fetation suspected of developing metabolic acidosis. This is achieved by applying the analysis and processing steps of the FHR signal. A set of features described in relation to this particular description problem. Then the SVM is proposed and used for management of difficult tasks to classify and identify the problem.

3rd point of research

AS it is known that a healthy pregnancy leads to a healthy birth, Some researches were made to classify the problems on the basis of association. There is an association-based classification model For access related to fetal health UCI cardiac tomography data. The above classifier is It can be interpreted from 4 versions and experiments Its CBA-M2 reduces the time required to build the model. It reduces the quality of service. After comparing with Some standard classifiers with Random Forest, XGBoost works well for the purposes required [3]. The classification model developed with XGBoost technology showed the highest predictive accuracy for adverse fetal outcomes. General health professionals in low- and middle-income countries can use this model to classify pregnant women in remote areas for early referral and further treatment [4].

Related work

Most of the studies conducted on finding the accuracy of fetal health of a child is done by validation of their results on different machine learning algorithms. But very few managed to judge the data while pre-processing it. They made it clear and it’s only natural for a raw set of data to be vivid of the amount of fetal outcome in each class. As the dataset was improvised to the contribution of fetal outcome which 3 certified obstetricians worked hard to analyze. But it was revised and new editions of the dataset came throughout the passing of years but while fitting it into a certain model in machine learning, the outcomes were questionable and although according to the numbers the accuracy was looking good but studies made it unsure for them if the dataset was well balanced or not [3][4][8].

Chapter 3
Methodology

Design of the System Model:

The methodology starts with a set of data collected which is the CTG data of pregnant women on their third trimester of pregnancy. We put out research through the following steps where the entire methodology after the paper reviews is shown in a constructive manner.

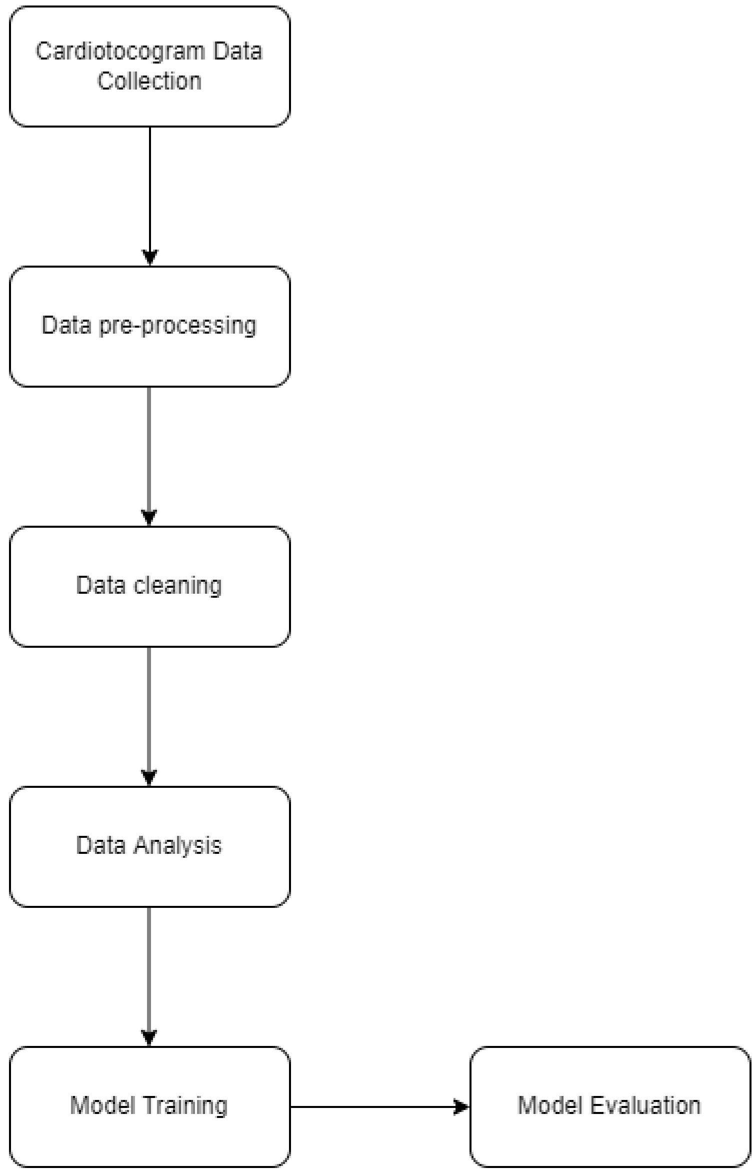


Fig. Research Methodology

In this structure after the step of visualizing the data the amount of the classified outcomes and verified what the previous research papers were pointing out was acknowledged. After this observation we remodeled the construction of the methodology and we preprocessed the data with SMOTE analysis This was the most important part of the programming section before entering the whole dataset in the model. Different research has made different approaches to balance the dataset. The most common among them is the bagging which has an exemplary accuracy but they seem to have problems with the fetal outcomes determined by the obstructions [2]. As bagging was used after the model was trained to avoid the result from getting biased. We used SMOTE analysis and pearson correlation of feature selection to remove data redundancy and biased outcome.

Dataset:

Dataset information:

This dataset originated in Ayres de Campos et al. (2000) SisPorto 2.0 A Program for Automated Analysis of Cardiotocograms. J Matern Fetal Med 5:311-318. Origination link (https://onlinelibrary.wiley.com/doi/10.1002/1520-6661(200009/10)9:5%3C311::AID-MFM12%3E3.0.CO;2-9)
The dataset consists of 2126 records of features extracted from Cardiotocogram exams which were then classified by three expert obstetricians into 3 classes:

- Normal
- Suspect
- Pathological

In view of the foregoing, cardiotocography (CTGs) are a straightforward and inexpensive way for healthcare practitioners to assess fetal health and take measures to reduce child and maternal mortality. The machine sends ultrasound pulses and reads the response, providing information on the fetal heart rate (FHR), fetal movements, uterine contractions, and more.

Dataset Requirement:

This research is based on figuring out the fetal outcome of a fetus in the 2nd or third trimester when the fetus is in the stage of having heartbeats. Based on medical techniques the features regarding the heartbeat of the fetus is to be monitored and the fetal outcome has to be predicted. When the word prediction is discussed we need to have some particular variable of data which can be processed according to the steps to reach our goal. Some of the major requirements are discussed below:

- The data needs to have variability in their respected features so that the outcome can be classified.

- Null values in consistency are harmful for the modeling of the research.
- Mean value of the variability because in plain numbers the record of long and short term variability differs from time to time.
- The Dataset needs to be licensed and ensured by certified professionals to ensure that the dataset is non-forged by AI.
- As the the accuracy of the result the data needs to be properly balanced as this is a multi classified data of more than 2 outcomes of the prediction.

Collection of data:

After knowing what kind of data we require for the thesis, the data was collected from open sources like kaggle UCI or many other platforms, in this study the data was collected from kaggle here is the source link.(https://www.kaggle.com/andrewmvd/fetal-health-classification/download)
After the attributes of the data is to be visualized in a development platform (Jupyter notebook in this research) the attributes of the data is shown below in a table format:

Table: Features and data types

Attributes	Data type
baseline value	int64
accelerations	float64
fetal_movement	float64
uterine_contractions	float64
light_decelerations	float64
severe_decelerations	float64
prolongued_decelerations	float64
abnormal_short_term_variability	int64
mean_value_of_short_term_variability	float64
percentage_of_time_with_abnormal_long_term_variability	int64
mean_value_of_long_term_variability	float64
histogram_width	int64

histogram_min	int64
histogram_max	int64
histogram_number_of_peaks	int64
histogram_number_of_zeroes	int64
histogram_mode	int64
histogram_mean	int64
histogram_median	int64
histogram_variance	int64
histogram_tendency	int64
fetal_health	int64

Data Pre-Processing:

The classificational dataset consisting of multiclass classification needs to be cleaned and checked for and to to remove any irregularities to avoid biased or error results.
Here we have pre processed the data through some steps and their discussion is as follows.

Feature Selection:

Then we executed the feature selection method through the Pearson coefficient.
The **Pearson correlation** is a form of correlation coefficient that shows how two variables measured on the same interval or ratio scale are related. The Pearson coefficient is a metric for determining the strength of a relationship between two continuous variables. The Pearson coefficient is a mathematical correlation coefficient that depicts the link between two variables

(X and Y). Pearson coefficients range from +1 to -1, with +1 representing a positive correlation, -1 representing a negative correlation, and 0 representing no relationship.

$$r = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}}$$

In this study, the execution threshold of the percentage of the correlation coefficient is 85% or 0.85.
Here, the features were held visible with respect to their correlation with one another.

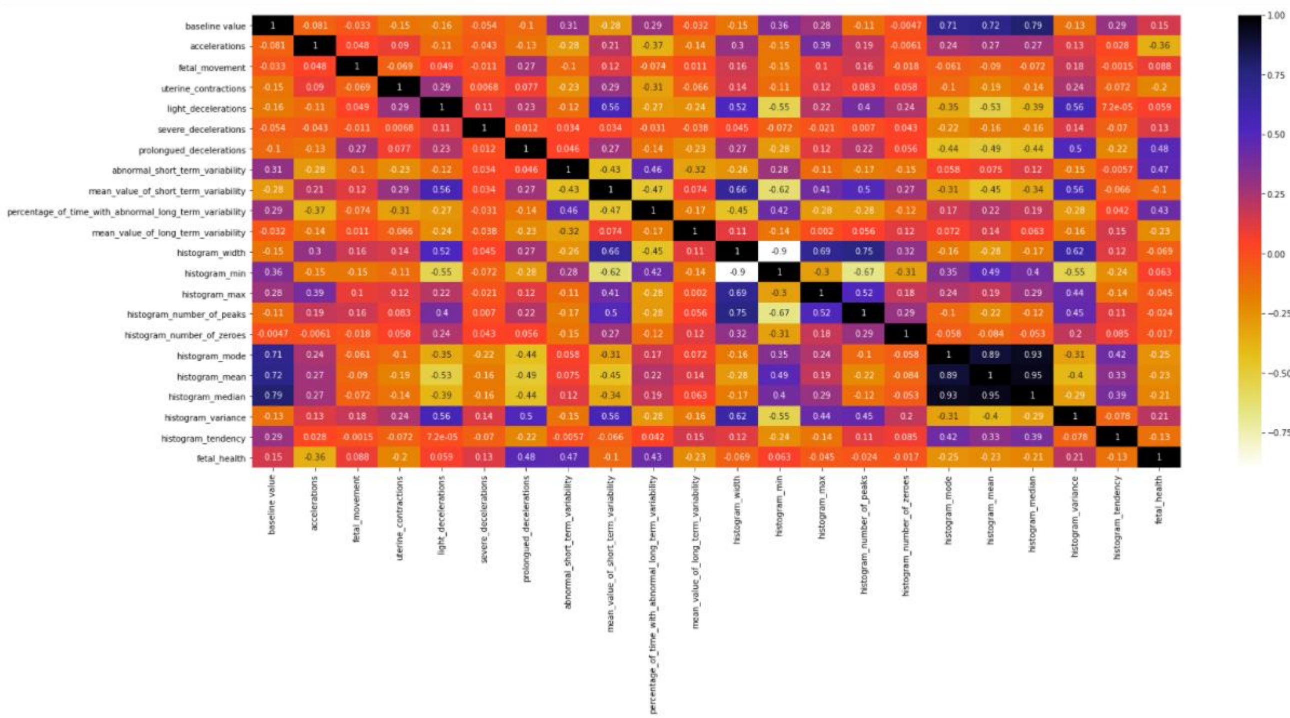


Fig. Correlation of the features

After visualizing the chart it was known that several features hold the correlation threshold, therefore the Pearson correlation algorithm was used using the threshold variable 0.85. By using that the correlation of two features was known to be more than 85% which are 'histogram_mean', 'histogram_median'. Histogram median was the selected feature that was dropped out from the dataset.

Data balance:

While balancing the dataset SMOTE analysis technique was followed. In order to avoid imbalance classification this technique creates multiple virtual rows in the dataset. Imbalanced classification involves developing predictive models on classification datasets that have a severe class imbalance.
Before that the Data was splitted into two parts i.e 80% for training and the rest of the 20% for testing.
When working with imbalanced datasets, the difficulty is that most machine learning techniques will overlook the minority class, resulting in poor performance, despite the fact that performance on the minority class is often the most significant.
SMOTE is an oversampling approach that creates synthetic minority class samples. It is used to create a synthetically or virtually class-balanced training set that is then utilized to train the classifier.The SMOTE samples are linear combinations of two similar samples from the minority class (x and x*R*) and are defined as

$$s=x+u\cdot(x^R-x),$$

with $0 \le u \le 1$; x^R is randomly chosen among the 5 minority class nearest neighbors of x.

Before the training set had 80% of the total rows of data i.e 1700 after we did the SMOTE analysis technique on the number of training set data the amount of data in the training set went up to 3987 and made the number of classes balanced.

Data Preparation:

From the above table it is visible that there are a total of 13 discrete values and 8 continuous values in the data set. The null values were ignored because the whole dataset was converted into a single data type to avoid the numerical error in data while pre-processing. As there weren't any other data to add or some secondary dataset to merge the preparation process was directly jumped into the sorting and clearing. While describing the dataset features and datatypes we avoid the 'fetal_state' feature because that is the feature that is yet to be used in testing our model.

Thus the data after being explored in such detail form was ready to enter into the next step of processing

Data Analysis:

Although the data is clean and it can be processed, observing the attributes is not enough to ensure the data to put into a model fitting the numbers and quantity inside the features helps to get more insight on the condition of classification of the data. The mean amount of the features gives the idea of the dataset condition.

Table. Mean value of the variables

Attributes	**Mean Value**
baseline value	133.303857
accelerations	0.003178
fetal_movement	0.009481
uterine_contractions	0.004366
light_decelerations	0.001889
severe_decelerations	0.000003
prolongued_decelerations	0.000159
abnormal_short_term_variability	46.990122
mean_value_of_short_term_variability	1.332785
percentage_of_time_with_abnormal_long_term_variability	9.846660
mean_value_of_long_term_variability	8.187629
histogram_width	70.445908
histogram_min	93.579492

histogram_max	164.025400
histogram_number_of_peaks	4.068203
histogram_number_of_zeroes	0.323612
histogram_mode	137.452023
histogram_mean	134.610536
histogram_median	138.090310
histogram_variance	18.808090
histogram_tendency	0.320320
fetal_health	1.304327

Fetal outcome of a fetus strongly depends on 4 basic features which are **baseline value, acceleration, variability and deceleration**. But in order to follow the variability which is the most important factor of the fetal outcome as the fetus is under the autonomic nervous system no particular variance value can be obtained from the reading. Because of that we need to assemble the histogram dots in variance and make predictions from the data derived from the long and short term variability.

But as we are talking about the dataset and the equilibrium of the outcome we need to visualize the outcomes in terms of classification

Machine Learning Algorithms:

The dataset applied here is a classificational dataset consisting of 8 continuous features and 13 discrete features. Having too many discrete values is troublesome for applying classification algorithms but as this dataset is classified into multiple classes we need to apply classification algorithms here.

The used models algorithms are described in detail below

[22]Used several machine learning algorithms one of them was SVM along with it's runtime sensitivity having an F1 accuracy score of 99%

[2]The most prosperous outcome was held by bagging with Random Forest 99.02% and they claim it to be biased dataset because of using bagging

[8]Here Hybrid K-SVM was used to obtain an accuracy of 90.64% with 10 fold cross-validation, where SVM has an accuracy of 76% and compared the results
[23]Simple Logistics was used in this research and the result obtained was far better than many other study having and accuracy of 98.74%
[3] IN this research most of the importance was given to Feature selection with puning where KNN algorithm was used they scored 83.54% with feature selection.

Decision Tree Classifier:

For classification and regression, Decision Trees (DTs) are a non-parametric supervised learning approach. The objective is to learn basic decision rules from data attributes to develop a model that predicts the value of a target variable. A tree is an approximation to a piecewise constant. In the shape of a tree structure, a decision tree constructs classification or regression models. It gradually cuts down a dataset into smaller and smaller sections while also developing an associated decision tree. A tree containing decision nodes and leaf nodes is the end result. There are two or more branches in a decision node (for example, Outlook) (e.g., Sunny, Overcast and Rainy). A categorization or judgment is represented by a leaf node (for example, Play). The root node is the topmost decision node in a tree that corresponds to the best predictor. Both category and numerical data may be handled by decision trees.

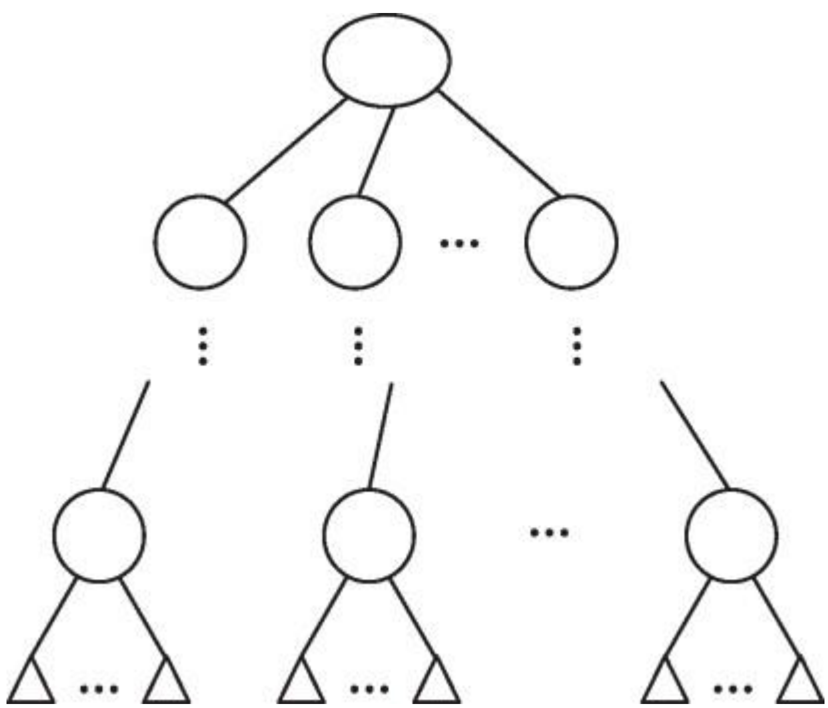


Fig. Decision tree

<u>Random Forest classifier</u>:

Random forest is a machine learning technique developed by Leo Breiman and Adele Cutler that combine**s** the output of numerous decision trees to produce a single outcome. Its popularity is due to its ease of use and adaptability, since it can handle both classification and regression problems.

The random forest algorithm is made up of a collection of decision trees, and each tree in the ensemble is made up of a bootstrap sample, which is a data sample obtained from a training set with replacement. One-third of the training sample is set aside as test data, referred to as the out-of-bag (oob) sample, which we'll discuss later. Using feature bagging, another instance of randomness is injected into the dataset, increasing the dataset's variety and decreasing the correlation between decision trees. The prediction will be determined differently depending on the type of difficulty. Individual decision trees will be averaged in a regression task, and a majority vote—i.e. the most frequent categorical variable—will produce the predicted class in a classification task. Finally, the oob sample is used for cross-validation, bringing the prediction to a conclusion.

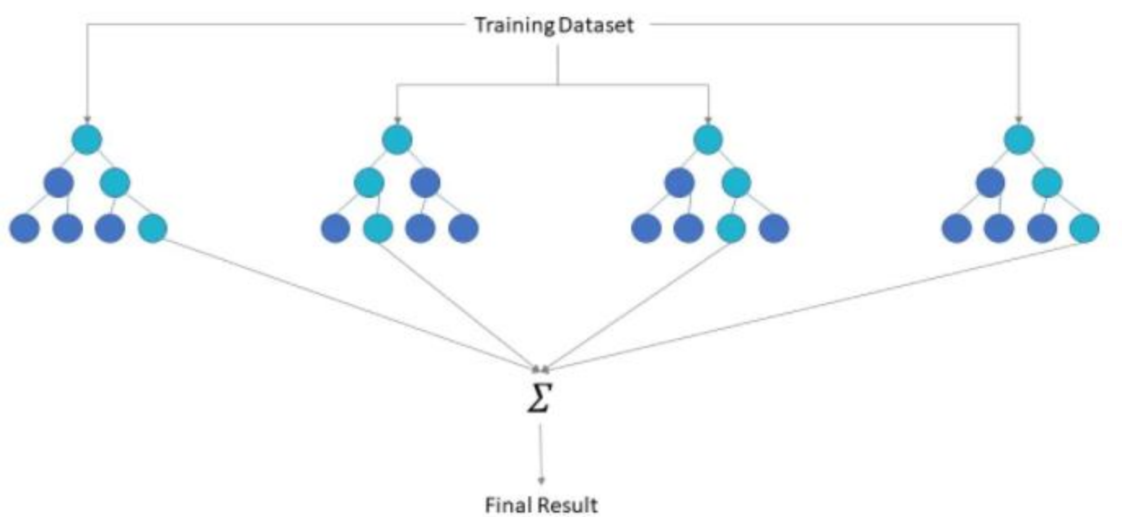


Fig. Random forest

Logistic regression:

The supervised learning classification method logistic regression is used to predict the likelihood of a target variable. Because the nature of the goal or dependent variable is dichotomous, there are only two classifications.

In this study we used Multinomial logistic regression for the used dataset.
Multinomial Logistic regression is an extended linear model that uses a collection of explanatory variables X to estimate the probability for the m categories of a qualitative dependent variable Y.

$$\Pr(Y_i = 1) = \Pr(Y_i = K)e^{\boldsymbol{\beta}_1 \cdot \mathbf{X}_i}$$
$$\Pr(Y_i = 2) = \Pr(Y_i = K)e^{\boldsymbol{\beta}_2 \cdot \mathbf{X}_i}$$
$$\cdots\cdots$$
$$\Pr(Y_i = K-1) = \Pr(Y_i = K)e^{\boldsymbol{\beta}_{K-1} \cdot \mathbf{X}_i}$$

Fig. Multinomial Logistic regression

Support Vector Machine:

The "Support Vector Machine" (SVM) is a supervised machine learning technique that can solve classification and regression problems. It is, however, mostly employed to solve categorization difficulties. Each data item is plotted as a point in n-dimensional space (where n is the number of features you have), with the value of each feature being the value of a certain coordinate in the SVM algorithm. Then we accomplish classification by locating the hyper-plane that clearly distinguishes the two classes.

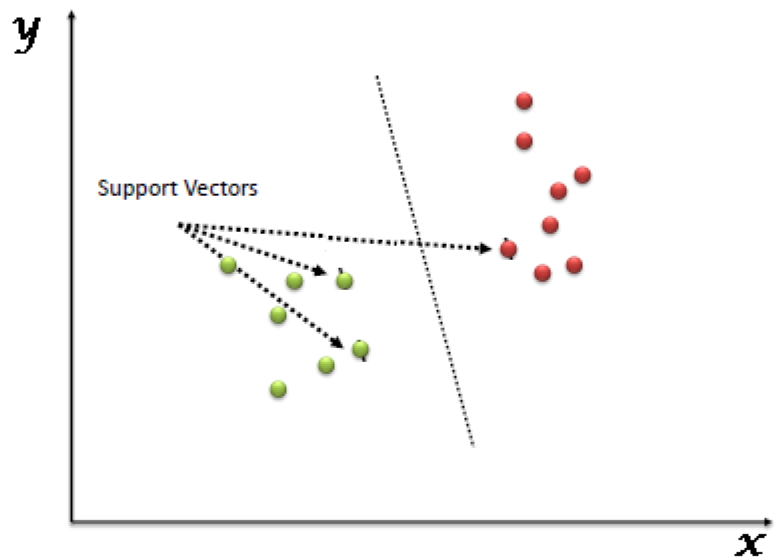


Fig Support Vector Machine

Making a straight line between two classes is how a simple linear SVM classifier works. That is, all of the data points on one side of the line will be assigned to one category, while the data points on the other side will be assigned to a different category. This implies that there may be an endless number of lines from which to pick.

K-Nearest Neighbour:

K Nearest Neighbour is a supervised machine learning method. It is commonly used in real-world contexts because it is non-parametric, which means it makes no underlying assumptions regarding data distribution. Which is depended on the input set of data and makes predictions by storing the training dataset and following the similarities between training and test set.

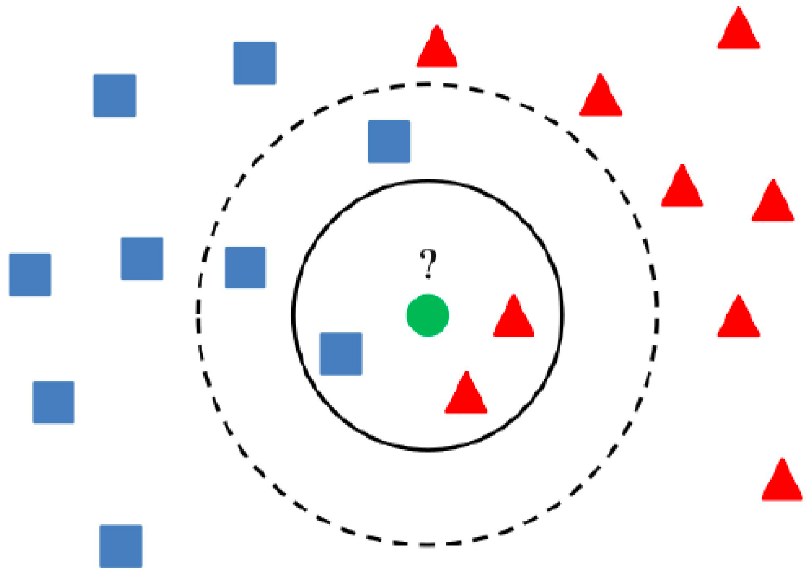


Fig. K Nearest Neighbour

We might be able to find some clusters or groupings if we plot these points on a graph. We may now assign an unclassified point to a group by looking at which group its nearest neighbors belong to. This indicates that a point near to a cluster of 'Red' points has a greater chance of being classed as 'Red.'

That is how when training features are stored, KNN scans the factors which are related to the outcome. Then when a similar problem from the test set enters the algorithm KNN tries to find the related factor from the test set and makes a prediction.

Chapter 4
Results and Discussions

In this topic we will discuss the result of this research and answer the research questions.
As it was mentioned in the above five different classification algorithms were used to predict the outcome of fetal state of a newborn baby or the fetus residing in the womb. A total of 2126 rows of datas were collected in this dataset and they were clinically approved and the outcome was determined by three obstetricians.

To see in detailed form whether the data is well balanced or not we need to go check the outcome which is 'fetal_state' in this case to ensure that all the classes of data are balanced with one another.

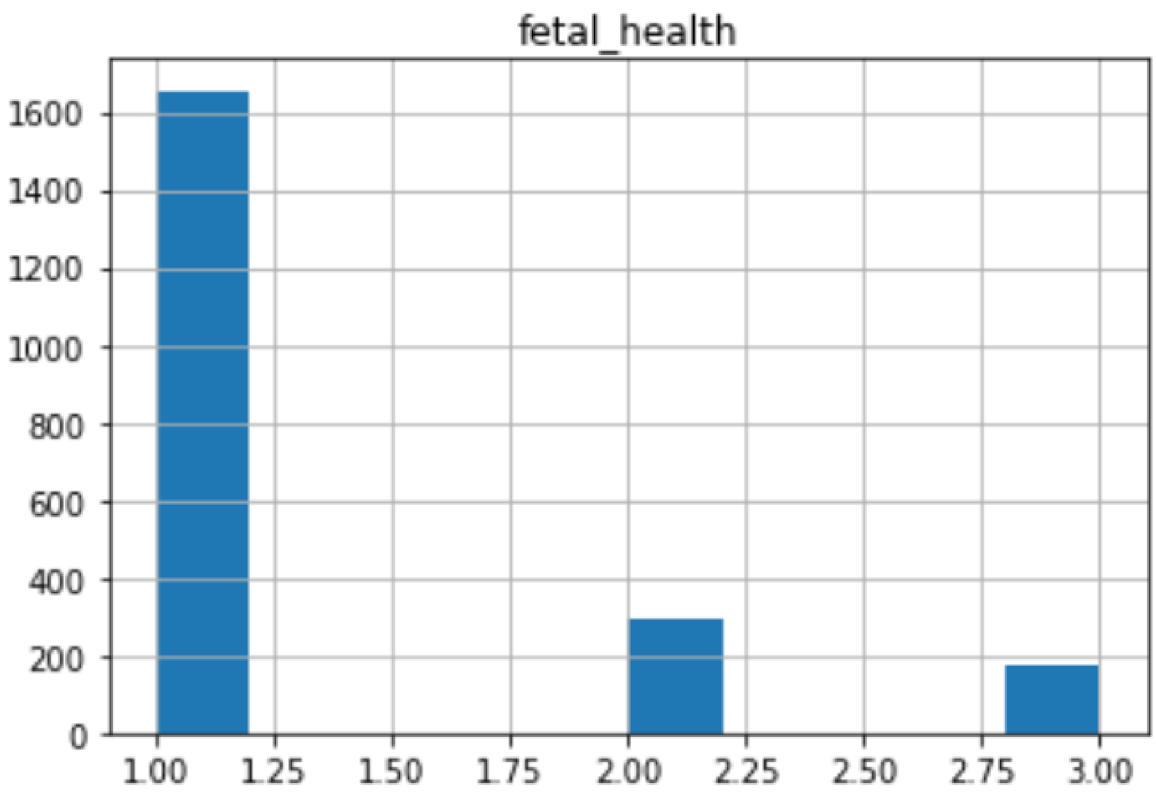


Fig. All classes of fetal outcome

In the histogram we see that three bars of three classes are visible which indicates three fetal outcomes of the dataset. They are classified in numerical forms in 1s, 2s and 3s which denote the normal, suspicious and pathological outcome of the fetus respectively.
But as we can see the outcomes of the above dataset is not even close to being balanced. The normal state containing the denotation '1' is almost more than 1600 (i.e 1655 to be exact) whereas the amount of suspicious and fetal health state cases is 590 and 528. Therefore, they do not balance with the normal state class.

After we applied smote analysis on the training set of the data the data the balanced histogram was something like this.

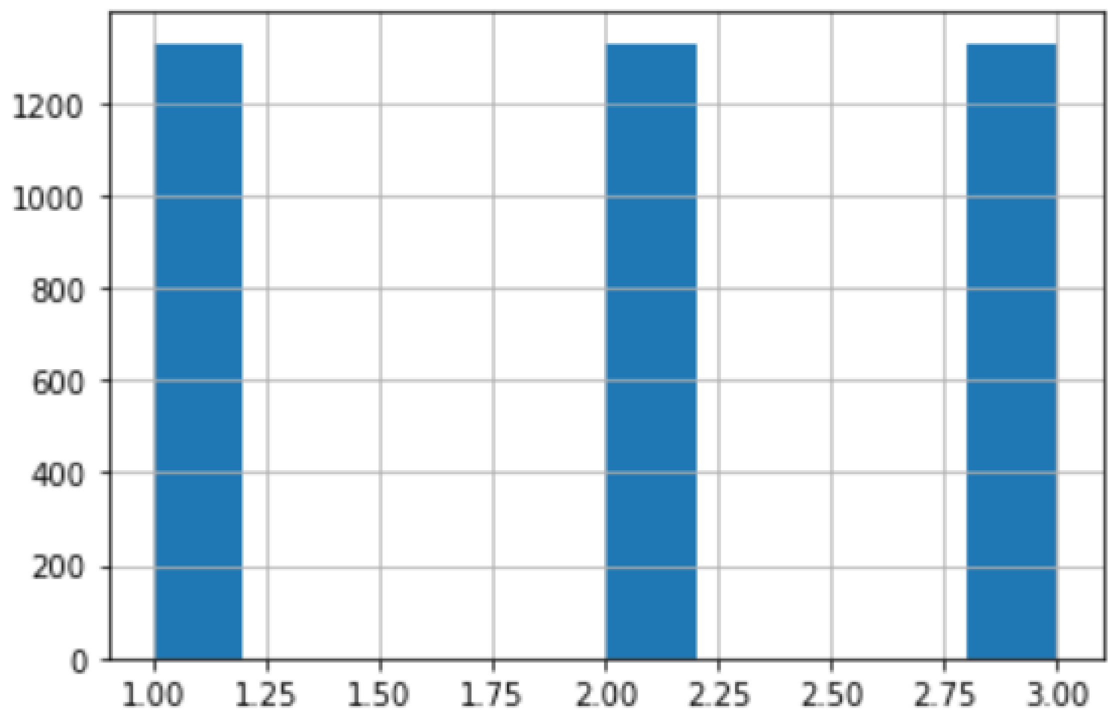


Fig. resampled fetal outcome

As it is visible that working on an outcome which is theoretically balanced and have the examplanitory outcome of all three states of our required classification is better and hope to give more of an unbiased result.

But as SMOTE analysis makes virtual rows we cannot analyze the numbers of the variation feature from the resample data set; the following graphs will demonstrate the key features in a visible pattern.

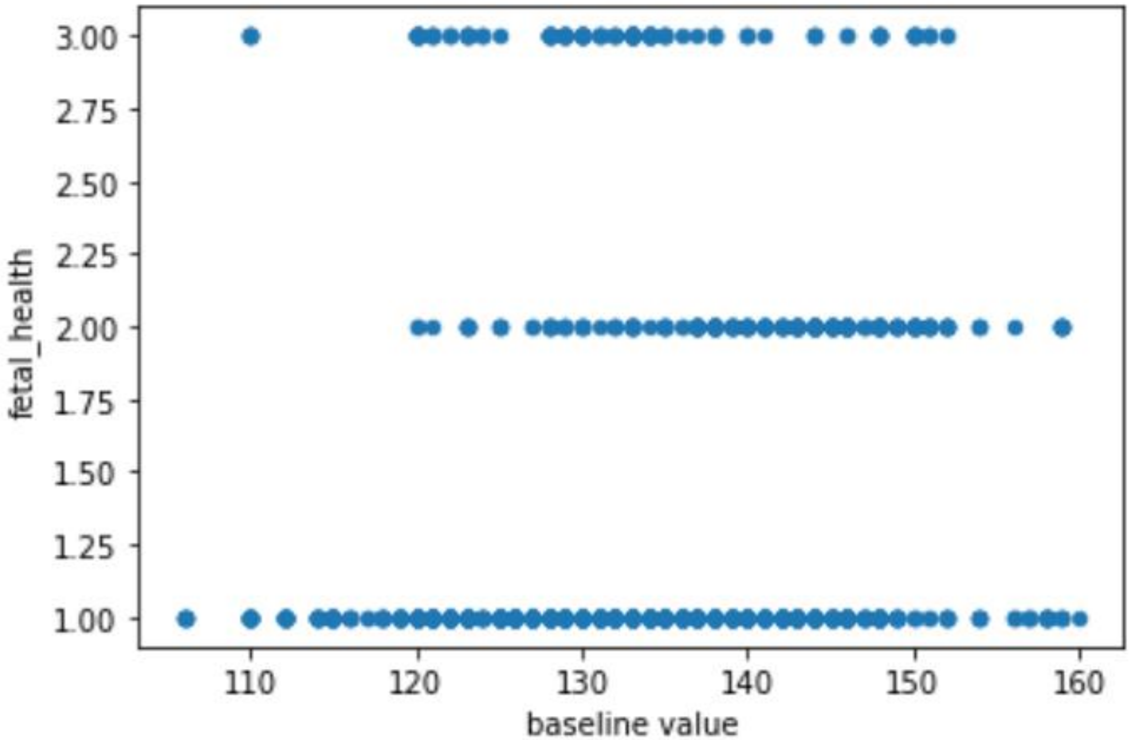


Fig. baseline vs fetal health scatter plot graph

As it was discussed before, baseline is one of the key components of determining fetal health of the fetus. In this dataset the baseline is known as the medium amount of heartbeat recorded per minute. From the figure we can determine that most of the normal state cases are at the middle of 120-160 bps. According to medical study if the heart beat goes close or exceeds below 120 bps and 160 bps then the situation is alarming.

Acceleration is the second key factor that helps to keep the fetus normal. The acceleration has some constant factor on the basis of which the rise of heartbeat is called an acceleration. First the heartbeat should be more that 15 beats per minute over the baseline and needs to have 15 seconds to go down if either one of these factors are not true that that phenomenon will not be acknowledged as an acceleration rather it would be called as a variability. The linear graph of the acceleration and fetal health condition is shown below.

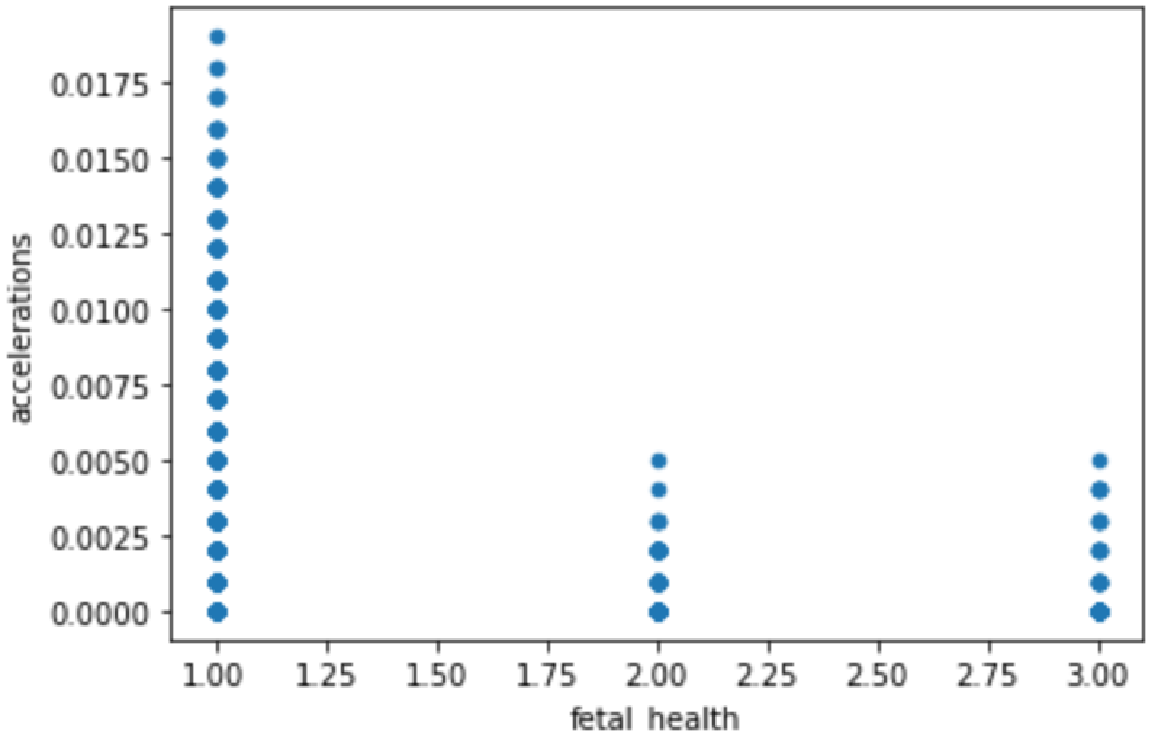


Fig. acceleration vs fetal health

According to the medical study it is known that the more the acceleration is the lesser is the chance of the fetus to be in an alarming situation. As we see in the graph a strong broad line of fetal health ‘1’ is visible, that is where most of the normal state lies at the highest peak of acceleration quantity.

And for the deceleration the condition is the opposite. If we look at the prolonged deceleration in a graphical manner then we will be able to see the negative impact of deceleration on the fetus’s health in a CTG.

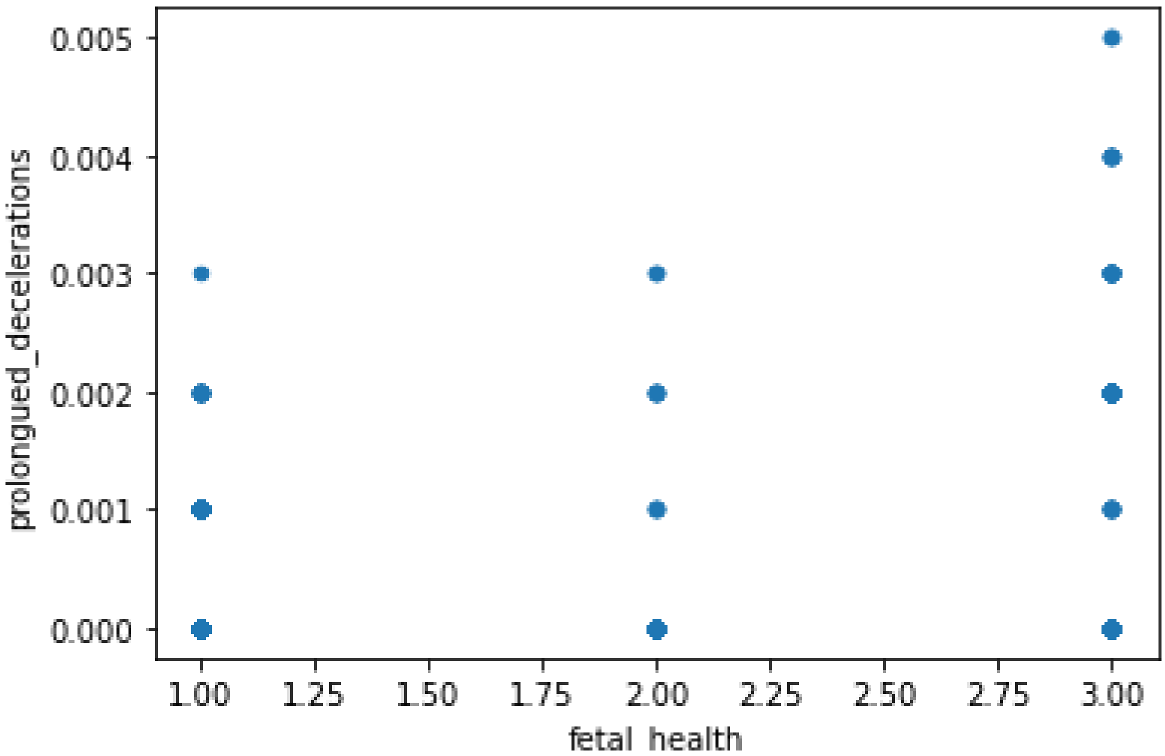
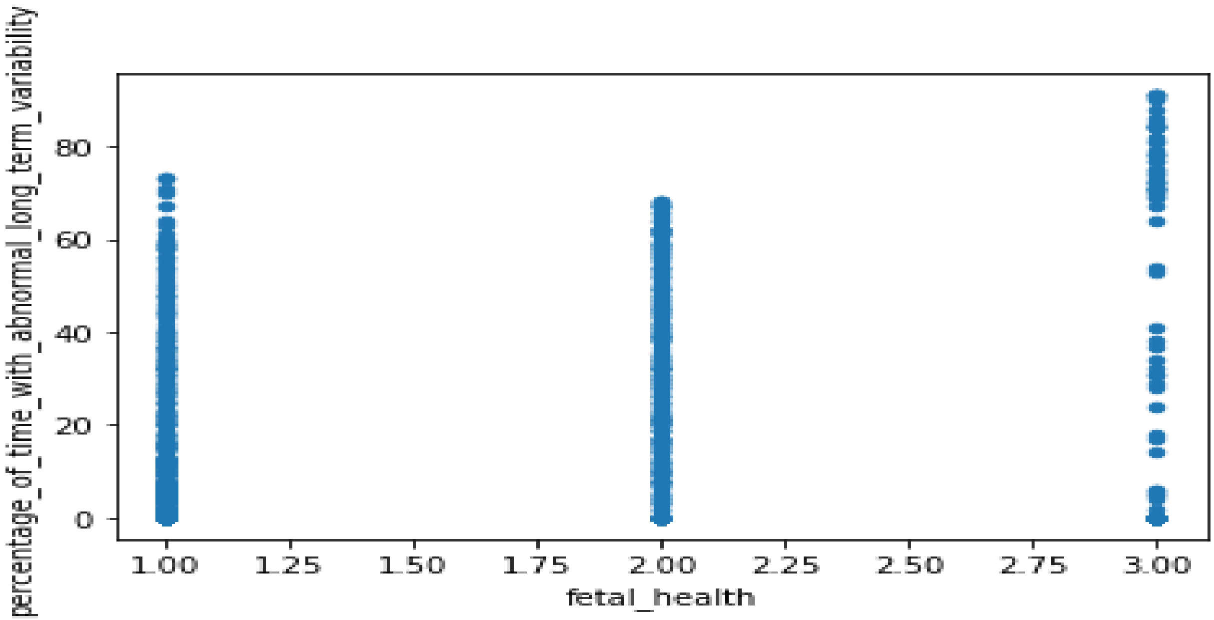


Fig. fetal health vs prolonged deceleration

Prolonged deceleration is the deceleration count from a longer period of time. From the graph it is seen that most of the state of pathological and suspicious state of the fetus occurs for the deceleration being too large in quantity.

Then lastly, variability of a fetus's heartbeat is the most important factor of determining the fetal state. Variability is somewhat like both acceleration and deceleration; it shows the variance of histogram of the heartbeat of a fetus; the middle line of the variability is known as the baseline value that was discussed above.



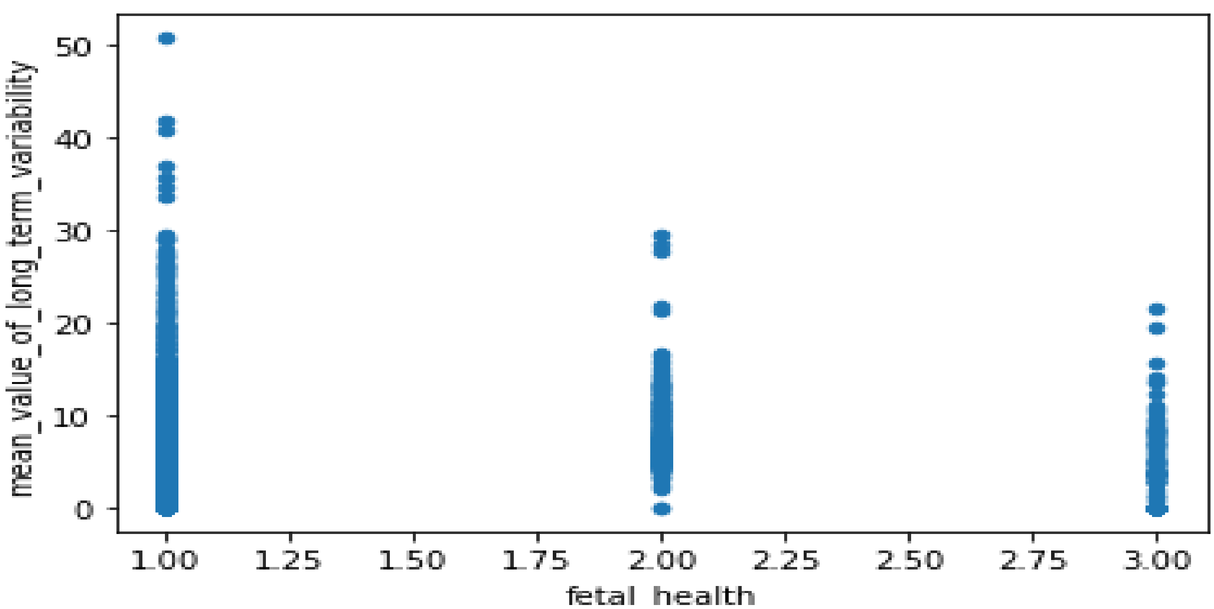


Fig. variability vs fetal health graph

The above figure shows the variability with respect to the fetal health in graphical visualization. Variability is normal for a fetus as it grows. A fetus has a larger quantity of heartbeat than a normal human being the heartbeat can sometimes exceed upto 200. But variability of the heartbeat generally shows the normal state of a fetus, that is only when the variable period is for a short amount of time, that is what it shows in the first graph.
It is mandatory for the fetus to have acceleration or deceleration in their heartbeat, but the alarming situation is when it goes on for a longer period of time.
We know that acceleration or deceleration comes down within 15 seconds from when it starts to happen the beat per second should be at least 15 above baseline, when it is lesser than that is known as a variability of the heartbeat. When it is more that 15 beats per second then it takes longer time to get into the base value and stays for a larger amount of time we denote that as a long term variability of the fetus. That is shown in the 2nd graph.

Therefore that was the overall visualization and definition of the dataset, as it was mentioned in the discussion the whole study is based on three fetal outcomes and the data is classified into several categories. The factors on which these outcomes are dependent are of abundance, so different techniques can easily fit to make a prediction outcome through machine learning.

Results:

At first from the feature selection we came to know that which features amongst this dataset were correlated with one another

```
In [19]: def correlation(dataset, threshold):
             col_corr = set()  # Set of all the names of correlated columns
             corr_matrix = dataset.corr()
             for i in range(len(corr_matrix.columns)):
                 for j in range(i):
                     if (corr_matrix.iloc[i, j]) > threshold: # we are interested in absolute coeff value
                         colname = corr_matrix.columns[i]  # getting the name of column
                         col_corr.add(colname)
             return col_corr

In [20]: corr_features = correlation(df, 0.85)
         len(set(corr_features))

Out[20]: 2

In [21]: corr_features

Out[21]: {'histogram_mean', 'histogram_median'}
```

The result showed that two of these features almost had the same outcome. The scaling was the same for the mean and median of the histogram of the dataset. So it was ok to leave either one of the features out of the dataset.

```
df=df.drop(['histogram_median'], axis=1)
```

Therefore the median was removed from the dataset.

From the machine learning models the most successful accuracy that we have got is 89.20% which is after applying Random forest classifier. Although, the acquired results of all five models are given below.

Table. Training Accuracy

Model Name	**Accuracy**
Random Forest	90.3%
Decision Tree	81.21%
Logistic Regression	80.03%
K nearest Neighbour	96.79%
Support vector Machine	99.84%

The training score of the data spilt of 80-20 is shown here expect of that of the KNN model because of it’s capability of storing the data inorder to train them.

Table. Testing Accuracy

Model Name	**Accuracy**
Random Forest	89.2%
Decision Tree	87.79%
Logistic Regression	77.23%
K-Nearest Neighbour	85.68%
Support vector Machine	82.15%

In the above table the decision tree could have done a better job but ensemble classifiers such as Random forest can out match the decision tree in performance. Krupa et al. [9] suggested a framework for utilizing EMD's statistical properties. They claimed that the test data was 86 percent accurate. Georgoulas et al. [6] suggested an SVM classification strategy for removing noise from FHR recordings by lowering the data dimension using PCA. They had a 75.61 percent overall classification accuracy and an AUC of 0.78.

Several automated methods with varied degrees of accuracy have therefore been created to aid in the analysis of CTG data; however, none of them has been generally used. Sundar C [10]one explanation for this might be that automated systems like SisPorto software, which can assist in clinical decision-making, are only compatible with particular equipment brands. Ayres-de C[11] This limits their general applicability, particularly in low-resource nations with limited access to technologically advanced devices. Vendor independence may be achieved using a machine learning model like the one described in this paper, but it would necessitate open interfacing of vendor-specific raw data with the machine learning algorithm.
After excluding features by pearson correlation and using SMOTE balance technique the study yielded better results than many other researches.

Now to look back on the previous result of the references or other previous research.

As it was discussed in [2] the Random Forest classifier through bagging ensemble technique has 99.02% accuracy where in this study Random forest with Smote Analysis and feature selection yielded 89.2% accuracy. Then again [22] SVM has accuracy of 99% which is said to be the hybrid K-Means SVM and the traditional SVM has the accuracy of 76% but in this research with the traditional SVM technique the accuracy which was obtained was 82.15%.where in this study we gained upto 82.15%. KNN algorithm scored 83.54% with feature selection[3]. And here with SMOTE analysis and feature selection respect to correlation accuracy gained was 85.68%. And lastly logistic regression in [23] had a score of 98.74% whereas this study got an accuracy of 77.23%.

<u>Addressing Research Questions</u>:

A01: The balencement of the dataset by SMOTE balancing technique the data set was balanced with a balanced set of outcomes in each class categorical data in the dataset.

A02: As for the Result two of the models successfully got us a more improved result which was discussed above.

Chapter 5
Conclusions and Findings

Findings:

Many similar studies were accomplished before this in order to improve the performance of medical technology. Though this study gave us an understandable result, the result is still unimproved and some accuracy is lesser compared to other kinds of techniques. One of the major backlash of this research is that there is no information about any specific patient in the dataset. Because of this the model can run the data but will ignore any external factors related to the patient. The machine learning system may show a varied accuracy in LMICs due to changes in sociodemographic features of pregnant women.

Contributions:

The major contributions of this researches shown below:

Good understanding of the categorical data of CTG readings and excluding the features after analyzing correlations.

Accuracy of the fetal risk through 5 different models from the balanced dataset and selected features.

Recommendations for future work:

The main focus of this study was to determine the accuracy of fetal risk prediction. IT was meant to make an improvement with the taken dataset and the selected features and somewhat it did. As it was mentioned Random Forest was the highest acquired accuracy. But Subasi, A.[2] made a better prediction using the model but that data was pre processed differently. Therefore, the research needs much work of refining and evaluating the outcome. By stating that the CTG dataset should be collected and validated by paying heed to the environment, because as it was mentioned above about the “changes in sociodemographic features of pregnant women”.

References

1. [Chow O Larbig A Kabartas B et al.Pitfalls of external FHR-monitoring in the second stage of labor *J Mat Fet Neonat Med* 201023(Suppl.)236]
2. 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.
3. 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.
4. 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.
5. Intan, P. R. D., Ma'sum, M. A., Alfiany, N., Jatmiko, W., Kekalih, A., & Bustamam, A. (2019, October). Ensemble learning versus deep learning for Hypoxia detection in CTG signal. In *2019 International Workshop on Big Data and Information Security (IWBIS)* (pp. 57-62). IEEE.
6. Georgoulas, G., Stylios, D., & Groumpos, P. (2006). Predicting the risk of metabolic acidosis for newborns based on fetal heart rate signal classification using support vector machines. *IEEE Transactions on biomedical engineering*, *53*(5), 875-884.
7. Gonçalves, H., Rocha, A. P., Ayres-de-Campos, D., & Bernardes, J. (2006). Linear and nonlinear fetal heart rate analysis of normal and acidemic fetuses in the minutes preceding delivery. *Medical and Biological Engineering and Computing*, *44*(10), 847-855.
8. Chamidah, N., & Wasito, I. (2015, October). Fetal state classification from cardiotocography based on feature extraction using hybrid K-Means and support vector machine. In 2015 international conference on advanced computer science and information systems (ICACSIS) (pp. 37-41). IEEE.
9. N. Krupa, M. Ali, E. Zahedi, S. Ahmed, and F. M. Hassan, "Antepartum fetal heart rate feature extraction and classification using empirical mode decomposition and support vector machine," Biomed. Eng. Online, vol. 10, no. 1, p. 6, 2011.
10. Sundar C, Chitradevi M, Geetharamani G. Classification of cardiotocogram data using neural network based machine learning technique. Int J Comput Appl 2012;47:14.
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. Rohani A, Taki M, Abdollahpour M. A novel soft computing model (Gaussian process regression with K-fold cross validation) for daily and monthly solar radiation forecasting (Part: I). Renew Energy 2018;115:411-22.
13. Dawson AJ, Buchan J, Duffield C, Homer CS, Wijewardena K. Task shifting and sharing in maternal and reproductive health in low-income countries: A narrative synthesis of current evidence. Health Policy Plan 2014;29:396-408.
14. Chandraharan E, Arulkumaran S. Prevention of birth asphyxia: Responding appropriately to cardiotocograph (CTG) traces. Best Pract Res Clin Obstet Gynaecol 2007;21:609-24.
15. K. Bache and M. Lichman, “Cardiotocography data set,” UCI Mach. Learn. Repos., 2010.
16. Costa MA, Ayres-de-Campos D, Machado AP, Santos CC, Bernardes J. Comparison of a computer system evaluation of intrapartum cardiotocographic events and a consensus of clinicians. J Perinat Med 2010;38:191-5.
17. R. Czabanski, J. Jezewski, A. Matonia, and M. Jezewski, “Computerized analysis of fetal heart rate signals as the predictor of neonatal acidemia,” Expert Syst. Appl., vol. 39, no. 15, pp. 11846–11860, 2012.
18. Grivell RM, Alfirevic Z, Gyte GM, Devane D. Antenatal cardiotocography for fetal assessment. Cochrane Libr 2015;9:CD007863.
19. Liu L, Oza S, Hogan D, Chu Y, Perin J, Zhu J, et al. Global, regional, and national causes of under-5 mortality in 2000-15: An updated systematic analysis with implications for the sustainable development goals. Lancet 2016;388:3027-35.
20. R. A. Johnson and D. W. Wichern, Applied multivariate statistical analysis, vol. 5. Prentice hall Upper Saddle River, NJ, 2002.
21. G.G. Georgoulas, C.D. Stylios, G. Nokas, P.P. Groumpos, Classification of fetal heart rate during labour using hidden Markov models, in: Neural Networks, 2004, Proceedings of IEEE International Joint Conference, vol. 2473, 2004, pp. 2471–2475
22. Cömert, Z., & Kocamaz, A. F. (2017). Comparison of machine learning techniques for fetal heart rate classification. Acta Physica Polonica A, 132(3), 451-454.
23. 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