Master Degree Project

UNIVERSITY OF SKÖVDE

C-section Birth Data Classification Using Ensemble Modelling Techniques And Their Performance Analysis

Maria Yousaf

Supervisor: Anton Kalen.
Examiner: Gunnar Mathiason.

Master Degree Thesis (60 ECTS) in Informatics
with a specialization in Data Science/
Privacy, Information and Cyber Security
15 ECTS
Spring term 2022

ABSTRACT

Data mining and machine learning techniques have a wide range of applications in businesses, healthcare, organizations, and academia, to name a few. Machine learning has been used by several academics to construct decision support systems, analyse major clinical features, extract useful information from trends in historical data, generate predictions, and classify diseases. Successful research gave doctors the ability to make the best decisions at the correct moment. We plan to use the learning potential of machine learning methods to classify birth data utilizing bagging, boosting, and stacking classification algorithms in the current work. Diversity in living styles, medical aid, religious connotations, and the place you live in all have an impact on the people who live in that culture. The current study is a complete comparison of the bagging, boosting, and stacking classification algorithms used on the government hospital's birth data. The caret library in R, which is regarded as an encompassing framework for developing machine learning models, is used for the experiments. Different evaluation measures are used to offer accuracy-based results. Boosting features with regard to, sensitivity, and specificity, the Gradient Boosting Machine (GBM) performed somewhat better.

Table of Contents

1. Introduction

A Caesarean section, sometimes known as a C-section, is a delivery method in which the intended child is delivered through a surgical operation that involves cutting through the mother's belly. Multiple infants in the uterus, a baby in the transverse position or in the breach, past C-section deliveries, a large infant, and so on all necessitate a C-section. If vaginal delivery puts the life or health of the pregnant mother or child in jeopardy, doctors consider C-sections. The reason for this study is that the obstructed labor and other emergency obstetrical situations need cesarean delivery, which saves lives and lowers child and mother mortality. However, because it is a surgical treatment, there are risks of complications, and excessive use can harm both mothers and newborns [1].

Cesarean birth is becoming a big public health issue as it grows more common. The worldwide cesarean rate reached 21.1% in 2018 and it has been steadily rising since 1990 when it was 6.7%[2]. In the previous two decades, Pakistan has become one of the countries with the highest rate of cesarean births [3]. Pakistan's infant mortality rate was 58.46 per 1,000 live births in 2020 [4]. According to the survey data [5], maternal causes accounted for 12% of fatalities associated with an increase in married women between the ages of 15 and 49 in the previous three years. Aside from this statistic, the number of deliveries made in private clinics, residences, and other settings is much higher. C-sections during childbirth have been linked to a higher risk of serious maternal and newborn complications[6]. In order to prevent maternal and newborn morbidity, predicting women in need of Cesarean (C-Section) is beneficial in obstetric management. The precision of this subject should be improved so that the kind of birth, i.e., normal(vaginal) or Cesarean(C-Section), is known in advance. Pregnant women are thought to have a similar social life and go through various similar pregnancy situations and medical conditions during their pregnancy. If these situations and medical conditions aspects of expecting mothers are thoroughly collected and documented, they can be useful in prediction-based studies [7].

Knowledge Engineering collects data in order to extract relevant information, make decisions, predictions, or analyze the data in some other way. Many decisions can be made with the use of knowledge and machine learning, which can be used in a variety of industries such as healthcare, education, manufacturing, business, image processing, and so on. In many circumstances in healthcare, machine learning may contribute significantly by assisting in decision-making based on previously acquired knowledge. Information technology applications such as prediction methodologies can also serve as a benchmark for achieving prediction in the event of a medical emergency.

Researchers are working in almost every area related to healthcare to determine relationships between different factors associated with that particular area. Furthermore, machine learning techniques can help find hidden patterns which can be beneficial for the research studies. In this research Data mining's application capabilities would aid in the development of intuitions from data by revealing connected patterns, while machine learning techniques such as Bagging, Boosting, and Stacking would allow us to build the perfect classification model for predicting the birth outcome.

1.1 Problem Definition

According to the World Health Organization, 810 women die every day in the world as a result of childbirth and pregnancy-related problems, with low and lower-middle-income nations accounting for 94 percent of all maternal deaths[8]. The prevalence of maternal mortality is decreasing as a result of recent technological advancements[9], [10], but it remains a difficult effort to assure the health of both mother and the child throughout pregnancy. Predicting difficulties and implementing preventive actions can help reduce pregnancy-related risks in this situation. As a result, predictive modeling became popular in saving the lives of thousands of mothers and babies.

This study focused on women's health issues, delivery preferences, diseases that affect the health of pregnant women, and factors that influence the decision to have a C-section in Pakistani women. During their gestational period, pregnant women with comparable social lives are thought to go through various related pregnancy symptoms and medical issues. The relationship involving maternal age and mode of birth is revealed to be an important feature among the socio-demographic variables of mothers[11]. Women with chronic health problems (such as persistent hypertension, cardiac illness, lung problems, or other health risk factors) are much more likely to have a cesarean section, according to research[12]. If these behaviours and medical conditions of expecting mothers are thoroughly collected and documented, they can be useful in prediction-based investigations. As a result, it's critical to perform studies that can aid in the discovery of physical or related aspects. This has the potential to be beneficial in two ways. For instance, such information may be used to make predictions regarding the current subject of study. Second, the expected outcomes may assist physicians in taking early action by addressing the conditions that are contributing to the situation that necessitates a C-section.

1.2 Aim and Objectives

The goal of this study was to evaluate the use of maternal factors in developing a machine learning model which suggests the best classification for unplanned C-sections in terms of accuracy and to classify locally gathered birth data by applying the ensemble classifiers. A set of different ensemble classifiers are used (bagging, boosting, and stacking) to analyse which technique have higher accuracy for this particular dataset.

The research objectives for this study are as follows:

- Construct a decision system for the accurate prediction of delivery methods.
- To analyse the results of various machine learning algorithms[1] that are commonly used in research literatures[2] to predict the birth mode.
- To see how various machine learning approaches perform on locally collected birth data.

The obtained measurements are referred to by the state-of-the-art research publications on examination of the application area and algorithms. In order to explain the problem, analyse the performance of selected models, and conclude results, this thesis comprises various chapters where Chapter 1 gives an overview of the area of study as well as the study plan. Chapter 2 explores more related studies in the subject area. The data and algorithms used in the study are thoroughly described in Chapter 3. The study's implementation is detailed in Chapter 4. The results and related discussion in Chapter 5 are backed by various evaluation measures. The overview and discussion are found in Chapter 6. Future directions are given in Chapter 7, followed by references.

---

[1] Ensemble Algorithm: Models of Bagging, Boosting, and Stacking details described in Chapter 2.
[2] Reference to chapter 2 Section2.1.

2 Background

2.1 Preliminaries

Machine learning techniques have been utilized in several types of research in the healthcare field to predict cesarean deliveries. Soleimanian et al. [13] employed a decision tree (DT) classifier as a prediction model, for example. Because of its capacity to generate multiple trees with varied methods and excellent accuracy in diagnosis, the authors [13] employed an extension of Quinlan's induction decision tree (ID3) [14], which is a C4.5 algorithm. They used a pregnancy dataset from the Tabriz health centre to create a model that had a 95% accuracy rate (86.25 percent). The tree formed has a high level of complexity, with a depth of 31 and 21 leaves. Large trees are expensive and can lead to inaccurate assumptions [15]. To increase accuracy, the authors suggested expanding the dataset and incorporating more related variables. Similarly, [16] employed the same dataset as [13] but used other predictive models, Support Vector Machine, kNN, Random Forest, NB, and logistic regression (LR)—resulting in 76.3 percent, 95 percent, 95 percent, and 77.5 percent accuracy, respectively. The best performers were RF and kNN, according to the results. Similarly, [7], [17] used ensemble techniques like Treebag, Adaboost, Gradient Boost, FDABag, XGBDart, XGBLinear, Support Vector Machine (SVM), Random Forest (RF) to do a performance analysis using birth classification data of cesarean sections and vaginal delivery. Another study [18] employed the C4.5 algorithm to estimate the probability of preterm birth during pregnancy. The C4.5 algorithm was used on both standardized and non-standardized pregnancy data by the authors. According to the study, the accuracy percentage for standardized and unstandardized data was 71.30 percent and 66.08 percent, respectively.

The aforementioned models are the most commonly used models by researchers in research related to birth data [13]- [18]. Some of these models produced results with high accuracy (<85%) for their datasets and a few of them used duplication of data to enlarge the datasets and obtained higher accuracy but their research limitations were that their data was not 100% original. This study used these models on original data (not duplicated) and analyse the performance of these models.

2.1.1 Ensemble Methods

“Unity is strength”. This old adage captures the underlying concept that underpins machine learning's extremely strong "ensemble methods"[19]. Simple decision trees, gradient boosting machines, RF, deep neural networks, and SVM are some of the modelling techniques used by Machine Learning practitioners[20]. Each approach has its own set of advantages and disadvantages, which is why it's common to combine Machine Learning algorithms to achieve even better-predicted results than any single Machine Learning method could achieve on its own. Ensemble is a technique for combining algorithms. We intend to use machine learning ensemble methods to do classification for imbalanced data[21]. To solve the classification task, we plan to use three different types of ensembles [22]–[25]: bagging, boosting, and stacking. Ensemble methods are those that have a group of weak learners working together to develop a strong learner. This quick learner is the potential of exceeding expectations.

2.1.2 Bagging

Bootstrap Aggregating is another name for Bagging[22], [24], [25]. Focusing on the rear end of bagging is based on combining the results of several models to provide a more generalized outcome. Someone might wonder if developing all of the algorithms on the same data and then merging them is a beneficial activity. Because all of these models receive the same input, the likelihood of them producing the same outcome increases. How can this issue be resolved? Bootstrapping is one of its answers. Bootstrapping is a sampling strategy that creates subsets of specified observations from the original dataset with replacements. The size of the subset remains the same as the original. These bags (subsets) are used in the Bagging/Bootstrap Aggregating process to gain a realistic understanding of the data distribution of the entire set. The total number of bags made with this procedure may be lower than expected.
The following is how bagging works:

- The original dataset is separated into bags (subsets), and the observations are replaced.
- Each bag uses the same base model. This is a flawed base model.
- All of the models are operating in parallel and are independent of one another.
- The final outcome, or projections, is acquired by connecting or combining all of the models' predictions.

In Figure 2.1, you can see a general view of Bagging.

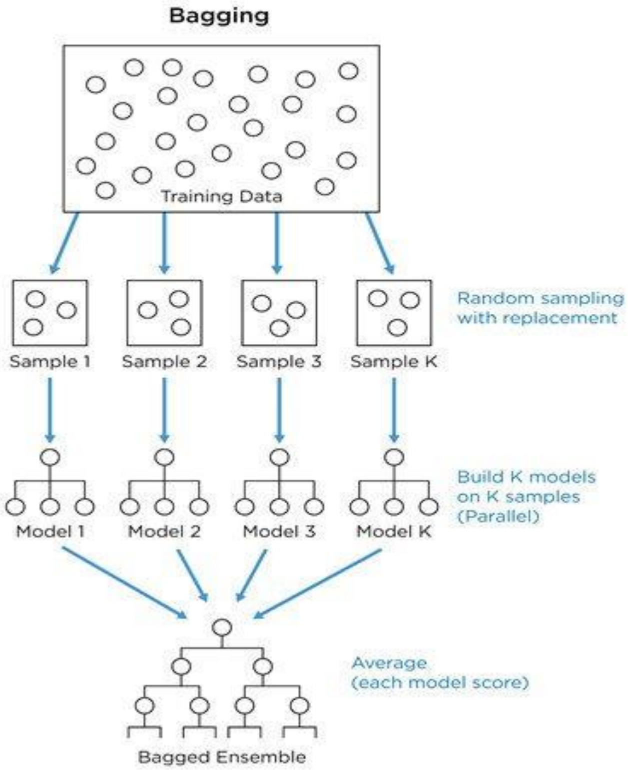

Figure 2.1: General working of Bagging from [20].

The following is the error (Bias) / variance decomposition:

Error (E) = Variance (V) + Bias (B) + Noise (N)

*The purpose of bagging is to lower the Variance (V).

Advantages

- It is simple to implement.
- Reduce overfitting in the model.
- It maintains the precision of missing data.
- Bagging allows you to easily handle and manage data with several dimensions.

Disadvantages

- For classification, a 100 % accuracy of predictions is not achieved because the final prediction is dependent on the mean prediction.

As the 100% accuracy often means overfitting a model, the bagging technique mitigates the chances of overfitting which leads to wrong predictions.

Although the distinction is frequently emphasized in clinical literature, validation is obviously important. Even when classification is good, inaccurate predictions can result in improper decisions or choices[26]. Bagging is used in many healthcare applications[27] despite the disadvantage of not having 100% accuracy.
The following are the bagging classifiers that were employed in the study.

1. Bagged Decision Tree

   The bagged CART model is known as Treebag. This is a Bagging model in essence. It resolves classification issues. None of the tuning settings are required for this procedure. It can be executed using the 'plyr' package.

2. Random Forest

   Random Forest[28] is a collection of decision trees. It creates and mixes numerous decision trees to improve prediction accuracy. It's a classification algorithm that isn't linear. When utilized alone, each decision tree is used. It is denoted as 'rf'.

2.1.3 Boosting

This is a sequential prediction procedure in which each prior model serves as the foundation for the next model's predictions. In order to accurately identify the data items, the next model in the sequence must fix the faults or satisfy the limitations of the previous model. If an input is improperly classified, its weight is enhanced so that it can be classified correctly in the next iteration. It's easy to see how each subsequent model is dependent on the last. The weak learners are thus prepared to create a strong model.
The following is how boosting works:

- As in Bagging, the original dataset is partitioned into subgroups.
- All of the sample points are given equal weights at the start.
- The base model for boosting is then constructed.
- Predictions for the entire dataset are made using the same model.
- Errors are indicated by the difference between projected and actual values.
- Data items with incorrect predictions are given higher weights.
- A new model is generated next to the previous one, and the prediction method is applied to the dataset once more. The major goal here is to repair the errors from the prior model.
- A series of models are produced, each with the task of fixing the flaws of the previous model.

- Finally, the final model, which is the strong learner, is really the weighted average of all of the series' preceding models (weak learners).

As a result, a group of weak classifiers is linked together to form a strong learner. The single learner cannot be applied to the entire dataset. As a result, each method is doing its part to improve group performance.
Figure 2.2 depicts the operation of Boosting.

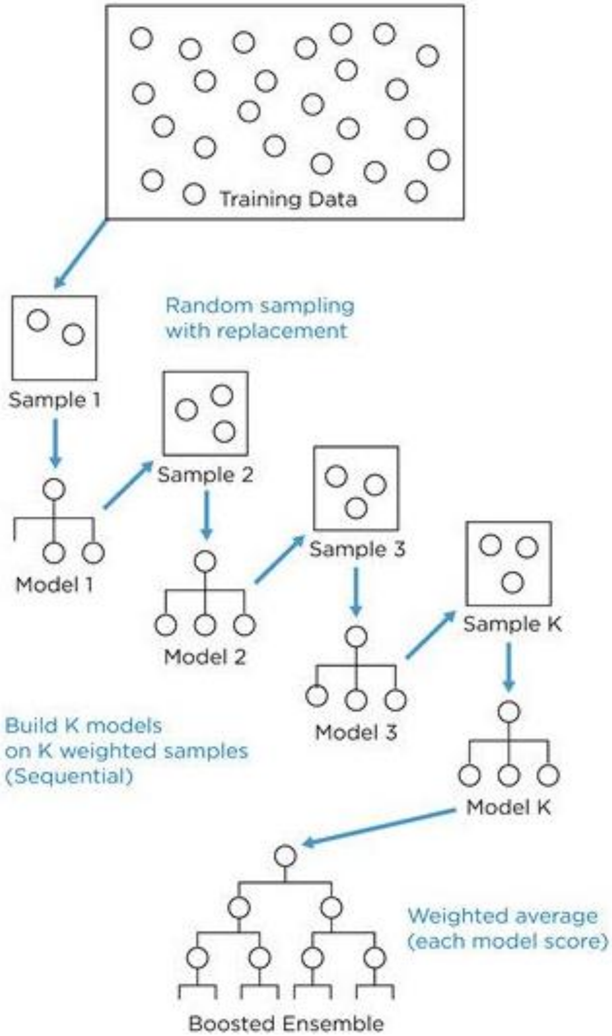

Figure 2.2: Working of Boosting Method from [20]

Advantages

- Boosting helps with a variety of loss functions.

- Works well when doing multiple jobs in a row.

Disadvantages

- Boosting can lead to over-fitting.

Although boosting can lead to boosting but the main causes of overfitting are unnormalized data, high variance, and uncleaned data. In this study, the data is cleaned from any unclear or missing values, and cross-validation is used to mitigate over-fitting. In healthcare, there are several areas[29]–[31] that are using boosting algorithms.
The following are the boosting classifiers that was employed in the study.

1. C50

   C50 is an R version of the decision tree-generating supervised classification algorithm C5.0. It is a successor of the C4.5 algorithm. By reducing the estimated entropy value, this approach employs an information entropy calculation to find the optimum rule for splitting the data at that node into purer classes[32].

2. Stochastic Gradient Boosting

   Freund and Schapire's AdaBoost algorithm and Friedman's gradient boosting machine are both implemented in the gbm R package[33].

2.1.4 Stacking

Stacking [34] is an ensemble learning approach that integrates the predictions made by numerous base classifiers created by different learning algorithms. The data used to train these classifiers is the same. Base classifiers predict things for the sample point in the first step. The meta-classifier then integrates the predictions given by the base classifiers to predict the final classifier in the second phase. The selection of a meta-classifier is a crucial aspect of stacking. In the past, comparative studies [35] were given to evaluate the performance of various learning algorithms as meta-classifiers. The most common learning technique used in stacking as a meta-classifier is logistic regression [36]. Figure 2.3 shows the general working of the stacking method.

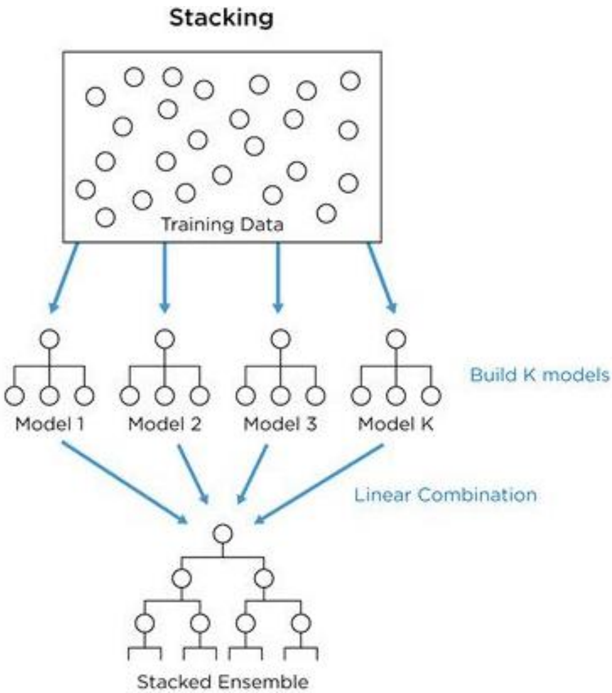

Figure 2.3: General working of Stacking from [20]

Advantages

- Combine the capabilities of a number of high-performing models to create predictions that outperform any single model in the ensemble on a classification or regression challenges.
- Model prediction accuracy is improved.

Disadvantage

1. Computing costs will be longer in the case of large datasets because each learner is operating individually on the entire dataset.

In this study, the dataset is small with only a little above 300 instances. Therefore, the high computation cost will not affect this study.

The following are the bagging classifiers that were employed in the study.

1. Logistic Regression:

   In essence, logistic regression is a classification algorithm. The word "regression" derives from its close relative, linear regression, in the regression domain. Because the classes in supervised classification issues are discrete, the algorithms' purpose is to determine the decision boundaries between them. Instances of one class are separated by decision boundaries. Decision boundaries can be complicated and nonlinear in geometric shape relying on the problem instance. Various machine learning algorithms make various assumptions about the geometry of decision boundaries in general. The assumption in logistic regression is that classification algorithm are linear[37].

2. RPart

   Recursive PARTitioning is the name of the R implementation. Rpart, like C50, employs a computation metric to find the optimum rule for splitting data into purer classes at that node[32].

3. KNN

   K-Nearest Neighbor (K-NN)[38] is another powerful classifier under the statistical learning umbrella, which assumes that examples in close proximity may have comparable attributes and that any unclassified instance's label can be deduced from its neighbors' labels. K-NN uses distance metrics such as Euclidean, Manhattan, Chebyshev, Minkowski, and others to reduce the distance between two similarly categorized examples while measuring the distance between occurrences of dissimilar classes.

4. LDA

   Linear discriminant analysis[39] is a technique for classifying, reducing dimensions, and visualizing data. It has been in existence for quite some time. Despite its simple structure, LDA frequently yields reliable, usable, and understandable classification results. When faced with actual classification problems, LDA is frequently used as a benchmarking tool before moving on to more intricate and flexible methods.

2.2 Research background

Different databases, including Google Scholar, IEEE explorer, ResearchGate, Wiley, Science Direct, and other data sources, were used to find studies linked to determining the mode of birthing. Major search terms included c-sections applying machine learning, Data Mining in Healthcare to Predict Cesarean Delivery, machine learning in maternity care, Artificial Intelligence in maternity care, and so on. The most relevant papers were found by searching various types of publications, such as journals, conference pieces, and open research articles.

Machine learning techniques have become increasingly prominent in a variety of computing fields. The number of academics interested in employing machine learning approaches to examine prognosis and diagnosis is steadily growing. This section discusses a few different sorts of research that are pertinent to the current topic.

The authors of [40] investigated the factors that influence neonatal and perinatal mortality. The newborn mortality rate was 31.40 per 1000 live births, while the prenatal mortality rate was 49.70 per 1000 pregnancies, according to the study. Infection was determined to be the leading cause of newborn death (43 percent). Cesarean section birth ($P = 0.049$), premature birth ($P = 0.003$), and twin pregnancy ($P = 0.001$) were all risk factors for death.

The authors[7] gathered data on births, C-sections, and regular deliveries from government hospitals in their home region, where pregnant women lacked adequate health infrastructure and amenities. They used bivariate analysis to identify the most important medical parameters linked to gestational anomalies and, as a result, the requirement for C-sections. They've also created a variety of machine learning-based birth classification algorithms.

Another study [41] in which the connection involving C-section deliveries and mother age was assessed by statistical analysis. The plan was to divide the pregnant women into different age groups. Expectant women under 35 years old, 35-39 years old, and over 40 years old were divided into three groups. According to statistics, women over the age of 40 are more susceptible to placenta abruption, placental Rupture, and C-section delivery. The findings also revealed that women aged 35 to 39 were more likely to have miscarriages and had chromosomal abnormalities.

Another work by [42] looked into how feature selection affected the effectiveness of naive Bayes for fetal heart rate variations and states. Correlation-based, ReliefF, Entropy, and Correlation were among the four feature selection approaches used. ReliefF performed better when it came to fetal state classification, but there was no discernible effect of feature selection approaches on fetal heart rate classification.

Another study [43] discovered that low birth weight is linked to parents with untreated celiac disease. To compare proportions and averages of normally distributed variables, they used the Pearson 2 test and students' t-tests. For the analyses of differences in the birth week, birth length, and birth weight, they employed multiple linear regression. The adjusted odds ratios (AOR) of the categorical variables preterm birth, C-section, low birth weight, and neonatal hospital care were calculated using binary logistic regression.

Another classification-based investigation involving numerous machine learning classifiers was undertaken by [44]. Bega Obstetrics and Gynecology Center provided a total of 2325 birth records. The researchers wanted to see if there was a link between the concentration of glucose in the umbilical cord blood, the newborns' cry, the mother's BMI before pregnancy, and the Apgar score. Dataset values were also used to create several categorization models. A specific Weka API-based application was also created to use the Logit Boost algorithm to categorize birth outcomes.

The authors [45] investigated the link between C-sections and maternal age using multivariate logistic regression. According to the data, pregnant women aged 44 and up have a higher risk of medical difficulties, which increases the likelihood of a C-section birth. On the contrary, the rate of C-sections amongst pregnant women aged 22 to 29 is lower than that of older women.

A study on preterm prediction has been proposed by [46]. For finding the premature prediction, they used RF, which is a machine learning technique, and regression. According to the research, RF has the highest accuracy rate of 97 percent.

According to Pakistan Institute of Development Economics (PIDE) data from 2015, the mode of distribution is influenced by two factors: 1) wealth and 2) Educate yourself. According to their findings, 23 percent to 35 percent of highly skilled women prefer C-section birth. Women with low education and income are less likely to have a C-section.

Data mining methodologies are helpful in obtaining knowledge from data, according to [47] Machine learning is one of the most popular approaches among these. Clustering is a commonly utilized technique. On the given day, they performed K-mean clustering and discovered that this method of clustering leads to better outcomes for their presented problem.

According to [48], categorization is essentially a sorting procedure of instances based on their feature values.

Research [49] suggested a new data classification technique. Kernel Fisher Discriminant was the proposed classification technique (KFD). They claimed that KFD is more effective than SVM.

Another review was conducted by researchers [50]. The aim of this review is to look at research that uses machine learning (ML) to predict the best mode of labor and detect various issues during childbirth. A total of 26 publications (published between 2000 and 2020) were chosen and reviewed using a Systematic Literature Review (SLR) method. As a result of this review study, the objectives or focuses of recent studies on birth outcomes using ML were highlighted; the adopted ML algorithms were explored, as well as their performances; and a synthesized perspective of characteristics used, attributes, data sources, and their characteristics were provided. The study also identified areas for future research to support existing activities aimed at lowering maternal mortality and complacency rates. The outcomes of this systematic review helped to form a theoretical foundation for the advancement of ML-based maternity healthcare. The review will present a cutting-edge paradigm for clinical decision-making, predicting pregnancy issues and delivery style, as well as medical diagnosis and therapy.

This research[51] created an advanced machine learning algorithm to predict the need for an emergency c - section surgery before labor begins. The study comprised a total of 6549 nulliparous women, with a 16.1% unexpected c-section rate. A nationwide multidisciplinary dataset for Korean foetal development was used to conduct external validation. KNN, Voting, XGBoost, Stacking, gradient boosting, random forest, LGBM, logistic regression, and SVM each have C-statistics of 0.6, 0.69, 0.64, 0.59, 0.66, 0.68, 0.68,0.7, and 0.69, respectively. With a prediction accuracy of 0.78, the logistic regression model performed the best. Individual risk of emergence of C-Sections during active labor could be predicted with machine learning methods.

Using real cases from a Tabriz health centre, researchers [52] experimentally tested multiple data-mining algorithms for determining the safest delivery mode for both mother and child. In addition, the implemented prediction models were evaluated using a cross-validation (CV) approach to ensure more accurate and dependable findings. With a 65 percent accuracy rate, the naive Bayesian classifier outperformed the other classifiers. Data on cesarean deliveries are limited, and expanding cesarean data is necessary for better prediction.

The research article [53] addresses the steps required to transition these exciting and promising techniques from research to clinical practice, as well as the main problems and limitations of AI in healthcare. The inherent challenges of machine learning science, logistical issues in deployment, and consideration of acceptance barriers as well as essential socio-cultural or route adjustments are all significant challenges for the application of ML systems in healthcare. Dataset shift, unintentional fitting of variables, unintended discriminating bias, the difficulty of adaptation to a new population, and the unexpected negative repercussions of algorithms designed on health outcomes are all potential concerns that AI algorithm developers must be aware of.

3 Method

The goal of this research is to classify locally gathered birth data by applying the ensemble classifiers and analysing the performance of these models. A set of different ensemble classifiers[23] are used (bagging, boosting, and stacking) to analyse which technique have higher accuracy for this particular dataset. For identifying the steps required for this project, the OSEMN (Obtain data, Scrub data, Explore data, Model, iNterpret results) method was used. The information was gathered at a government hospital in Lahore, Punjab's capital. After the information is recorded, it is cleaned and put into the structure for the experiment under the guidance of an Ob-Gyn. With a cross-validation approach, data is partitioned into different train and test sets, and the training data is then partitioned using k-fold cross-validation so the results from training the data could be compared with predictions from a separate test data which is not trained at all. The caret package in R software is used, which is regarded as a sophisticated machine learning model creation tool after the training and testing sets were available. The study uses a variety of classification models from both categories to conduct classification. Bagging (Bagged Cart, Random Forest), Boosting (Stochastic Gradient Boosting, C5.0), and Stacking (Linear Discriminate Analysis (LDA), Classification and Regression Trees (CART), Logistic Regression (via Generalized Linear Model or GLM), k-Nearest Neighbors (kNN), and Support Vector Machine with a Radial Basis Kernel Function) are some of the classifiers that are included. Several metrics were used to assess the models' reliability during the evaluation process. The study's plan is outlined below in Figure 3.1.

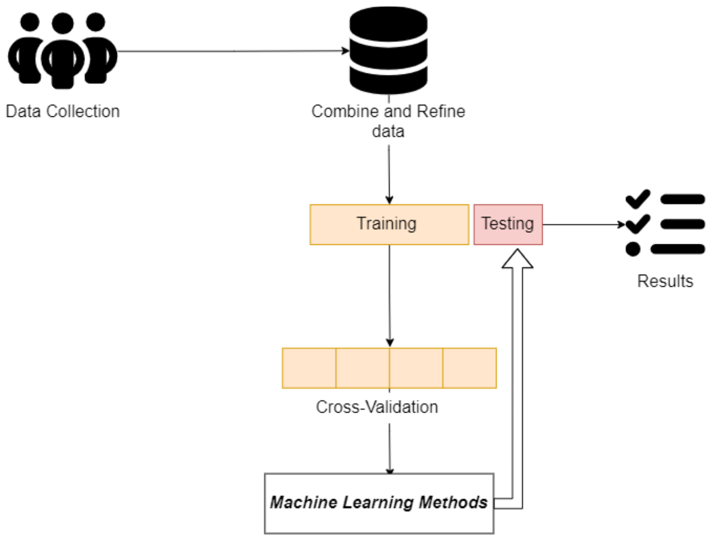

Figure 3.1: Scheme of the Study

4 Implementation

4.1 Dataset

The data for the experiment was obtained from government hospitals in Lahore, Pakistan's capital city. In the presence of an OB-Gyn, questionnaires were completed. 62 subjects were studied for various historic, maternal, cultural, and physical aspects. The study aims to investigate the impact of maternal care variables on birth mode. The current study does not include a cause-and-effect examination of lifestyle influences and birth outcomes. Using bagging, boosting, and stacking machine learning algorithms, 24 parameters (maternal history, throughout pregnancy) and 1 dependent variable are chosen to accomplish classification.

During the period of data collection, women of various ages were treated, ranging from 17 to 45 years old. C-sections were reported in greater numbers in younger patients and women over 40 years of age. In the current pregnancy, all women who had a previous c-section were delivered surgically. Mothers with diabetes are at increased risk to have a c-section birth than a normal birth. C-sections are also linked to high blood pressure. C-sections are associated with greater blood pressure in women. Natural births have been documented among middle-aged women, between the ages of 25 and 30. According to research, women who have had no past surgeries, have used less medicine, and have a high iron intake are more likely to give birth spontaneously.

Table 4.1: Dataset Features Description

Features	**Details**
class	Labeled class for vaginal or C-Section
abortion	Any history of Abortion (Yes/ No)
miscarriage	Any history of Miscarriage (Yes/ No)
last_mode	History of previous C-Sections (Yes/ No)
inheritdH	Any inherited disease (Yes/ No)
menstrual	Menstrual (regular or not)
days_mentcal	Days/length in a menstrual cycle
bleeding	Yes/ No (during the first trimester)
fatigue	Yes/ No

diabetic	History of Diabetics (Yes/ No)
breathing issues	Issues related to breathing (Yes/ No)
headache during pregnancy	Any type of headache during pregnancy (Yes/ No)
surgery (before preg)	Any other surgery before pregnancy (not C-Section)
BP patient	History of Blood pressure problem (Hypo or Hyper Tension) (Yes/ No)
Medicine intake	History of any kind of medication before pregnancy (Yes/ No)
headache (normal days)	History of Headache during normal days (Yes/ No)
hyper tension	History of Hyper Tension (Yes/ No)
use of folic acid	Patient taking FA (Yes/ No)
iron deficient	Patient have iron Deficiency (Yes/ No)
Bp MAX	Systolic value
BP MIN	Diastolic value
hemoglobin	Hemoglobin level in the body
Age	Age of the Patient

4.2 Modelling the algorithms

As described in 2.1 Preliminaries regarding importance of using these selected models this section complements the necessities of models by describing value addition in context of hyper-parameters and selection of evaluation matrices for Birth data analysis.
According to study [54], the most important component of developing the predictive model is identifying the ideal values of the hyperparameters to get the lowest test set error. The selected algorithms were modelled by using a setup where following hyper-parameters were selected based on different aspects of used dataset i.e., using cross-validation, and evaluation metrics (Accuracy, ROC and confusion matrix).
Cross-validation[55] is the statistical method for evaluating and analysing learning algorithms that divides data into two segments: one for learning or training a model and the other for validating the model. Typically, the train and validate sets must crossover over in subsequent cycles so that each piece of data gets an opportunity to be validated against. k-fold cross-validation is the most basic type of cross-validation. Other types of cross-validation

are variations on k-fold cross-validation or include many iterations of k-fold cross-validation.
The dataset is first divided into k equally (or relatively equally) sized sections or folds in k-fold cross-validation. Following that, k repetitions of training and validation are undertaken, with each iteration holding out a different fold of the data for validation while the remaining k-1 folds are used for learning.
An enhanced approach known as repeated cross-validation[56] is achieved by creating multiple sets of k-folds. In addition, cross-validation error is computed as the average of the repeated partitions. Over the classic single cross-validation method, repeated cross-validation can increase prediction accuracy and selection accuracy. It can be used when the data is limited (as in this case) and to avoid over-fitting of data.
In this study, 5- folds were selected for repeated cross-validation as the data is limited and with a greater number of folds can result in less amount of data in each fold which prevents effective training as the data is imbalanced.

An effective ensemble performs well on test data. The fundamental reason for this concept is that accuracy rate is usually the end goal when developing prediction models such as for ensembles[54]. Another reason for adopting accuracy as the performance measure definition is that various studies have been conducted to determine the extent to which diversity measurements may be utilized as performance parameters, and most variety metrics are described in terms of accuracy. When accuracy solely is not a good metric for performance analysis, consider performance in terms of how expertly ensembles can rank properly predicted examples ahead of incorrectly predicted instances. If ranking competence was the objective, the area under the ROC curve could be utilized[54] to define performance. Moreover, confusion matrix can be used to understand the sensitivity and specificity of the models.

4.3 Implementation of Bagging Technique

Bagging is an ensemble learning method that entails repeatedly training the same algorithm using different subsets of the training data. The final output forecast is then averaged over all sub-model predictions. Bagged Decision Trees[32] and Random Forest are the two most prevalent bagging ensemble approaches used in similar research as mentioned in 2.1.

- Resampling Method: Repeated Cross-Validation
- Repeats: 5 times
- Folds/partitions: 5
- Metric: Accuracy, Confusion Matrix, ROC Curve
- Method: Treebag, rf (Random Forest)

4.4 Implementation of Boosting Technique

Several models are trained consecutively in boosting, and each model improves from its predecessors' mistakes. I used a gradient boosting approach in this method.

- Resampling Method: Repeated Cross-Validation
- Repeats: 5 times
- Folds/partitions: 5
- Metric: Accuracy, Confusion Matrix, ROC Curve
- Method: c5.0, gbm

4.5 Implementation of Stacking Technique

The caretEnsemble package is used to combine the predictions of many caret models in this method. The caretStack() function is then used to define a higher-order model based on the list of models to learn how to optimally merge the predictions of sub-models. I have used a list of methods in the technique namely (Linear Discriminate Analysis (LDA), Classification and Regression Trees (CART), Logistic Regression (via Generalized Linear Model or GLM), k-Nearest Neighbors (kNN), and Support Vector Machine with a Radial Basis Kernel Function.

- Resampling Method: Repeated Cross-Validation
- Repeats: 5 times
- Folds/partitions: 5
- Metric: Accuracy, Confusion Matrix, ROC Curve
- Method: first level {lda, rpart, glm, knn, svmRadial} 2nd level {logistic regression}

5 Results

For evaluating the results initially, the data was divided into training and testing subsets. With the 5-fold repeated cross-validation (CV) scheme, each classifier was utilized to create scores for the training dataset. The training dataset was divided into five sections of equal size. Each subset served as a validation dataset for a trained model on all cases, with an equivalent number of non-cases drawn at random from the remaining 4 data subsets. This method of cross-validation was repeated five times, enabling each subset to act as the test dataset once.

5.1 Results for Bagging technique

The first technique used in the study is Bagging which includes two algorithms Random Forest and Treebag. These two methods were the most commonly used ensemble bagging techniques in the related literature.
After training both models the accuracy achieved with Treebag was 67.6% for training data and 61.38% for testing data. The accuracy of random Forest after training is 70.9% and the accuracy achieved after testing was 64.8%.

5.2 Results Boosting technique

The second technique used for evaluation is Boosting which includes C5.0 and Stochastic Gradient Boosting (GBM) algorithms. The accuracy achieved after training of both models is 73.32% and 72.46% and the testing accuracy is 66.08% and 68.51% respectively.

5.3 Results Stacking technique

The third and last technique is stacking and a number of algorithms are used in this technique which is (Linear Discriminate Analysis (LDA), Classification and Regression Trees (CART), Logistic Regression (via Generalized Linear Model or GLM), k-Nearest Neighbors (kNN), and Support Vector Machine with a Radial Basis Kernel Function. Logistic regression[36] was used as the meta estimator (second layer estimator) and the accuracy achieved after training and testing of the overall model is 73.8%, and 67.26% respectively.

5.4 Analysis of the joint results

The proposed 5 approaches' accuracy with regard to training and testing data is demonstrated.

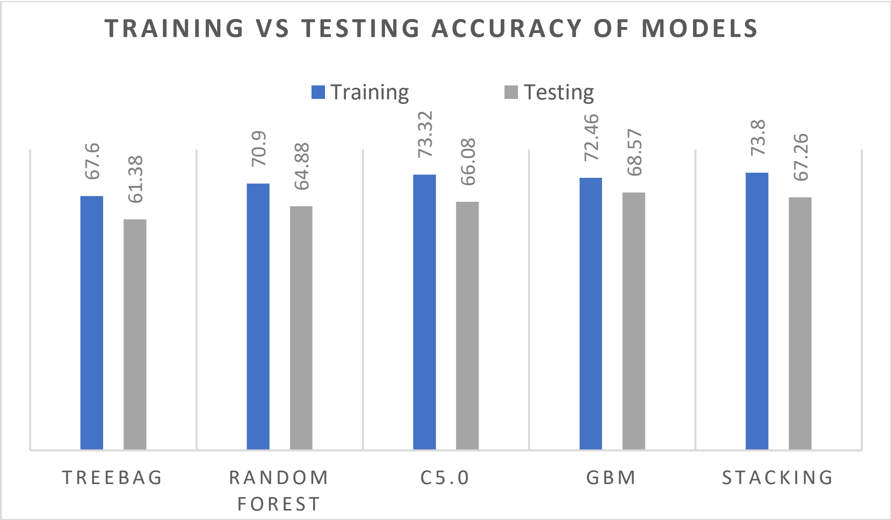

Figure 5.1: Accuracy comparison of training and testing models.

Inter classifier variance can be quantified using Kappa statistics in any circumstance when two or more distinct classification techniques are investigating the same thing. The computation is measured by the difference between the amount of agreement that is actually present ("observed") and the amount that would be predicted to be there by chance alone ("expected" agreement). Table 5.1: Interpretation of kappa [57]Table 5.1 shows how the Kappa statistics values are interpreted. When Kappa is more than 0.81, it means there is a lot of agreement between what is observed and what is expected[57].

Table 5.1: Interpretation of kappa [57]

	Poor	*Slight*	*Fair*	*Moderate*	*Substantial*	*Almost perfect*
Kappa	0.0	.20	.40	.60	.80	1.0

Kappa	Agreement
<0	Less than chance agreement
0.01-0.02	Slight agreement
0.21-0.40	Fair agreement
0.41-0.60	Moderate agreement
0.61-0.80	Substantial agreement
0.81-0.99	Almost perfect agreement

Figure 5.2: Accuracies v/s Kappa values for models.Figure 5.2 shows a summary of the results based on accuracy and kappa score of the selected models and the kappa score could be interpreted with the help from Table 5.1.

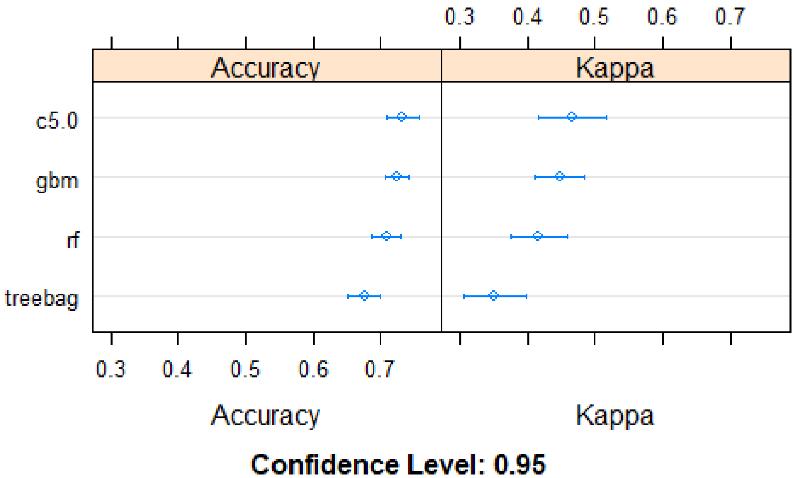

Figure 5.2: Accuracies v/s Kappa values for models.

The following are the confusion matrices for the calculated predictions.

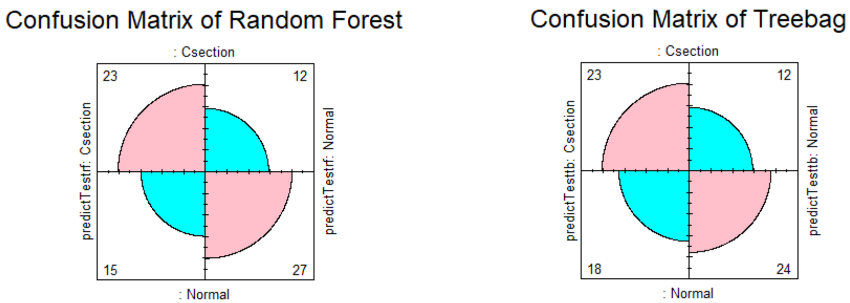

Figure 5.3: Confusion matrix of testing data Bagging models

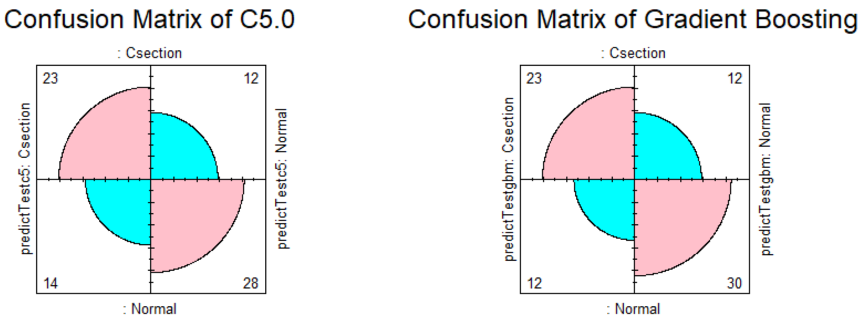

Figure 5.4: Confusion matrix of testing data Boosting models

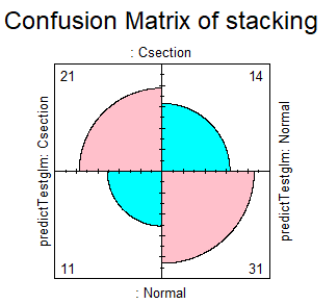

Figure 5.5: Confusion Matrix of testing data Stacking model

The accuracy of a classifier's analytic performance in distinguishing C-section cases from normal instances is assessed using Receiver Operating Characteristic (ROC) curve analysis [58]. It shows all possible outcomes of relative rates of different types of correct and wrong decisions. The area under the curve (ROC) is used to establish a link between the probabilistic estimates of the percent of true positive (sensitivity) and true negative (specificity) occurrences.

With 100 % specificity and sensitivity, a test with no overlapping in the predicted values has a ROC curve that passes through the top left corner. As a result, the closer the ROC is to the top left corner, the more accurate the classifier/test is[58].

Calculating the ROC curve's area-under-the-curve (AUC) and using it as a performance metric for the classifier is another typical application. It's a value between 0 and 1 and 1, with 1 suggesting that the predictor will always accurately classify a randomly provided case. A score of 0.5 shows that the predictor is no better than guessing at random. The AUC of a good classifier should be (slightly) greater than 0.5. The application determines how big an AUC will be to be considered "good enough." In some cases, an AUC of 0.7 may be sufficient, while in others, 0.95 may still be inadequate. In some circumstances, a lower AUC may be clinically acceptable if the sensitivity, for example, is high[59].

The ROC curves of the proposed models are as follow.

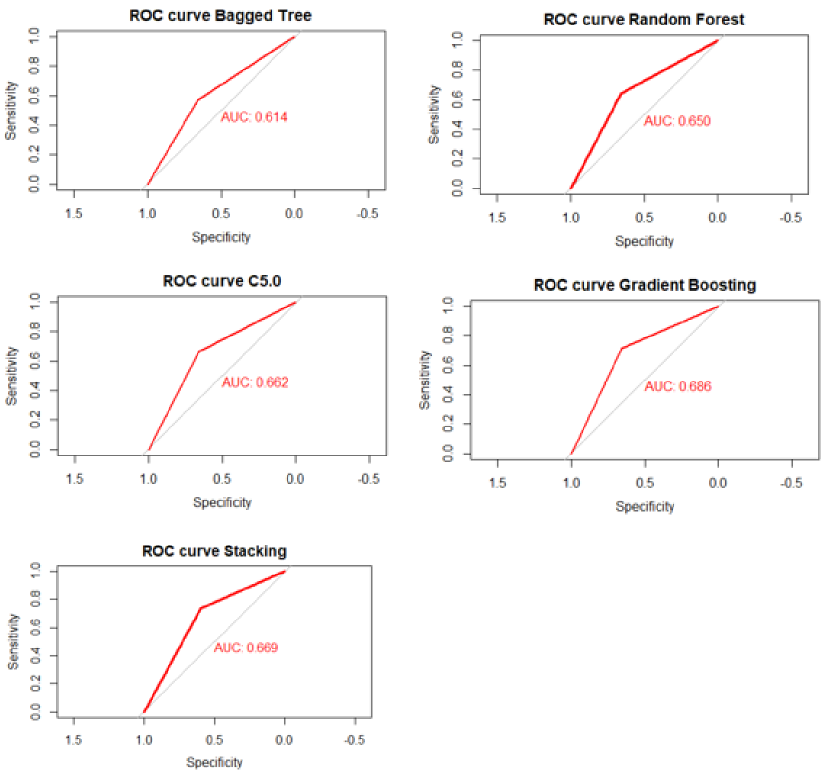

Figure 5.6: ROC curve of all models with AUC

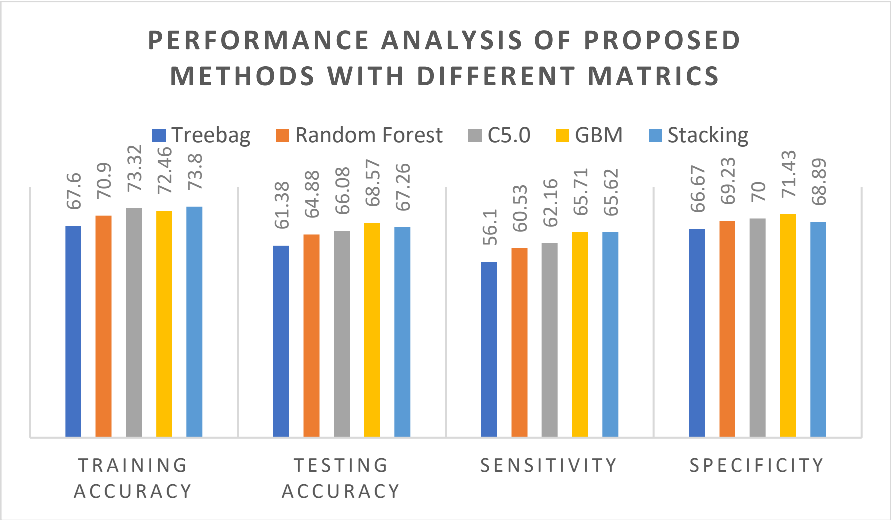

Figure 5.7: Comparison of different metrics of the proposed models

The above Figure 5.7 depicts different performance metrics of the proposed models/algorithms which can help to understand the models' performance. The boosting algorithm C5.0 has the highest training accuracy of 73.32% and the testing accuracy of this model is 66.08% which is the second-highest among the proposed models. Similarly, Gradient Boost Machine (GBM) performed better in both accuracy and Sensitivity than other models. Moreover, Treebag model had the least performance in all the above criteria.

6 Discussion and Conclusion

6.1 Methods, implementation, and results

The methods proposed for this study are selected through careful study of related work and how different researchers implemented these methods in their studies which are similar to this study. The cross-validation method used to split the training data and performing 5- fold validation increased the training capacity of the data as the data was not abundant which enabled the model to achieve higher accuracy.

6.2 Ethical and societal aspects

Any research project requires a high level of ethics. Ethics are upheld in the study by maintaining the answers obtained totally confidential. Many scientists and institutional ethics boards perceive maternity as a near-automatic reason for removal from research trials, despite the fact that the risks are minor and the study tackles an issue of essential importance to mother or fetal health[60]. Women are not well educated and normally gets really conscious about their personal data.

Before performing the research, consent was obtained from the respondents, and no inaccurate information was provided. The research was properly referenced in the theoretical analysis. Throughout the investigation, the ethical guidelines[61], [62] ,[63] were followed. According to these guidelines the data was anonymous, it is not publicly available and only used for research purpose, the data is used for beneficence purpose only.

6.3 Conclusion

We have demonstrated the use of several ML Bagging, Boosting, and Stacking algorithms in the realm of medical science in this paper. Machine learning approaches are used in conjunction with knowledge engineering to extract relevant information from publicly available birth data.

The research's main objectives are as follows:

Firstly, several models were selected from previous research to construct a decision system that predicted the class of birth data (C-Section or Normal) from the data collected locally. To fulfil the first goal, a self-explanatory questionnaire was filled out by researchers and members of the research team in the presence of the physician and the patient. The majority of the responses were categorical in nature. As a result, the acquired data are converted and standardized in SPSS, with necessary modifications made to deal with any inaccurate or missing information.

Secondly, the results of the selected models were described and analysed using metrics like accuracy, ROC, and confusion metrics. For classification which included objectives 1 and 2, Treebag, Random Forest, C5.0, Gradient Boost Machine (GBM), and Stacking algorithms were used. The research work was done in RStudio, and prediction methods were created by following the steps below.

Random number generation without replacement utilizing sample function and Dataset split, with 75% training and 25% test.

1. Apply control Cross-validation on training data.
2. Using training functions on classifiers that have been proposed.
3. Obtaining accuracies after the proposed classifiers have been fitted.
4. Making bar graphs and ROC curves.
5. The Confusion Matrix.
6. Creating various plots.
7. Plotting different results together.

The results suggest that GBM produces the best results with the highest accuracy (68.51%), while stacking produces better results with the highest accuracy (67.26%). The C5.0 accuracy is about (66.08%), the Random Forest accuracy is approximately (64.80%), and the Treebag accuracy is approximately (61.38%) which is the lowest of all proposed models.

Lastly, the research's purpose is to use known factors in machine learning algorithms in order to create a categorization model. The goal is to see whether applying classification algorithms to locally obtained data is practical and produces reliable results. The analysis presented in this study establish that the results obtained from locally collected data satisfied previous research findings.

6.4 Future work

The data sets used in the majority of the studies were not freely available. If the dataset containing pregnancy features is provided open source while personal information is hidden, researchers can successfully contribute to these types of studies. According to WHO, the majority of maternal mortality occurs in Africa and South Asia [64], despite the fact that relatively little study has been undertaken in these areas. Data can be gathered from these areas, and analysis can be conducted to determine the association between various elements in order to determine the causes of these regrettable deaths and problems.

The selected studies related to ensuring maternal safety have been carried out. Though the reviewed studies do not present any software framework, desktop, or mobile application. In the future, ML-based applications can be developed for predicting maternal complications along with the. In the future, besides conducting research on predicting mode of delivery, maternal complications, etc., ML-based applications can be developed. Consequently, these researches can be utilized in practical applications. Moreover, the usability and usefulness of these software systems can be evaluated to examine how effective they are in aiding doctors in clinical decision-making.

7 References

[1] G. Molina *et al.*, "Relationship between cesarean delivery rate and maternal and neonatal mortality," *JAMA - Journal of the American Medical Association*, vol. 314, no. 21, pp. 2263–2270, Dec. 2015, doi: 10.1001/jama.2015.15553.

[2] A. P. Betran, J. Ye, A. B. Moller, J. P. Souza, and J. Zhang, "Trends and projections of caesarean section rates: global and regional estimates," *BMJ Glob Health*, vol. 6, no. 6, Jun. 2021, doi: 10.1136/BMJGH-2021-005671.

[3] A. Amjad *et al.*, "Trends of caesarean section deliveries in Pakistan: secondary data analysis from Demographic and Health Surveys, 1990–2018," *BMC Pregnancy and Childbirth*, vol. 20, no. 1, pp. 1–13, Dec. 2020, doi: 10.1186/S12884-020-03457-Y/TABLES/4.

[4] "Pakistan Infant mortality rate, 1950-2021 - knoema.com." https://knoema.com/atlas/Pakistan/topics/Demographics/Mortality/Infant-mortality-rate (accessed Apr. 01, 2022).

[5] "Maternal mortality ratio in Pakistan still too high, shows report - Pakistan - DAWN.COM." https://www.dawn.com/news/1595085 (accessed Apr. 01, 2022).

[6] A. Pires-Menard, C. Flatley, and S. Kumar, "Severe neonatal outcomes associated with emergency cesarean section at term," *J Matern Fetal Neonatal Med*, vol. 34, no. 4, pp. 629–633, 2021, doi: 10.1080/14767058.2019.1611770.

[7] S. A. Abbas *et al.*, "Performance Analysis of Classification Algorithms on Birth Dataset," *IEEE Access*, vol. 8, pp. 102146–102154, 2020, doi: 10.1109/ACCESS.2020.2999899.

[8] UNFPA, UNICEF, W. H. Organization, W. B. Group, and the U. N. P. Division, "Trends in Maternal Mortality: 2000 to 2017 | UNFPA - United Nations Population Fund," *WHO, UNICEF, UNFPA,World Bank Group and theUnited Nations Population Division*, p. 104, 2019, Accessed: Jun. 02, 2022. [Online]. Available: https://www.unfpa.org/featured-publication/trends-maternal-mortality-2000-2017

[9] "Maternal mortality." https://www.who.int/news-room/fact-sheets/detail/maternal-mortality (accessed Jun. 02, 2022).

[10] "Maternal deaths decline slowly with vast inequalities worldwide." https://www.who.int/news/item/19-09-2019-maternal-

deaths-decline-slowly-with-vast-inequalities-worldwide (accessed Jun. 02, 2022).
[11] M. Carolan, M. A. Davey, M. A. Biro, and M. Kealy, "Older Maternal Age and Intervention in Labor: A Population-Based Study Comparing Older and Younger First-Time Mothers in Victoria, Australia," *Birth*, vol. 38, no. 1, pp. 24–29, Mar. 2011, doi: 10.1111/J.1523-536X.2010.00439.X.
[12] M. Neuman *et al.*, "Prevalence and determinants of caesarean section in private and public health facilities in underserved South Asian communities: cross-sectional analysis of data from Bangladesh, India and Nepal," *BMJ Open*, vol. 4, no. 12, p. e005982, Dec. 2014, doi: 10.1136/BMJOPEN-2014-005982.
[13] "Application of Decision Tree Algorithm for Data Mining in Healthcare Operations: A Case Study | Farhad Soleimanian Gharehchopogh - Academia.edu." https://www.academia.edu/26292461/Application_of_Decision_Tree_Algorithm_for_Data_Mining_in_Healthcare_Operations_A_Case_Study (accessed Jun. 02, 2022).
[14] J. R. Quinlan, "Induction of decision trees," *Machine Learning 1986 1:1*, vol. 1, no. 1, pp. 81–106, Mar. 1986, doi: 10.1007/BF00116251.
[15] L. Rokach and O. Maimon, "Top-down induction of decision trees classifiers - A survey," *IEEE Transactions on Systems, Man and Cybernetics Part C: Applications and Reviews*, vol. 35, no. 4, pp. 476–487, Nov. 2005, doi: 10.1109/TSMCC.2004.843247.
[16] S. ÇELİK, "Determination and Classification of Importance of Attributes Used in Diagnosing Pregnant Women's Birth Method," *Alphanumeric Journal*, vol. 8, no. 2, pp. 261–274, Oct. 2020, doi: 10.17093/ALPHANUMERIC.757769.
[17] P. Fergus, M. Selvaraj, and C. Chalmers, "Machine learning ensemble modelling to classify caesarean section and vaginal delivery types using Cardiotocography traces," *Computers in Biology and Medicine*, vol. 93, pp. 7–16, Feb. 2018, doi: 10.1016/j.compbiomed.2017.12.002.
[18] B. N. Lakshmi, T. S. Indumathi, and N. Ravi, "A Study on C.5 Decision Tree Classification Algorithm for Risk Predictions During Pregnancy," *Procedia Technology*, vol. 24, pp. 1542–1549, 2016, doi: 10.1016/J.PROTCY.2016.05.128.

[19] T. G. Dietterich, “Ensemble Methods in Machine Learning,” *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 1857 LNCS, pp. 1–15, 2000, doi: 10.1007/3-540-45014-9_1.
[20] “Many Heads Are Better Than One: The Case For Ensemble Learning - KDnuggets.” https://www.kdnuggets.com/2019/09/ensemble-learning.html (accessed Jun. 02, 2022).
[21] Z. Sun, Q. Song, X. Zhu, H. Sun, B. Xu, and Y. Zhou, “A novel ensemble method for classifying imbalanced data,” *Pattern Recognition*, vol. 48, no. 5, pp. 1623–1637, May 2015, doi: 10.1016/J.PATCOG.2014.11.014.
[22] R. Maclin and D. Opitz, “An Empirical Evaluation of Bagging and Boosting,” 1997, Accessed: Apr. 01, 2022. [Online]. Available: www.aaai.org
[23] “Ensemble methods: bagging, boosting and stacking | by Joseph Rocca | Towards Data Science.” https://towardsdatascience.com/ensemble-methods-bagging-boosting-and-stacking-c9214a10a205 (accessed Apr. 01, 2022).
[24] S. B. Kotsiantis, “Bagging and boosting variants for handling classifications problems: a survey,” *The Knowledge Engineering Review*, vol. 29, no. 1, pp. 78–100, Jan. 2014, doi: 10.1017/S0269888913000313.
[25] P. Bühlmann, “Bagging, Boosting and Ensemble Methods,” *Handbook of Computational Statistics*, pp. 985–1022, 2012, doi: 10.1007/978-3-642-21551-3_33.
[26] B. van Calster, L. Wynants, D. Timmerman, E. W. Steyerberg, and G. S. Collins, “Predictive analytics in health care: how can we know it works?,” *J Am Med Inform Assoc*, vol. 26, no. 12, p. 1651, Nov. 2019, doi: 10.1093/JAMIA/OCZ130.
[27] “What is Bagging? | IBM.” https://www.ibm.com/cloud/learn/bagging (accessed Jun. 22, 2022).
[28] L. Breiman, “Random Forests,” *Machine Learning 2001 45:1*, vol. 45, no. 1, pp. 5–32, Oct. 2001, doi: 10.1023/A:1010933404324.
[29] J. M. Luna *et al.*, “Building more accurate decision trees with the additive tree,” *Proc Natl Acad Sci U S A*, vol. 116, no. 40,

pp. 19887–19893, Oct. 2019, doi: 10.1073/PNAS.1816748116.

[30] C. Iwendi *et al.*, “COVID-19 patient health prediction using boosted random forest algorithm,” *Frontiers in Public Health*, vol. 8, p. 357, Jul. 2020, doi: 10.3389/FPUBH.2020.00357/BIBTEX.

[31] Z. Che, S. Purushotham, R. Khemani, and Y. Liu, “Distilling Knowledge from Deep Networks with Applications to Healthcare Domain,” Dec. 2015, doi: 10.48550/arxiv.1512.03542.

[32] “Classification of data using decision tree and regression tree methods.” http://mercury.webster.edu/aleshunas/R_learning_infrastructure/Classification%20of%20data%20using%20decision%20tree%20and%20regression%20tree%20methods.html (accessed May 16, 2022).

[33] A. Natekin and A. Knoll, “Gradient boosting machines, a tutorial,” *Frontiers in Neurorobotics*, vol. 7, no. DEC, 2013, doi: 10.3389/FNBOT.2013.00021/FULL.

[34] D. H. Wolpert, “Stacked generalization,” *Neural Networks*, vol. 5, no. 2, pp. 241–259, Jan. 1992, doi: 10.1016/S0893-6080(05)80023-1.

[35] S. Džeroski and B. Ženko, “Is Combining Classifiers with Stacking Better than Selecting the Best One?,” *Machine Learning 2004 54:3*, vol. 54, no. 3, pp. 255–273, Mar. 2004, doi: 10.1023/B:MACH.0000015881.36452.6E.

[36] S. le Cessie and J. C. van Houwelingen, “Ridge Estimators in Logistic Regression,” *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, vol. 41, no. 1, pp. 191–201, Mar. 1992, doi: 10.2307/2347628.

[37] V. N. Gudivada, M. T. Irfan, E. Fathi, and D. L. Rao, “Cognitive Analytics: Going Beyond Big Data Analytics and Machine Learning,” *Handbook of Statistics*, vol. 35, pp. 169–205, 2016, doi: 10.1016/BS.HOST.2016.07.010.

[38] H. Yigit, “A weighting approach for KNN classifier,” *2013 International Conference on Electronics, Computer and Computation, ICECCO 2013*, pp. 228–231, 2013, doi: 10.1109/ICECCO.2013.6718270.

[39] F. Song, D. Mei, and H. Li, “Feature selection based on linear discriminant analysis,” 2010, doi: 10.1109/ISDEA.2010.311.

[40] P. Fergus, M. Selvaraj, and C. Chalmers, “Machine learning ensemble modelling to classify caesarean section and vaginal delivery types using Cardiotocography traces,” *Computers in Biology and Medicine*, vol. 93, pp. 7–16, Feb. 2018, doi: 10.1016/J.COMPBIOMED.2017.12.002.
[41] J. Cleary-Goldman *et al.*, “Impact of maternal age on obstetric outcome,” *Obstet Gynecol*, vol. 105, no. 5, pp. 983–990, 2005, doi: 10.1097/01.AOG.0000158118.75532.51.
[42] M. E. B. Menai, F. J. Mohder, and F. Al-mutairi, “Influence of Feature Selection on Naïve Bayes Classifier for Recognizing Patterns in Cardiotocograms,” *Journal of Medical and Bioengineering*, vol. 2, no. 1, pp. 66–70, 2013, doi: 10.12720/JOMB.2.1.66-70.
[43] J. F. Ludvigsson and J. Ludvigsson, “Coeliac disease in the father affects the newborn,” *Gut*, vol. 49, no. 2, pp. 169–175, Aug. 2001, doi: 10.1136/GUT.49.2.169.
[44] R. Robu and Ş. Holban, “The Analysis and Classification of Birth Data.”
[45] M. Dulitzki, D. Soriano, E. Schiff, A. Chetrit, S. Mashiach, and D. S. Seidman, “Effect of very advanced maternal age on pregnancy outcome and rate of cesarean delivery,” *Obstet Gynecol*, vol. 92, no. 6, pp. 935–939, Dec. 1998, doi: 10.1016/S0029-7844(98)00335-4.
[46] I. O. Idowu *et al.*, “Artificial Intelligence for Detecting Preterm Uterine Activity in Gynecology and Obstetric Care,” *2015 IEEE International Conference on Computer and Information Technology*, pp. 215–220, Jan. 2015, doi: 10.1109/CIT/IUCC/DASC/PICOM.2015.31.
[47] J. Thomas and G. Kulanthaivel, “Data Mining Approach in Preterm Birth Prediction,” *Mapana - Journal of Sciences*, vol. 9, no. 1, pp. 18–30, May 2010, doi: 10.12723/MJS.16.3.
[48] S. K. Murthy, “Automatic Construction of Decision Trees from Data: A Multi-Disciplinary Survey”.
[49] H. Wang, X. Lu, Z. Hu, and W. Zheng, “Fisher discriminant analysis with L1-norm,” *IEEE Transactions on Cybernetics*, vol. 44, no. 6, pp. 828–842, 2014, doi: 10.1109/TCYB.2013.2273355.
[50] M. N. Islam, S. N. Mustafina, T. Mahmud, and N. I. Khan, “Machine learning to predict pregnancy outcomes: a systematic review, synthesizing framework and future research

agenda," *BMC Pregnancy and Childbirth 2022 22:1*, vol. 22, no. 1, pp. 1–19, Apr. 2022, doi: 10.1186/S12884-022-04594-2.

[51] J. H. Wie *et al.*, "Prediction of Emergency Cesarean Section Using Machine Learning Methods: Development and External Validation of a Nationwide Multicenter Dataset in Republic of Korea," *Life (Basel)*, vol. 12, no. 4, p. 604, Apr. 2022, doi: 10.3390/LIFE12040604.

[52] M. M. Jamjoom, "Data Mining in Healthcare to Predict Cesarean Delivery Operations using a Real Dataset", Accessed: Jun. 02, 2022. [Online]. Available: https://orcid.org/0000-0001-9149-2810

[53] C. J. Kelly, A. Karthikesalingam, M. Suleyman, G. Corrado, and D. King, "Key challenges for delivering clinical impact with artificial intelligence," *BMC Medicine*, vol. 17, no. 1, pp. 1–9, Oct. 2019, doi: 10.1186/S12916-019-1426-2/PEER-REVIEW.

[54] "On Effectively Creating Ensembles of Classifiers Studies on Creation Strategies, Diversity and Predicting with Confi-dence Tuwe Löfström".

[55] P. Refaeilzadeh, L. Tang, and H. Liu, "Cross-Validation," *Encyclopedia of Database Systems*, pp. 1–7, 2016, doi: 10.1007/978-1-4899-7993-3_565-2.

[56] Y. zhong, J. He, P. Chalise, Y. zhong, J. He, and P. Chalise, "Nested and Repeated Cross Validation for Classification Model With High-dimensional Data," *Revista Colombiana de Estadística*, vol. 43, no. 1, pp. 103–125, Jan. 2020, doi: 10.15446/RCE.V43N1.80000.

[57] "[PDF] Understanding interobserver agreement: the kappa statistic. | Scinapse." https://www.scinapse.io/papers/1444168786 (accessed Jun. 03, 2022).

[58] K. Hajian-Tilaki, "Receiver Operating Characteristic (ROC) Curve Analysis for Medical Diagnostic Test Evaluation," *Caspian Journal of Internal Medicine*, vol. 4, no. 2, p. 627, 2013, Accessed: Jun. 03, 2022. [Online]. Available: /pmc/articles/PMC3755824/

[59] J. Tohka and M. van Gils, "Evaluation of machine learning algorithms for health and wellness applications: A tutorial," *Computers in Biology and Medicine*, vol. 132, p. 104324, May 2021, doi: 10.1016/J.COMPBIOMED.2021.104324.

[60] A. D. Lyerly and R. R. Faden, “Mothers matter: Ethics and research during pregnancy,” *Virtual Mentor*, vol. 15, no. 9, pp. 775–778, 2013, doi: 10.1001/VIRTUALMENTOR.2013.15.9.PFOR1-1309.

[61] A. This, J. Cleland, N. Scott, K. Harrild, and M. Moffat, “Using databases in medical education research: AMEE Guide No. 77,” *Medical Teacher*, vol. 35, no. 5, pp. 1103–1122, 2013, Accessed: Jun. 22, 2022. [Online]. Available: www.amee.org

[62] L. W. T. Schuwirth and S. J. Durning, “Ethics approval for health professions education research: are we going too far down the barrel?,” *Medical Education*, vol. 53, no. 10, p. 956, Oct. 2019, doi: 10.1111/MEDU.13942.

[63] H. Forester-Miller and T. Davis, “A Practitioner’s Guide to Ethical Decision Making,” 1996.

[64] “Why Maternal Mortality Is So High in Sub-Saharan Africa.” https://www.globalcitizen.org/en/content/maternal-mortality-sub-saharan-africa-causes/ (accessed Jun. 03, 2022).

8 Appendices

8.1 Apendix A

Questioners
Health Questionnaire: Pregnant Woman
Note: (If this is your second marriage and you have been pregnant in your first marriage relation, please quit filling the questioner)

This questioner is intended to gather information of pregnant women in the region. Information you provide will be used solely for research purpose while keeping the ethical boundaries in mind. You are requested to provide information with best of your knowledge.

Date:____________
Age: Years_____ Months____ Education____________
Blood Group(patient):
Resident of : City __________ Area__________

1.Are you a house wife? If No then go to question # 6
Yes_____ No_____

2.What is your Occupation?
1. Medicine__________2. Teaching_________3. Law_______
4. Aviation___________5.Banking_________
6. Information Technology_____ 7.Designer_________
8. Beaution _____
10. Entrepreneurship __________11 others___________

3.What are your Working days________ and working hours per day_________?

4.Is it your job needs physical movement more than sitting?
Yes_____ No_____

5.What was your age when married? _______________

Note: Please consider your pregnancy a pregnancy even if it has been resulted in abortion or miscarriage.

6.Is it your first pregnancy? If answer is Yes, go to question #28
Yes_____ No_____

7.How many times you have been pregnant before the current pregnancy?
1 2 3 4 5 6 or more__________

8. How many successful births you have delivered?

1 2 3 4 5 6 or more___________

9.Birth weights of your child/children, as best of your knowledge?
Child#1_______ Child#2______ Child #3______ don't remember_____

Encircle the correct choice

10.How many boys born among your previous pregnancies?
0 1 2 3 4 5 or more______

13. How many abortions occurred in your married life?
0 1 2 3 4 or more______

14. How many miscarriages occurred in your married life?
0 1 2 3 4 or more______

16.Number of premature births? If answer is 0, go to question # 21
0 1 2 3 4 or more______

17.How many times you gave birth to a premature?
1 2 3 or more____________
18. How much premature was your baby? Answer in days from expected delivery date?
Child #1___________ Child # 2____________ Child #3__________
Or
more____

19.What is the current age of your premature baby?

Child # 1__________ Child #2___________ Deceased(details)_________

20.Number of normal deliveries?
0 1 2 3 4 5 or more______

21. Number of C section deliveries?
0 1 2 3 4 5 or more______

22. Mode of last delivery?
Normal___________ C-Section___________

23.Did you visit prenatal care centers in previous pregnancies?
Yes_______ No________

24. Did you breast feed?
Yes_________ No________ Occasionally ________

25.Is there any history of increase B.P in previous pregnancy?

Yes_________ No________

26.Did you use any family planning method during your married life?
Yes______ No_______

27. Is your menstrual cycle regular?
Yes ______ No_______

28. How many days your menstrual cycle has? _________

29.What was the Date of your last day of last menstrual cy-
cle________________

30.Did you have implantation bleeding in current pregnancy?
Yes_____ No_______
31.Are you suffering from fatigue?
Yes_______ No________

32.Are you diabetic? If yes, how long? _____________

33. Do you have breathing issues?
Yes ______ No________
34.Do you feel more headaches than your non pregnant days?
Yes ______ No________

35. Do you have strong change in breathing and fast heart beat?
Yes______ No__________
_
36. Have you ever had any surgery? Other than C-section?
No______ Yes(de-
tails)____

37. What is your Hemoglobin rate? ________________.

38. Is your Palpitation per minute normal
Yes_____ No________

39. What is your palpitation rate?_____________

40. What is your Heart beat rate?_______________

41. What is your blood pressure? Max______ Min_______

42. Are you a patient of blood pressure?
Yes_____ No________

43. Are you on some medication other than pregnancy?
No ________

Yes(details)________________

44. DO you have headache, epigastric, pain, fits?
Yes______ No_______

45. Are you taking any any Rx for hypertension in this pregnancy?
Yes______ No________

46. Are you a booked or non-booked? ________

47. Have you taken F/A in 1st trimester?
Yes_____ No_________

48. Have you taken iron supplements in last two trimesters?
Yes______ No_________

49. Are you vaccinated?
Yes______ No_________

50. Have you done your ultrasound?
Yes______ No________

51. Do you have hemorrhoids?
Yes______ No_______

52. Do you have ever bleed stained diarrhea/vomiting?
Yes_______ No_______

53. Is there any history of bleeding disorder or thalassemia trait?
Yes_____ No_______

54. Do you have ever bleeding form nose moth are any part of body?
Yes_____ No________

55. Do you have blurning of vision?
Yes_____ No________

56. Do you have numbness of limbs?
Yes_______ No________

57. What is your Expected delivery date?
____

58. Patient's weight? _________________________

59. Do you have cravings to eat items which are not diet? For example dirt, ice, clay?

Yes ______ No________

60. How many times you eat meat weekly? ______________

61. Do you have deficiency in buying adequate food like milk, meat or fruit?

Yes______ No________

62. Which of the following food items you consider hard/challenging to eat during pregnancy?

1. Milk and related
2. Proteins
3. Fruits
4. Vegetables
5. Bread
6. rice
7. Other: _________________