American International University-Bangladesh (AIUB)

Fetal Health Prediction using Machine Learning Algorithm

Mifa, Afia Farjana (18-39114-3)

Ara, Anjuman (18-39015-3)

Latif, Nayeem Bin Abdul (17-35801-3)

Nusrat,Fatema Tuj Zohora (18-39165-3)

Department of Computer Science Faculty of Science & Information Technology American International University, Bangladesh

7 January, 2022

Declaration by author

This thesis is composed of our original work, and contains no material previously published or written by another person except where due reference has been made in the text. Others' commitments to our proposal as an entire, such as factual help, overview plan, information investigation, basic mechanical forms, proficient altering exhortation, and any other special inquire about exertion utilized or archived in our proposition, are recognized, and a list of references is given.

Afia Farjana

Mifa, Afia Farjana (18-39114-3)

Anjuman Ara

Ara, Anjuman (18-39015-3)

Nayeem

Latif, Nayeem Bin Abdul (17-35801-3)

Nusrat

Nusrat, Fatema Tuj Zohora (18-39165-3)

.

Approval

The thesis titled **“Fetal Health Prediction using Machine Learning Algorithm”** has been submitted to the following respected members of the board of examiners of the department of computer science in partial fulfillment of the requirements for the degree of Bachelor of Science in Computer Science on **(June 26)** and has been acceptedas satisfactory.

. . . .

Dr. S. M. Hasan Mahmud
Assistant Professor & Supervisor
Department of Computer Science
American International University-Bangladesh

. . . .

DR. Muhammad Firoz Mridha
Associate Professor
Department of Computer Science
American International University-Bangladesh

. . . .

Dr. Md. Abdullah-Al-Jubair
Assistant Professor & Head (Undergraduate)
Department of Computer Science
American International University-Bangladesh

. . . .

Professor Dr. Tafazzal Hossain
Dean
Faculty of Science & Information Technology
American International University-Bangladesh

. . . .

Dr. Carmen Z. Lamagna
Vice Chancellor
American International University-Bangladesh

Acknowledgments

The researchers were appreciative for the clinic's support, as well as Dr. Jobeda Hasan, the nurse, the team, and the practitioners of Tanha Health Care and Hospital, Shafipur, Bangladesh. The dataset utilized in this project can be requested through the mail. Work to improve and expand this data set will continue. We want analysts who need to bolster this investigation to communicate with us in this setting. I express my appreciation to the Almighty Allah for the most plentiful blessings. Despite a problematic outbreak, we could finish the paper with His assistance. We seem to have been able to finish our paper without even any significant delays. We would also like to thank Jupiter notebook for making a great tool available for free. This thesis would not be achievable without it. We also want to express our gratitude to our thesis supervisor, Dr. S. Mr. Hasan Mahmud, for his kind assistance, advice, and encouragement. When we needed some help, he was always keen to support us. Finally, to begin, I express my appreciation to the Almighty Allah for the most plentiful blessings. We would also like to express our gratitude to Jupiter notebook for making a great tool available for free. When we needed some help, he was always keen to support us.

Abstract

Congenital disorders are morphological or functional abnormalities that develop during fetal life. These conditions, sometimes called congenital disabilities, congenital disorders, or birth defects, occur during pregnancy and can be detected before or after birth. About 6% of babies are born with a genetic defect, causing thousands of deaths worldwide. Machine learning is applied in a variety of aspects of fitness care, including finding of latest clinical care, affected person facts and file management, and continual ailment treatment. Frequent contact with the uterus, the health and development of the fetus of pregnant women are referred to as fetal health. Maximum complications of the pregnancy period cause the fetus to have serious difficulty in limiting its growth, causing deficiency or death. The risk-free pregnancy period is expanded by foreseeing chance levels some time recently the onset of difficulties. Forecasting fetal wellbeing and development from a set of pre-classified designs is fundamental in creating a prescient classifier show with Machine Learning Calculations. Routine examinations and early treatment are increasingly being shown to be effective in preventing death or serious illness. However, these aspects are often overlooked in health care, particularly in low-income countries where inadequate medical services frequently result in multiple fatalities. To discover the mutation, women must have ultrasound tests and fetal assessments to improve detection and definition of these abnormalities. In this article, we have tried to create a prediction system with an online health care website so that every woman can use it. For this reason, we use nine popular machine learning algorithms. These are SVM, AdaBoost, Light, Boost, CNN, Decision Tree, Decision Forest, RF, KNN, Logistic regression, and Naive Bayes. They were all prepared utilizing the 183 pregnant women's clinical dataset, which was utilized to dissect information and anticipate fetal variation from the norm status based on parental and medical studies. The dataset was collected through a maternal survey and a nitty gritty assessment by Tanha Health Care and Hospital, Shafipur, Bangladesh. Based on data that was separated into balance data and imbalance data, we derived our result. The outcome was very pleasing. Random forest had the highest accuracy of 89 percent for balance data. In terms of accuracy, DT had the highest accuracy for imbalance data 81%. The rate of sound fetuses was 81.97 percent, and the rate of unhealthful fetuses was 18.03 percent. In this paper our purpose is to help pregnant women and their households get sounders to indicate fetal mutation. Using ml algorithms and the Fetal Health Checker website, we exclude the conventional pregnancy tests.

Keywords

Fetal health, Pregnancy, Classification, Comparison, congenital disorders, Random forest, Machine Learning, Fetal Heart Rate

Table of Contents

CHAPTER 1 .... **11**

Introduction 11

1.1.1 Thesis topic *13*

1.1.2 Clinical dataset *13*

1.1.3 Motivation *13*

1.1.4 Objective *14*

1.1.5 Orientation *14*

CHAPTER 2 .... **15**

Literature review 15

CHAPTER 3 .... **20**

Methods 20

3.1.1 Proposed Model *20*

3.1.2 System Architecture *21*

3.1.3 Web application *22*

3.1.4 Prediction system *22*

3.1.5 Dataset introduction *25*

3.1.6 Data Visualization *26*

3.1.7 Histogram *26*

3.1.8 Bivariate histogram *28*

3.1.9 Corelation heatmap *30*

3.1.10 Data preprocessing *31*

3.1.11 Undersampling *31*

3.1.12 Standard Scaler *32*

3.1.13 SMOTE *32*

3.1.14 Train-test split *33*

3.1.15 K-fold Cross validation *34*

3.1.16 Tools and Libraries Used *46*

CHAPTER 4 .... **50**

Results or findings 50

4.1.1 Evaluation of Algorithm performance *50*

CHAPTER 5 .... **69**

Discussion 69

CHAPTER 6 .... **71**

Conclusion 71

BIBLIOGRAPHY **73**

List of Figures

FIGURE 1: TECHNICAL APPROACH VISUALIZATION 20
FIGURE 2: TECHNICAL APPROACH DIAGRAM 21
FIGURE 3: SYSTEM ARCHITECTURE 22
FIGURE 4: WEBSITE INTERFACE 24
FIGURE 5: FETAL AGE HISTOGRAM 27
FIGURE 6: BLOOD SEROTYPE OF MOTHER HISTOGRAM 28
FIGURE 7: FETAL AGE AND STATUS BIVARIATE HISTOGRAM (HEATMAP) 29
FIGURE 8: FETAL AGE AND STATUS BIVARIATE HISTOGRAM 29
FIGURE 9: CORRELATION HEATMAP 30
FIGURE 10: K-FOLD CROSS VALIDATION ON IMBALANCED DATASET 35
FIGURE 11: KNN ALGORITHM 37
FIGURE 12: LR ALGORITHM 38
FIGURE 13: DECISION TREE ALGORITHM 39
FIGURE 14: RANDOM FOREST ALGORITHM 40
FIGURE 15: SUPPORT VECTOR MACHINE ALGORITHM 41
FIGURE 16: LOGISTIC REGRESSION ALGORITHM 42
FIGURE 17: CONVOLUTIONAL NEURAL NETWORKS ALGORITHM 43
FIGURE 18: LIGHT GRADIENT BOOSTING MACHINE ALGORITHM 44
FIGURE 19: EXTREME GRADIENT BOOSTING ALGORITHM 45
FIGURE 20: CONFUSION MATRIX 53
FIGURE 21: ACCURACY OF VARIOUS ALGORITHMS 56
FIGURE 22: SENSITIVITY SCORE FOR BALANCED AND IMBALANCED DATA 57
FIGURE 23: ROC CURVE FOR IMBALANCED DATA 58
FIGURE 24: ROC CURVE FOR BALANCED DATA 59
FIGURE 25: COMPARISON OF AUC SCORE AMONG VARIOUS ALGORITHMS 59
FIGURE 26: CONFUSION MATRIX KNN FOR IMBALANCED DATASET 61
FIGURE 27: CONFUSION MATRIX KNN FOR BALANCED DATASET 61
FIGURE 28: CONFUSION MATRIX LINEAR REGRESSION FOR IMBALANCED DATASET 62
FIGURE 29: CONFUSION MATRIX LINEAR REGRESSION FOR IMBALANCED DATASET 62
FIGURE 30: CONFUSION MATRIX LOGISTICS REGRESSION FOR IMBALANCED DATASET 62
FIGURE 31: CONFUSION MATRIX LOGISTIC REGRESSION FOR BALANCED DATASET 62
FIGURE 32: CONFUSION MATRIX DECISION TREE FOR IMBALANCED DATASET 63
FIGURE 33: CONFUSION MATRIX DECISION TREE FOR BALANCED DATASET 63
FIGURE 34: CONFUSION MATRIX RANDOM FOREST FOR IMBALANCED DATASET 64
FIGURE 35: CONFUSION MATRIX RANDOM FOREST FOR BALANCED DATASET 64
FIGURE 36: CONFUSION MATRIX SVM FOR IMBALANCED DATASET 64
FIGURE 37: CONFUSION MATRIX SVM FOR BALANCED DATASET 64
FIGURE 38:CONFUSION MATRIX LGBM FOR IMBALANCED DATASET 65
FIGURE 39: CONFUSION MATRIX LGBM FOR BALANCED DATASET 65
FIGURE 40: CONFUSION MATRIX XGBOOST FOR IMBALANCED DATASET 66
FIGURE 41: CONFUSION MATRIX XGBOOST FOR BALANCED DATASET 66
FIGURE 42:CONFUSION MATRIX CNN FOR IMBALANCED DATASET 66
FIGURE 43: CONFUSION MATRIX SVM FOR IMBALANCED DATASET 66
FIGURE 44: FEATURE RANKING IMBALANCED DATA 67
FIGURE 45: FEATURE RANKING BALANCED DATA 68

List of Tables

TABLE 1: COMPARISON OF PREVIOUS RELATED WORKS 19
TABLE 2: DATASET SUMMARY 26
TABLE 3: COMPARISON OF SAMPLE BEFORE AND AFTER UNDERSAMPLING 32
TABLE 4: COMPARISON OF SAMPLES BEFORE AND AFTER USING SMOTE 33
TABLE 5: COMPARISON OF TRAIN-TEST SPLIT AMONG BALANCED AND IMBALANCED DATASET....34
TABLE 6: K-FOLD CROSS VALIDATION SCORE FOR BALANCED AND IMBALANCED DATASET 35
TABLE 7: ACCURACY, AUC, SENSITIVITY AND SPECIFICITY FOR IMBALANCED DATA....55
TABLE 8: ACCURACY, AUC, SENSITIVITY AND SPECIFICITY FOR BALANCED DATA 55
TABLE 9: PRECISION, RECALL, F1 SCORE FOR IMBALANCED DATA 60
TABLE 10: PRECISION, RECALL, F1 SCORE FOR BALANCED DATA 60

List of Abbreviations and Symbols

Abbreviation	
CS	Computer Science
CSE	Computer Science and Engineering
NN	Neural Network
CNN	Convolutional Neural Networks
ML	Machine Learning
LGBM	Light gradient boosting machine
LR	Logistic regression
DT	Decision tree
RF	Random Forest
SVM	Support vector machine
XGBoost	Extreme gradient boosting
TPR	True positive rate
FPR	False positive rate
TP	True positive
FP	False positive
AUC	Area under the curve
ROC	Receiver operator characteristics

Chapter 1

Introduction

Birth defects in children are cause of agony for families who are already victim of this predicament. According to WHO, birth defects also known as congenital abnormalities, congenital disorders or congenital malformations can be defined as structural or functional anomalies that occur during intrauterine life and can be identified prenatally, at birth, or sometimes may only be detected later in infancy. Broadly, congenital refers to the existence at or before birth. An estimated 240 000 newborns die worldwide within 28 days of birth every year due to birth defects. Birth defects cause a further 170 000 deaths of children between the ages of 1 month and 5 years. (WHO) Although, identifying the exact cause of birth defect is complicated due to the complex nature of the possibilities, it can be attributable to various factors such as: genetics, socioeconomic, demographic, and environmental.

An early detection of fetal birth defects can provide a possible opportunity to mitigate the severity of the outcome. Medical professionals can infer the presence of birth defects among the unborn child by considering the mother's medical history, drug intake, and various maternal or familial factors. For detecting and diagnosing the defect ultrasound scanning is the most reliable and widely used. Studies show that the detection rate of such scans is approximately 60–87% [1]. Early detection of birth defects can provide an opportunity to provide proper care to the mother which can reduce the occurrence of birth defect. Proper primary health care along with secondary or tertiary level referral care could correct over 40% birth defects [2]. Unfortunately for the people belonging in less developed and developed countries such necessities are a privilege. In the countries where people mostly live on an extremely low income receiving the diagnosis and aftercare is unaffordable to them.

Bangladesh is one such developing country where majority of the population has a low to middle income range. Being one of the most densely populated countries with 1,115 peoples/Km^2, makes the living condition distressful for the majority of the population. The country has a pollution of 161 million with 21.8% living below the poverty line [3]. Although, the country has a developing economy many its people cannot properly avail the benefits of the growth. The healthcare system in Bangladesh is not adequate enough to provide proper healthcare to the pregnant woman in need. Majority of its medical facilities are situated within the urban locations for which preventative measures for ensuring the health of a newborn is rarely a possibility for the people living in the rural area with a low income.

In Bangladesh, no national population-wide data on the congenital disorders is available. Most of the deduction has to be drawn from the available tertiary level healthcare facility-based data. As in Bangladesh approximately 72% of all deliveries are conducted by traditional birth attendant at home in a lack of proper diagnostic facilities for which the majority of the fetal or early childhood deaths relating to birth defects remain unknown [4]. A specialized facility-based study conducted during January to December 2007 has found that among 1630 deliveries including live births, stillborn and abortions found 60 congenital abnormalities which make 3.68% of the total deliveries [5].

Although exact cause of congenital birth defects cannot be identified the correlated risk factors can provide a means to detect and prevent birth defects. Apart from genetics, various social, environmental and maternal health factors can contribute to the development of birth defects. A study reported that maternal exposure to the following known risk factors: antiepileptic drugs, vitamin-A, maternal diabetes, history of smoking, previous delivery of malformed baby, young mother and consanguineous marriage can increase as risk factors for congenital anomalies in the fetus [6]. Furthermore, maternal infections, diabetes, hypertension also have a correlation to the development of congenital abnormalities. Avoiding the risk factor can make 70% of the birth defects preventable. [7]

As proper healthcare is still a privilege for many in Bangladesh, most people avoid going to the healthcare centers without obvious and severe symptoms. If the families are informed early about the risks regarding their unborn child early without the hassle and cost of traveling to a healthcare center they will become willing and prepared to provide proper healthcare to the expecting mothers. An effective solution is the use of telemedicine. Bangladesh has begun taking steps in telemedicine services [8]. Although it is not as widespread or efficient but can become an alternative to the strenuous visit to the hospital for many.

So considering the situation in Bangladesh, the idea of our research is to use machine learning to predict the health status of the fetus based on the risk factors which are relevant to Bangladeshi population. The model will be deployed to a web application where patients can provide answers to the question based on the relevant maternal clinical history and familial attributes and receive a prediction based on that. The parents can use the web application even before the pregnancy for being informed about the risks regarding the birth of their children.

1.1.1 Thesis topic

The classification of fetal states using machine learning has been the subject of much research [19]. Fetal mortality is common in developing and underdeveloped countries because of fetal health issues. Machine Learning (ML) algorithms have made significant advances in disease diagnosis, treatment, and prognosis since their inception in healthcare. Most challenging and demanding medical processes monitor fetal health growth during pregnancy. Even if preventative steps have been implemented, approximately 810 pregnant women die every day, according to the World Health Organization (WHO) [20]. The findings indicate that machine learning will become a common foundation for monitoring early fetal health in automated systems [21].

1.1.2 Clinical dataset

The majority of health and medical research relies on patient records. During the course of patient care, we gathered clinical data. We went to the hospital called Tanha Health Care and Hospital. Under Dr. Jobeda Hasans's patient list, we collect all the data based on our features. It is a text dataset in CSV format that contains features in columns and samples in rows. The dataset contains 16 distinct features. The 16 features are General menstrual cyclic status of mother, Fetal age, Type of pregnancy, Blood serotype of mother, Past delivery Abortion status, Diabetic’s history, Hypertension, Other illness, Operation history, Presence of disabled children, Presence of disabled children in father family, Covid history, Previous child home or hospital delivery, Any other illness, Fetal heath Status.

1.1.3 Motivation

Fetal birth defect is a prevalent problem worldwide. But, the population coming from the less developed and developing countries suffer immensely due to having a lack of option to prevent or mitigate the severity of the consequences. Birth defects also correspond to the occurrence of stillbirths. In 2019, there had been around two million stillbirths worldwide, with more than 64.5 percent of cases from low- to middle-income households (Hug et al., 2020). As the majority of the population in Bangladesh belongs to the low-to-middle income group the problem of fetal birth defect is an extremely relevant issue for this country. The risks of fetal birth defects can be seen to have a correlation to the clinical history of the mother. This provides a way to predict the probability of having birth defects in the unborn child beforehand so that, such patients can be monitored and provided the needed care. But, living with low income and large population does not provide the people of a country as Bangladesh to have the opportunity, information and resources to be informed of the risks earlier by a healthcare professional. Our main goal is to use the technology of machine learning to predict the health status of the fetus as early as possible and provide an online web based platform to the expected mothers for them to be informed about the possibilities and risks regarding the health of their unborn child by themselves.

1.1.4 Objective

The main objective of this research is to provide a system which can possibly aid the expecting mothers in Bangladesh by informing them about the risks regarding the health of their unborn child. The women living in a developing country is often deprived of necessary healthcare during pregnancy. Although, most birth defects are preventable and reducible many bear the torments due to the lack of accessible information. Our aim is to assist the expecting mothers of Bangladesh by providing them access to a web application which can predict the health status of the unborn child early on.

- **Sub objective 1:** Compare various machine learning technologies to analyze their performance on the fetal health dataset.
- **Sub objective 2:** To contribute to the research field of fetal health in Bangladesh as the amount of pre-existing studies are scarce.
- **Sub objective 2:** Increase the amount of data availability relating to fetal health for advancing further research in Bangladesh

1.1.5 Orientation

The second chapter of this paper discusses past work and its outcomes, as well as the principles of machine learning and fetal health. Moreover, the third chapter will discuss the mechanisms of the proposed algorithm, including implementation, data preprocessing, and dataset visualization. We organize all of the results, as well as their influence and percentage, in Chapter 4. We talked about the paper and the algorithm analysis in Chapter 5. Lastly final chapter is all about authors and we summarized our work

Chapter 2

Literature review

People's health and wellness are among the most essential things, so the healthcare industry has always received the best attention and been kept up to date. The medical sector, which has traditionally been given higher priority, has always been benefited from the development of new drugs and technologies. Artificial intelligence and machine learning, which are components of computer science, are currently applied in a variety of medical sectors. Medical therapy, fetal status determination, chronic diseases and early disease identification are a few examples that are substantially helping society. The managerial side of the healthcare sector is another area where machine learning is tremendously benefiting from doctor patient records and data. [8]. It has the potential to disrupt the medical business by providing new ways to manage healthcare data, improving patient care, and reducing administrative processes. Terabytes of medical records that previously required human interpretation can now be used as machine learning input data in healthcare initiatives

Medical diagnosis in terms of more efficient and faster result is one of the most important applications of machine learning in the healthcare industry. Advances in both imaging and computers working together have resulted in a significant increase in the potential use of artificial intelligence in a variety of radiological imaging activities, such as diagnosis, prognosis, risk assessment, detection, and therapy response, as well as multi-omic approaches to diseases. [7]. The end-user or patient will benefit from early problem detection using our suggested system. Many features were used to feed the dataset into the system to get a result that would assist pregnant women in determining their fetal status. The system was created by putting various machine learning algorithms to the test, with the most accurate method being chosen.

Next, some of the most well-known research studies on machine learning techniques to predict pregnancy risks and fetal health will be shown. The cardiotocograph is the most extensively used method for detecting the fetal condition in a routine clinical examination. Jiaming Li compared the smooth voting integration approach to the stacking integration process for integrating some of the best in this case 4 best models to produce the Blender Model in this paper. The Blender Model outperformed typical machine learning models in a variety of Classification Model valuations. [9].

The Stacking EL technique, according to Pankaj Bhowmik's paper [10], can be a significant example in machine learning field to increase the accuracy of the model while the error rate reduces. Their proposed meta learner of the EL system (ensemble learning) classifier model uses the 10-fold cross- validation and that's why it was picked, with an accuracy of roughly 96.05 percent.

Yarlapati et al. used machine learning approaches to predict low birth weight (LBW) problem in pregnancies, which can be very riky for the patient and the fetal health [11]. They managed to classify fetal status as LBW or NOTLBW with almost 96.77 percent accuracy with the method 'Bayes minimum error rate classifier' on health care dataset in Indian. The fetal heart rate (FHR) is crucial in determining the fetus's health.

Jianqiang Li and his colleagues divided high one-dimensional FHR records into ten d-window segments. They used a machine learning method known as (CNN) convolutional neural network to parallelly process the dataset to increase the accuracy of fetal status evaluation [12]. Finally, the voting method was employed to classify fetal heart rate recordings. CNN had the highest classification accuracy of 93.24 percent, using their EFM system to test 3012 standards, 1024 suspect, and 437 aberrant records.

In another study [13], ultrasound scans of lumbar spines of pregnant woman's are used to determine where the needle would go. A Support Vector Machine (SVM) with a Gaussian kernel is used to identify the interspinous region. The suggested model was evaluated with 640 photographs from the second group of 16 pregnant patients after training with 800 images from 20 pregnant individuals, and it attained a success rate of 95.00 percent.

In the examination of predicting the health of a fetus [14], antepartum cardiotocography (CTG) data were used with various machine learning algorithms, including ANN, SVM, k-NN, RF, CART, Logistic Regression, C4.5, and RBFN. The uterine contraction (UC) and fetal heart rate (FHR) were extracted from the baby's heart rate, collected from the mother's abdomen, and used in classifiers. C4.5 decision tree classification method and Naive Bayes Kernel algorithm are examples of earlier machine learning research in medical sector for disease diagnosis and prediction, with applications such as predicting and displaying gestational hazards [3,4] and normal or pathological phases of pregnancy.

The connectivity between the whole brain function was examined with 105 preterm newborns in a study replicated fetal brain development [15]. The SVM approach was used to estimate connections with an accuracy of 80%. In this regard, data were provided for locations that magnetic resonance imaging (MRI) analysis could not fully interpret.

Ocak [16] used the CTG approach to analyze a UCI dataset containing uterine contractions and fetus heart rate and created a fetal health assessment model using the SVM method. The SVM's classification result is improved by allowing the genetic algorithm to remove features that doesn't contribute much from the whole dataset. His suggested method correctly predicted fetal condition as normal or abnormal 99.3% of the time and 100% of the time, respectively.

A comprehensive prediction of pathologic cases was performed using a modular neural network using a separate UCI dataset with 2126 fetal CTG recordings [17]. On the same dataset, the traditional cnn neural network method also obtains a great degree of accuracy.

Shenglong Li [41] explains that there is hardly any auxiliary diagnosis method for orthopedic disorders in his investigations on orthopedic auxiliary. In a study they co-wrote to advance the diagnosis of orthopedic diseases using wisdom, they presented the XGBoost algorithm. Nearly 150 patient cases with femoral neck data were used in their experiments. They discovered that the XGBoost model exhibits a definite advantage in running speed and has a good accuracy and recall rate in orthopedic conditions. It is clear that the XGBoost algorithm processes medical data more rapidly and effectively. The algorithms of the widely utilized Decision Tree Classification approach are best suited for medical diagnosis.

One of the well-liked and efficient classifiers employed in the current study for the classification of pregnancy data is the C4.5 Decision Tree algorithm. Lakshmi.B. N and her team's paper's [42] primary goal is to emphasize the value of standardizing the parameters chosen for data collecting in a study and analyze the performance of the C4.5 methods by comparing the results produced by the C4.5 classifier on both standardized and non-standardized datasets. With the use of 16 features, they had gathered over 600 datasets from pregnant patients. These datasets contained details about the patients' height, weight, BMI, age, and family history. It is clear from the research results on both datasets that the C4.5 decision tree classifier has a higher potential for accuracy when used on standardized data (accuracy 71%) as opposed to un-standardized data.

Fetal movement patterns are frequently used to evaluate the health of the fetus. A suitable tool to recognize and track fetal movement patterns is currently lacking. As a result, a wearable device containing an INS sensor was created in this study [43] to track fetal movement. a hybrid approach that employed CNN along with a common signal processing algorithm. With an average accuracy of 73%, the direct deep learning system identified fetal movements. The hybrid technique, which included CNN and STFT, had an average accuracy of 88 percent in identifying fetal movement. 77 pregnant women between the ages of 28 and 40 were asked to participate in this study.

In a study by RUOWEI QU [44], 155 fetal patients between the ages of 16 and 34 weeks were represented by 30,000 2D ultrasound images. To automatically distinguish the six fetal brain standard planes (FBSPs) from the non-standard planes, they suggested a differential convolutional neural network, or differential-CNN. According to the experimental findings, this approach has an accuracy rate of 92.93 percent. Their research demonstrates how the differential-CNN can be used to simplify the implementation of automatic FBSP identification.

In a study [45] conducted by Liang Xu and his team, the goal was to use genetic algorithms (GA) as a feature selection approach to choose the best feature subset from 64 FHR (fetal heart rate) data and then integrate these best features to identify unfavorable FHR patterns. The experiment included three classifiers, including linear regression, linear SVM, and RBF SVM, with one of the largest datasets, 7568. The AUC scores of the three classier cases from these cases are, respectively, 74%, 75%, and 76% from RBM SVM, which is the highest. Medicines are crucial during pregnancy and may also be harmful to the health of the fetus. FDA "category C" pharmaceuticals, which often recommended medications during pregnancy lack fetal safety recommendations. The goal of this study is to divide these medications into categories of danger and safety. The researcher on this study [46] focused on two unique outcomes: fetal loss and congenital abnormalities, and trained a random forest to categorize medications of uncertain pregnancy class into harmful or safe groups. The dataset as a result had 33,043 pregnancies without fetal loss and 14,922 pregnancies with fetal loss.

The assessment of fetal well-being is a global issue that is studied from a variety of perspectives. Some studies predict fetal state, while others concentrate on fetal premature birth rate and brain development. The preferred machine learning algorithm and dataset are the two most critical factors. Because each country has its own set of differences, the dataset and the results differ from one country to the next. Most of the research we have mentioned so far has involved physiological measurements, biochemical data, MRI images, and ultrasound images, all of which are primarily concerned with fetal development and the state of the fetus during pregnancy. These measurements are difficult to grasp from a patient's perspective and are used mainly by medical professionals. Patient focus studies are rare where they will be able to get medical assistance or advice by themselves. Akhan Akbulut [1] proposed a model implemented on an e-Health system, which is the most significant distinction between their work and the other alternatives. They used a variety of patient-specific features, such as whether the patient smokes, drinks, or has ever had a kid. These features allow patients to seek advice and answers from the comfort of their own homes. Our research in that scenario is similar because of the implementation of an e-health system and the use of datasets to support patients. However, the difference is in the dataset and machine learning method. Our research involved 349 Bangladeshi pregnant women, and since the dataset varies per country, we feel our research will be of great use. There is some accuracy comparison between different algorithms by studying all of the papers. Based on this comparison, the following table is provided:

Technique of ML	Research	No of features	Size of dataset	Accuracy
SVM	Ball et al.	N/A	105	80.2%
BMERC	Yarlapati et al.	18	101	96.77%
GP	Kenny et al.	N/A	87	98%
RF	Sahin et al.	21	1831+	99.2%
ANN	Huang et al.	23	2126+	99.5%
ule Based	Woolery et al.	214	9419	88%
SVM	Ocak	21	1831+	99.7%
LSVM	Czabanski et al.	N/A	51	90.1%
SA	Cheng et al.	14	412	88.1%
Rule Based	Woolery et al	214	9419	88%
MNN	Jadhav et al.	23	2126+	99.5%
SVM	Yu et al.	N/A	36	96.15%

Table 1: Comparison of previous related works

The table lists some of the studies described along with some of their characteristics. The dataset and the accuracy value stand out the most. It is important to note that the correctness of our study has no relationship to the studies listed on the table and neither the results. The percentage of accuracy varies depending on the methodology given for each study, the number of datasets used, and other factors. The dataset is one of the key factors in accuracy; the bigger the database, the more likely it is that accuracy will increase. As we can see from the chart, some of these studies utilized over a thousand datasets, the largest of which was 30000 and was an image dataset created using the CNN approach. A relatively common CTG dataset that many researchers have used consists of 2126 dataset with 23 features which we can see in table.

Chapter 3

Methods

3.1.1 Proposed Model

This study aims to predict the health of fetus depending on mother's health and genetic attributes. The data has been self-collected and contains 300 samples of expecting mothers. The data had a significant amount of imbalance. Some of the imbalance was tackled through undersampling which reduced the sample size to 183. The data has been standardized using standard scaler. For the remaining imbalance synthetic minority oversampling technique was used to generate artificial data points. The classifiers have been built for both balanced and imbalanced dataset to compare their performance based of the sample distributions. K-fold cross validation was used to evaluate the performance of the machine learning models being used. From the total number of samples 80% has been used for training and 20% has been used for testing the classifiers. For testing the classifiers accuracy, sensitivity, specificity, precision, F1 score, ROC and AUC have been used. The technical approach has been summarized in the visual diagram below:

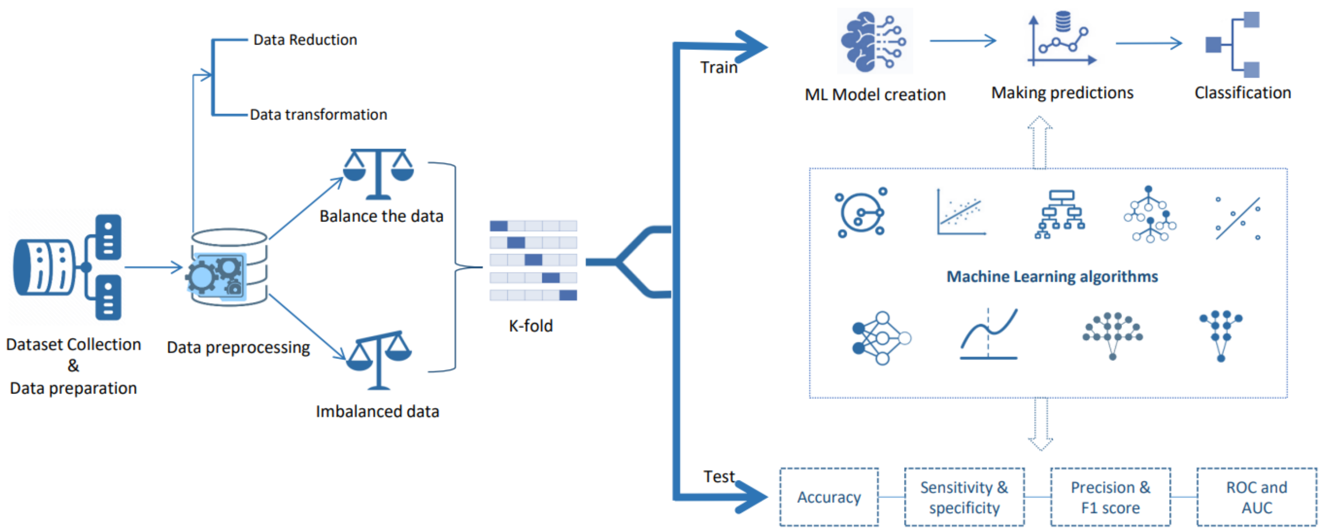


Figure 1: Technical approach visualization

The technical approach has also been summarized in the table below:

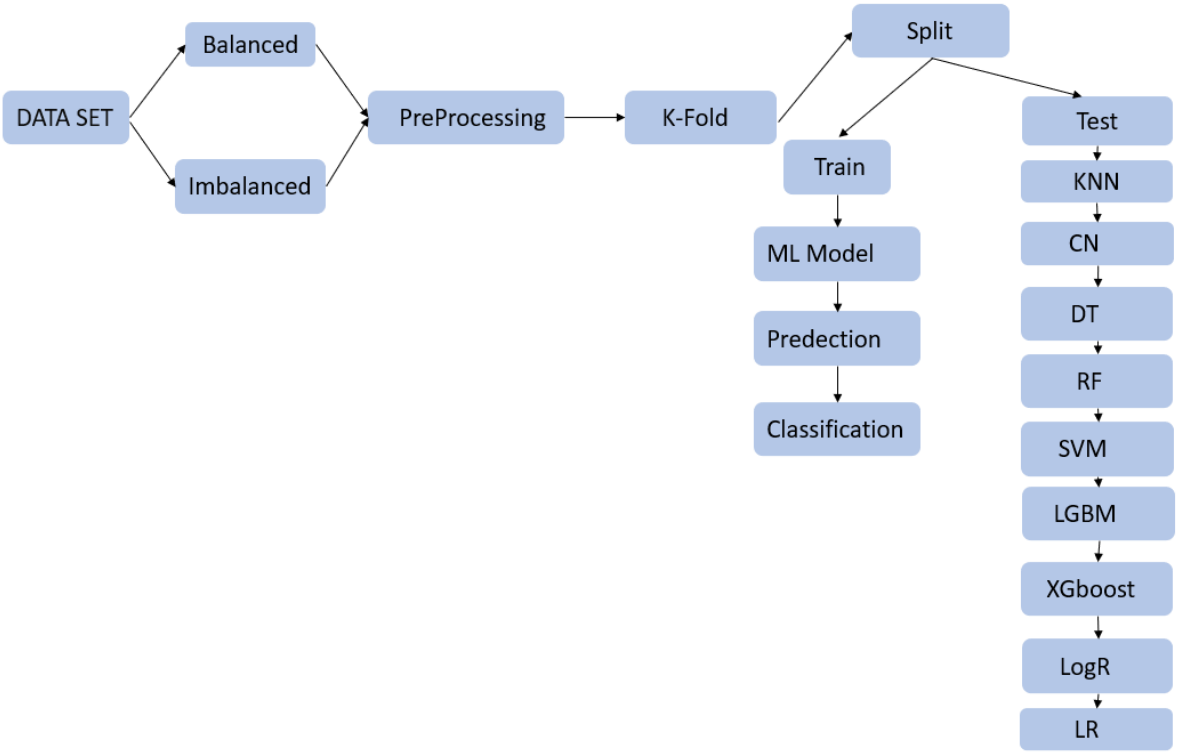


Figure 2: Technical approach diagram

3.1.2 System Architecture

The proposed system includes a website where the user can input their health details through the questions which are based on the features of the dataset. The input given by the user is fed into the machine learning algorithms which has been previously trained and saved in the server. The user will receive the result immediately after submitting the questioners. The result will be based on the model's prediction. This system will provide an opportunity to take measures as early as possible if there is any risk involving the health of the fetus. The mother's whose fetuses are at risk of having birth defects are advice to consult with their health care provider as soon as possible.

Though the system does not notify any health care provider automatically to save the patients from hassle but it can still prove to be efficient by informing the expected mothers and the families about the risks early on. This can motivate them be more aware and cautious about the complications.

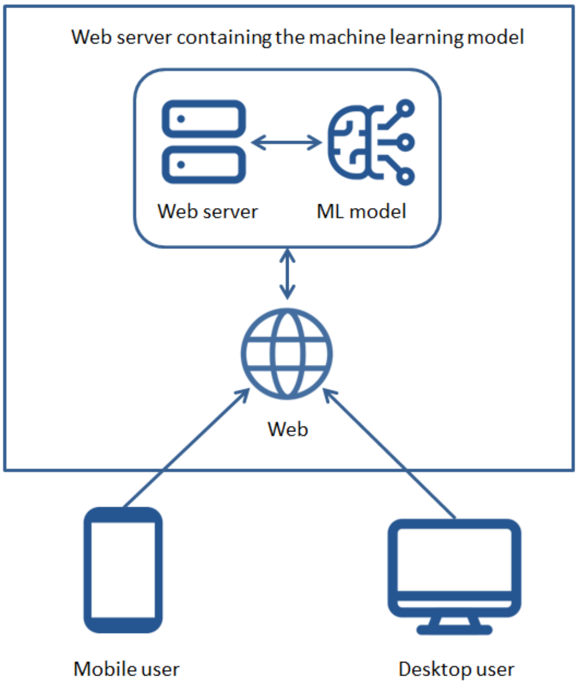


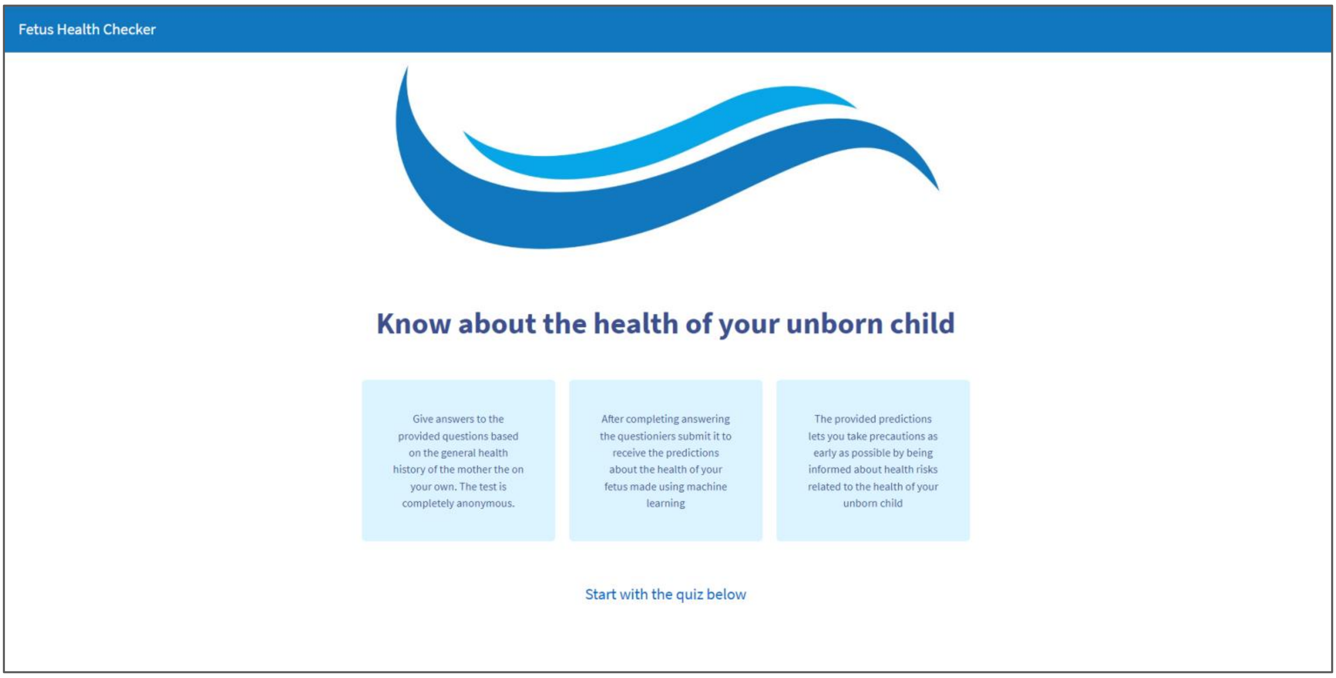
Figure 3: System architecture

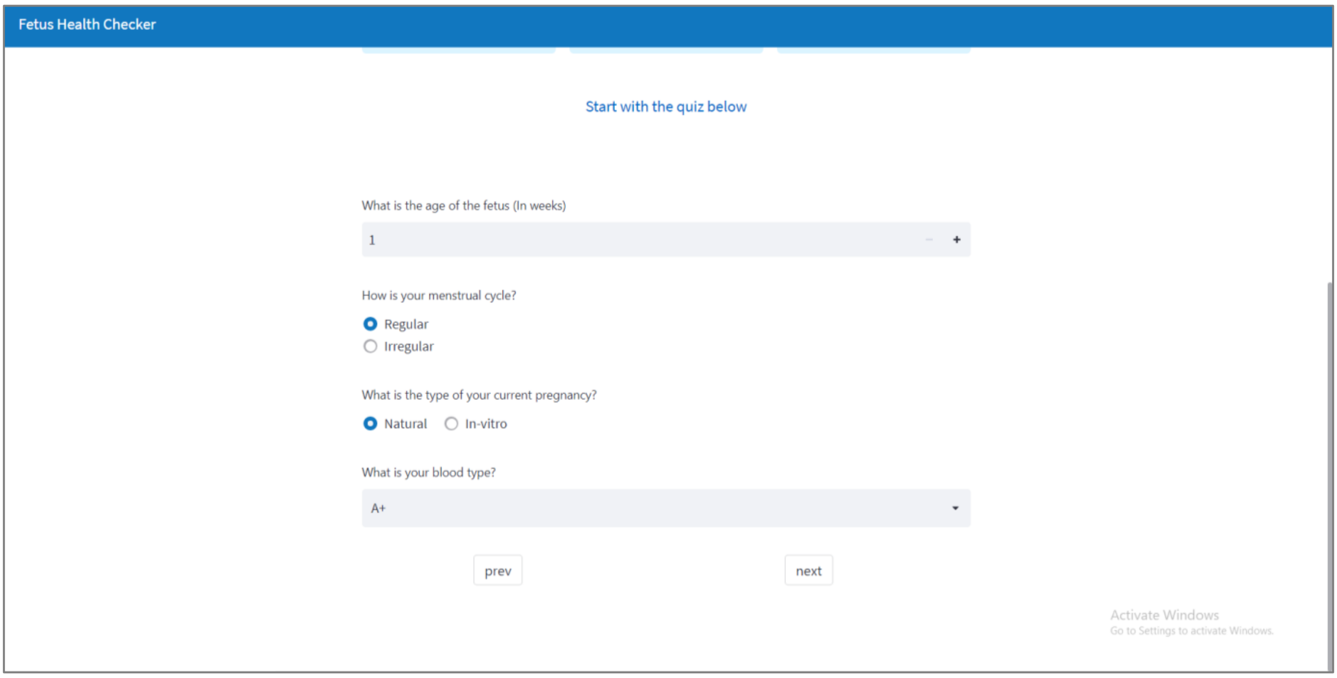
3.1.3 Web application

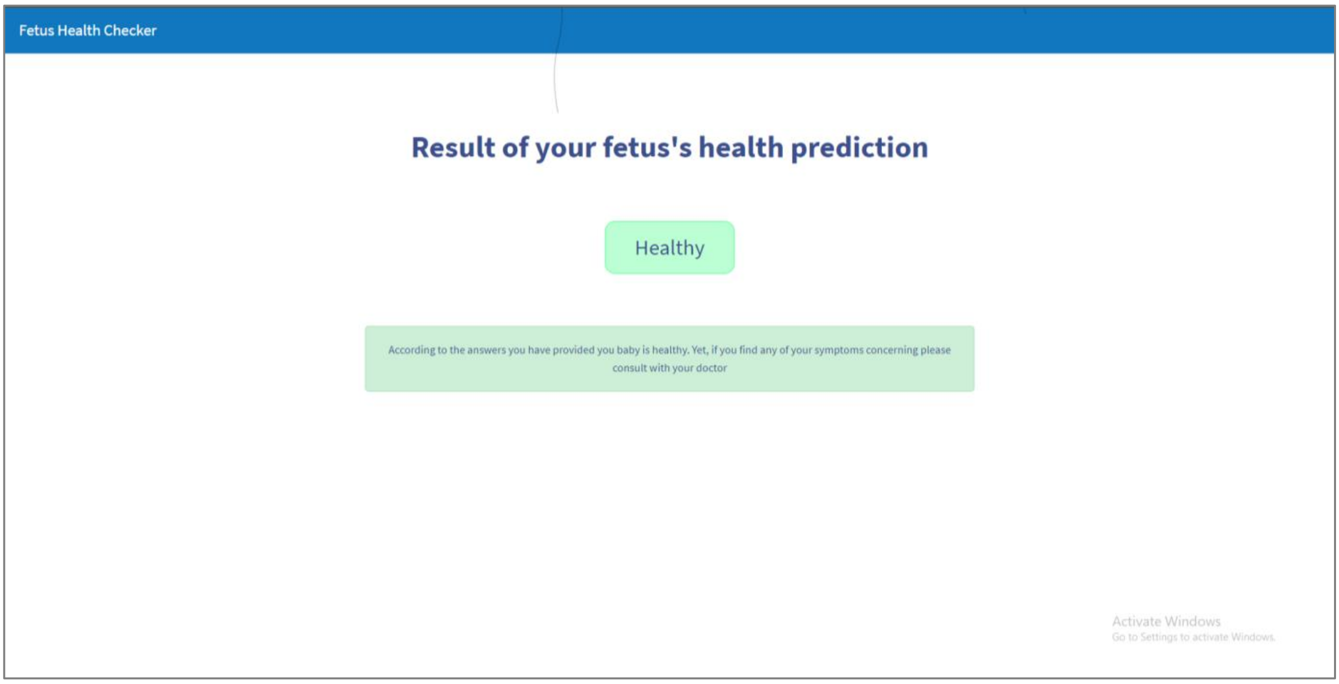
The web application is made as a system where patients can input their health details to know the health of their fetuses. The website is written using the python package called streamlit. Streamlit was chosen as it is extremely efficient for integrating machine learning model with a website. A small amount of Html and css was used to customize the interface accordingly. The website is very simple to use for any user. It allows the patients to input their data and get the results based on it instantly. The result is shown in binary. 'Healthy' for a fetus that has little to no risk of having a birth defect and 'Unhealthy' for a fetus that had quite a good amount of risk of having one. As the machine learning model is not a 100% correct on the predictions, the mothers who are predicted to have a healthy child are also advised to consult with the doctor in case the find any symptoms suspicious. The quiz is completely anonymous so it is comfortable for the users.

3.1.4 Prediction system

The machine learning model that had the most efficiency in predicting the health of the fetus while comparing the models has been saved as a python pickle file. This file is kept in the server to be loaded by the website and used for making predictions based on the input given by the users. It is a simple and effective approach for the website. As the website is written using streamlit the machine learning model and prediction system could be integrated effortlessly with the website.


Fetus Health Checker
Know about the health of your unborn child
Give answers to the provided questions based on the general health history of the mother the on your own. The test is completely anonymous.
After completing answering the questioniers submit it to receive the predictions about the health of your fetus made using machine learning
The provided predictions lets you take precautions as early as possible by being informed about health risks related to the health of your unborn child
Start with the quiz below


Fetus Health Checker
Start with the quiz below
What is the age of the fetus (In weeks)
1
How is your menstrual cycle?
Regular
Irregular
What is the type of your current pregnancy?
Natural
In-vitro
What is your blood type?
A+
prev
next
Activate Windows
Go to Settings to activate Windows.



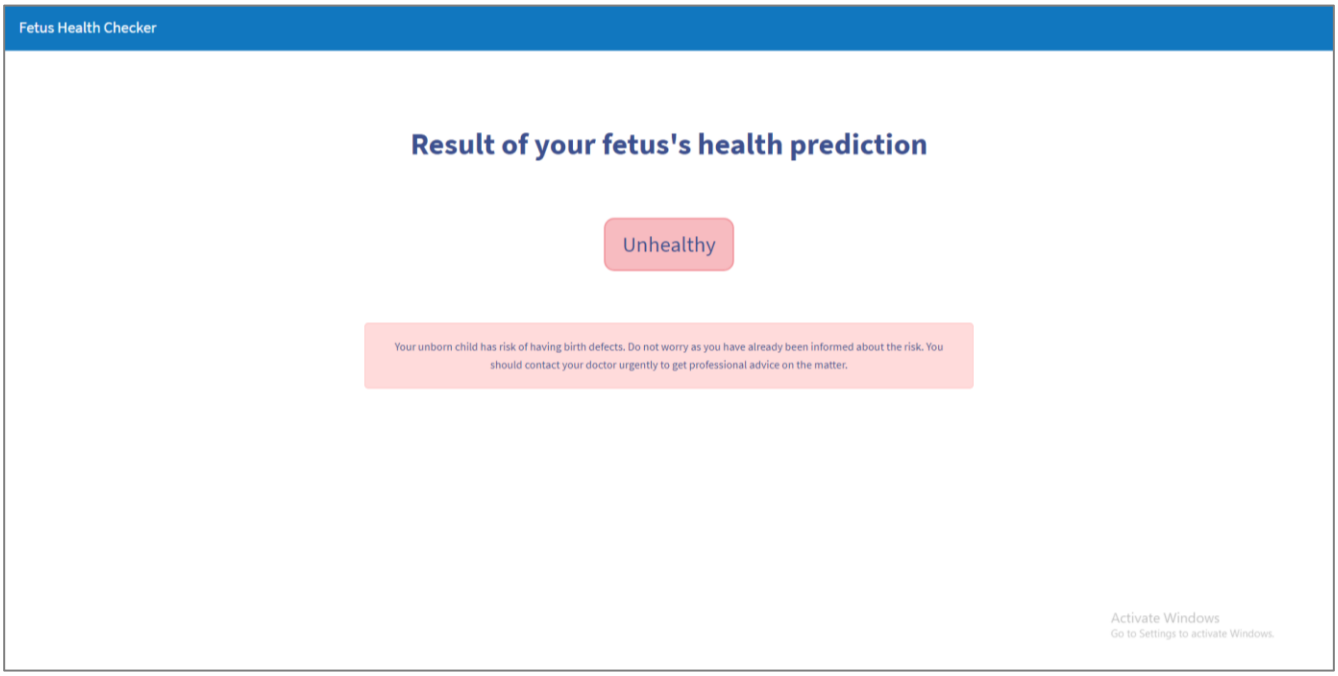


Figure 4: Website interface

3.1.5 Dataset introduction

This research was aimed towards building a machine learning model which can predict the health of a fetus based on various parameters of mother's relevant medical history and hereditary attributes. With that intend data was searched throughout the sources which provide publically available data. Within the search scope we were not able to find a suitable dataset which could satisfy the requirements. Hence, we had to collect our own dataset for training and testing our model. The data has been collected from a clinic named "Tanha hospital and health care" under the supervision of Dr. Jobeda Hasan with the assistance of her team. It is a text dataset in CSV format which contains the features along the columns and the samples along rows. The dataset has 16 unique features. The features of the data were chosen with the intention of being able to extract questions for building a questioner service to predict the health status of a fetus. For using this service expecting mothers will provide the medical, health and related hereditary history of themselves and their families to know the predicted health status of their unborn child. Majority of the features have binary values where 0 represents yes and 1 represents no. The last column belongs to the target class representing the health of the fetus where 0 represents a healthy fetus and 1 represents an unhealthy fetus. The machine learning models will use the feature values to find a specific pattern that it can learn to predict the health status for a new instance provided the values for the feature. The data originally had 348 samples in which the samples that belonged to each of the 2 classes were hugely imbalanced. Within the originally collected samples 315 samples belonged to healthy fetus class, in contrast only 33 samples belonged to unhealthy fetus class. Due to this imbalance in classes, the data did not have the ability of producing satisfactory result while used for building models. For achieving satisfactory results the data was under sampled. After under sampling 183 samples belonged to healthy fetus class and 33 samples belonged to the unhealthy fetus class. The dataset achieved through under sampling was used for running all of the tests for this research. The dataset has no missing values. For some of the features the value is negative for all of the collected samples such as "Presence of disabled child in father's family", "Tobacco/alcohol consumption" which has to be taken into consideration while making predictions.

Features	Type	0	1	Minimum	Maximum
Fetal age	Numeric				
General menstrual cyclic status of mother	Binary				
Type of pregnancy	Binary				
Blood serotype of mother	Numeric				
Past delivery	Binary				
Abortion status	Binary				
Diabetic's history	Binary				
Hypertension	Binary				
Other illness	Binary				

Operation history	Binary				
Presence of disabled children	Binary				
Presence of disabled children in father family	Binary				
Covid history	Binary				
Previous child home or hospital delivery	Binary				
Any other illness	Binary				
Fetal heath Status	Binary				

Table 2: Dataset Summary

3.1.6 Data Visualization

Data visualization provides a way to observe patterns, trends and outliers through visual representation. It is an effective way to understand the overall behavior of the dataset. The dataset used in this research contains data mostly in a binary format having a 0 corresponding to a yes and 1 corresponding to a no, only exception being fetal age and blood serotype of mother. To visualize fetal age range of the collected samples a histogram has been created. For the features having binary values pie chart has been created as it offered observable and useful visual representation.

3.1.7 Histogram

Histogram is a widely used graph for data visualization. It groups discrete or continuous data into logical bin. Histogram is a convenient tool for visualizing the distribution of data present within each bin or interval. In our data majority of the features have binary values except for ‘Fetal age’ and ‘Mother’s blood serotype’. To visualize these two features effectively, histogram is the optimal choice. The histogram of ‘Fetal age’ provides an illustration of how many fetuses belong within a specific interval. In our dataset, fetal age ranges from 5-40 weeks. The plotted histogram has an interval of 5 which produced 7 logical bins. Analyzing the histogram shows that among the 183 samples around 38 samples have fetal age ranging from 10-15 weeks, which is the maximum number of fetus belonging to a specific interval. The second majority belongs to the interval 15-20 weeks with 34 samples belonging to this range. Majority of the samples lies between the fetal ages of 5-20 weeks which represents the initial weeks of fetus’s lifetime. Among the fetus age range of 20-40 weeks, 17 samples belong to the age range 25-30 which is the highest count. Samples counts are lower around the later weeks of the fetus’s lifetime. By analyzing the histogram, it can be observed that the samples cover the almost totality of a fetus’s lifetime starting from the detection till birth. The dataset has a decent amount of samples belonging to each of the age range. The histogram has been plotted with a kernel density estimator which contracts a curve to represent a smooth version of the histogram. It fits the curve over the population providing information about the distribution.

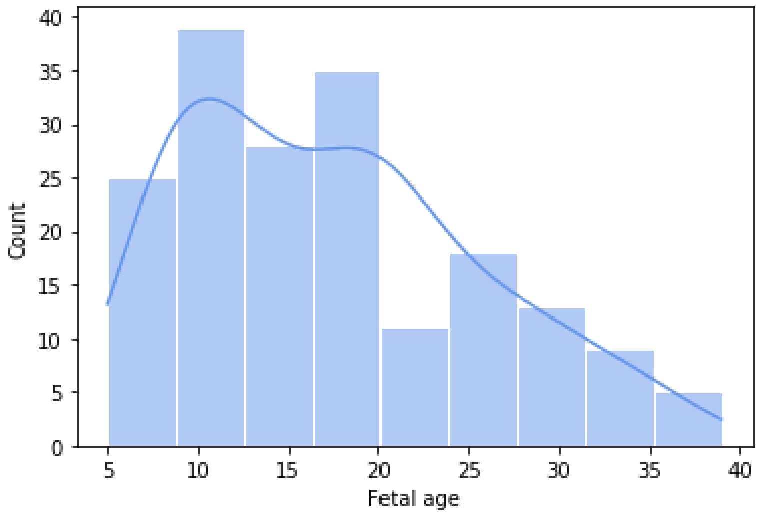


Figure 5: Fetal age histogram

For visualizing the distribution of the feature “Blood serotype of mother” another histogram has been plotted along with the kernel density estimator. For this feature the values along the x axis are discrete. There are 7 blood types in total. By analyzing the histogram, it can be seen that blood type 0,5 and 6 ha the least amount of samples associated with them. Majority of the samples have blood type 1 and 4 for which the count is 60 and 55 respectively. After these two serotypes, 39 samples belong from blood type 2 and 25 samples belong to blood type 3. The rest of the three blood types 0,5 and 6 has less than 10 samples associated with them. According to a study by Rai et al. based on 5098 people from Sikkim India the frequency of blood group A (35.34%) was found to be the highest, followed by blood group O (35.18%), B (21.99%) and AB (7.49%). The results also indicated that 99.47% of individuals were Rh positive and 0.53 % was Rh negative. In our sample, according to the distribution, 34% samples belong to the blood group O which is the highest, second highest percentage is for the blood group A with 32% sample displays the sample distribution being congruent to the findings of the research paper by Rai et al. which implies that our dataset has an appropriate distribution among all possible blood types being associated with it. Blood group B and AB has 21% and 12% samples respectively. If Rh factor is considered, 95% samples belong to positive Rh factor and 4% belong to the negative Rh factor. The analysis our dataset does not contain any sample belonging to B negative blood group.

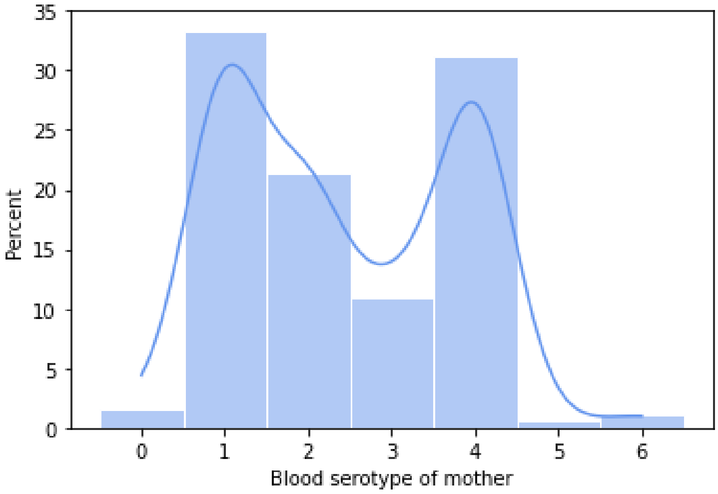


Figure 6: Blood serotype of mother histogram

3.1.8 Bivariate histogram

The bivariate histogram shows the distribution while comparing two sets of data to find a relationship between the two variables. The distribution depicted by bivariate histogram enables the analysis of one variable in relation to the other. For our data set, a bivariate histogram has been created using “Fetal health status” along the x axis and “Fetal age” along the y axis. Fetal age has been divided into logical bins and the quantity of samples in each bin corresponding to a health status has been depicted using a heatmap. With more samples being in each of the bins, the color gets darker and with less samples the color gets lighter. Based on the concept of bivariate histogram the relation between fetal age and fetal health status can be analyzed. The plot shows that most samples that belong to the healthy class has fetal age ranging from 5-20 weeks. The heatmap for unhealthy fetus class is significantly lighter due to the amount of samples belonging to unhealthy fetus class being lower. Within the fetal ages belonging to the unhealthy fetus class the distribution is similar to that of the healthy fetus class. Majority of the samples are within the age range of 5-20 weeks for the unhealthy fetus class as well. The unhealthy fetus class does not contain any sample from the age range 25-40 weeks.

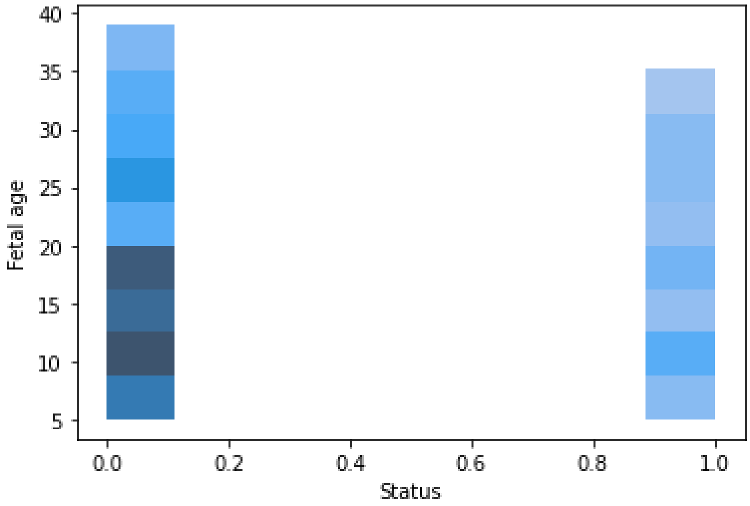


Figure 7: Fetal age and status bivariate histogram (Heatmap)

Another plot is created using the two variable "Fetal age" and "Status" to visualize the percentage of samples belonging to each of the class depending on fetal age range. Through analyzing the graph a similar pattern can be observed as the (graph-3) which was a heatmap representation based on the relationship among the stated variables. The imbalance within the dataset is conspicuos. Majority of the samples among various age ranges belong to the healthy fetus class. Another key observation is the unhealthy fetus class having a decent distribution, containing few samples belonging to wach of the age range. Similar to (graph-3), unhealthy fetus class does not contain any samples within the age range 35-40 weeks. Kernel density estimator has been applied to visualize the distribution through the smooth curve fitted around the histogram.

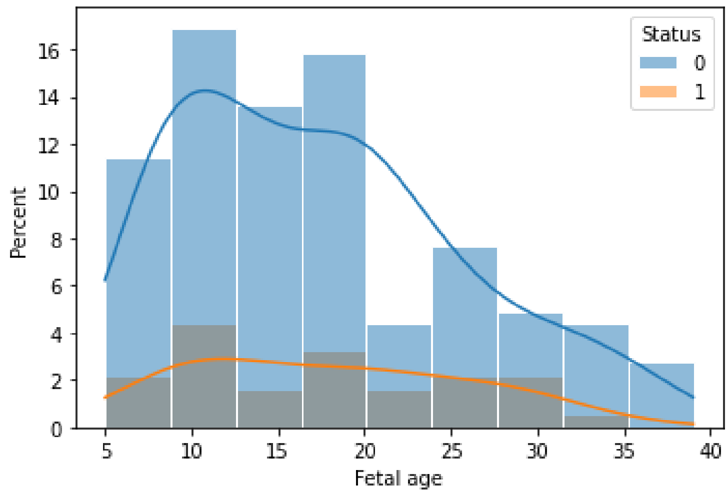


Figure 8: Fetal age and status bivariate histogram

3.1.9 Corelation heatmap

Corelation heatmap provides a visual representation of the linear relation and the intensity of the relation among any two variables. The value representing the corealtion ranges from -1 to +1. The higher the value the stronger is the corealtion. If the corelation value is positive the variables are positively corelated and if the value is negative the variables are negatively corelated. It simplifies the corelation and eases visulization process. By analysing the heatmap produced by the dataset used in this paper, it can be observed that fetal health status is positively corelated mostly with delivery status, previous child delivery location, diabetics, type of pregnancy, hypertension and fetal age. It is negatively corelated with general menstrual cycle, covid history, presence of disabled children and presence of disabled children in mother’s family.

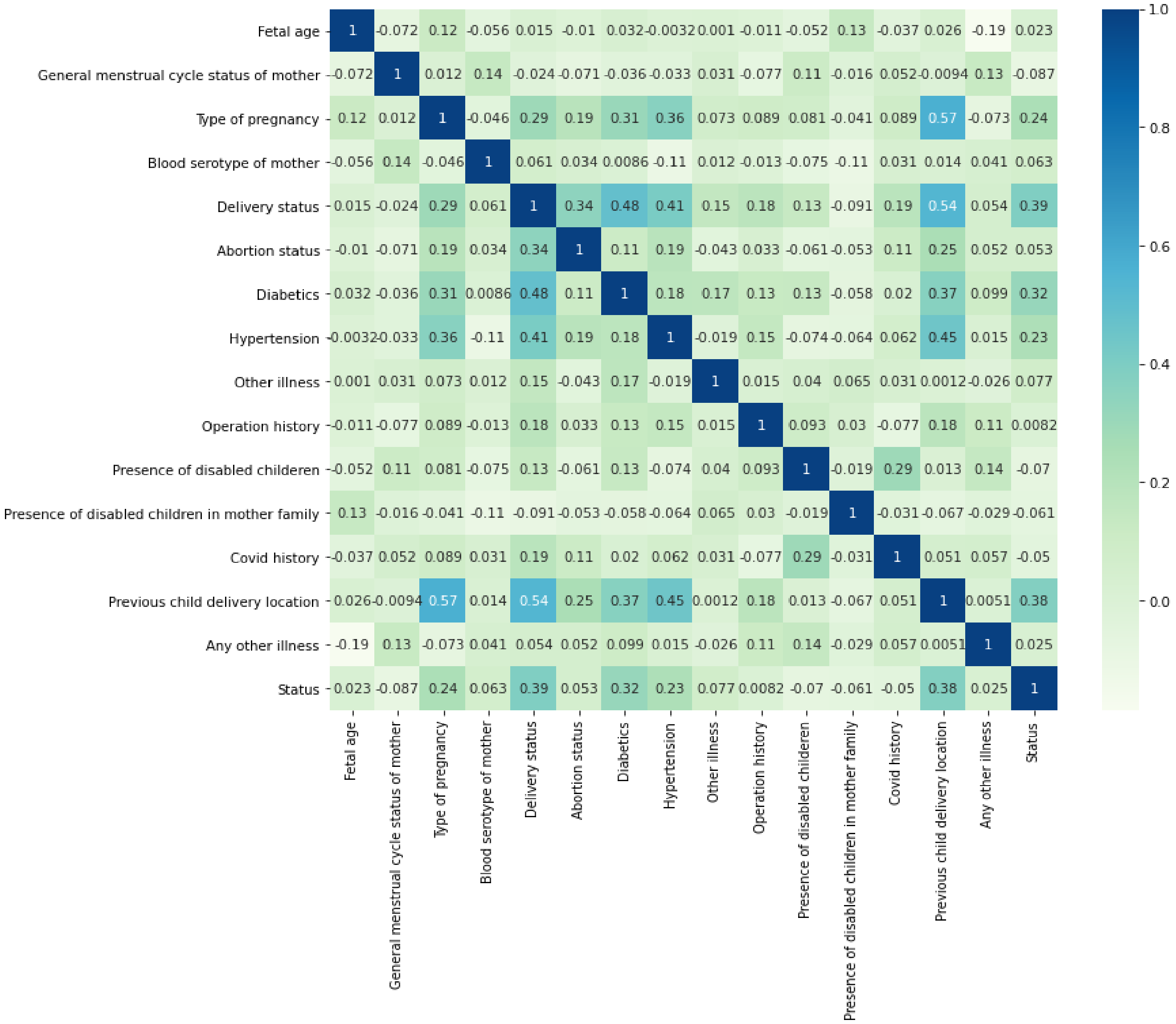


Figure 9*:* Correlation Heatmap

3.1.10 Data preprocessing

Any type of processing or manipulation carried out on raw data before putting the data to use in order to achieve improvement in performance is known as data preprocessing. Data can be preprocessed through cleanse, integration, transformation and reduction of the data.

Data cleaning is carried out to handle missing or incomplete, noisy and inconsistent data. For missing data, the data points can either be removed or filled manually by calculating the mean value or the value with highest probability. For handling noisy data methods like binning, clustering, regression are used. Binning works by separating the data in multiple bins and computing the bin's mean, median, boundary value as to use it as a replacement for the noisy data. Clustering works by creating clusters of similar data points from which the outliers can be detected and removed. Regression method fits one or multiple lines along the data points to smooth out the noise in the data.

Data integration is used for combining data from various sources and to store them in an integrated and consistent manner. Tight coupling and loose coupling are the methods used for data integration. During tight coupling data from different sources are merged into single source. For loose coupling data is transformed using queries from the user while keeping the original source unchanged.

Data transformation is used for transforming the raw is data in order to increase comprehensibility, coherence and improves performance. Some of the techniques used for transforming the data are smoothing, normalization, aggregation, attribute selection, discretization, and generalization.

Data reduction performed to reduce the overall size of the data while preserving the core behavior and pattern present in the raw data in order reduce storage requirements. Some methods that used for data reduction are data cube aggregation, data compression, dimensionality Reduction

The data we have used in our research has been self-collected. It does not have any missing values so, data cleansing was not required. Data has been prepared in a way which makes all the analysis possible. The data has quite substantial amount of imbalance which is an obstruction in attaining desirable results. After reducing the data, the features and the target class has been separated to prepare the training and testing dataset and labels. Afterwards the data was balanced using SMOTE With intend of resolving the negative impact of the imbalance present in the data. As the algorithms were applied on both the balanced and the imbalanced dataset both the reduced dataset and oversampled dataset was trained and tested separately. The results were compared based on the accuracy, sensitivity, specificity, precision, F1 score, ROC and AUC score.

3.1.11 Undersampling

Undersampling is a data transformation method applied on imbalanced dataset where some of the data belonging to the majority class is removed to minimize its difference with the minority class. Undersampling can be both random or through the use of an algorithm. Some of the most widely used under sampling techniques include Tomek Links method, edited nearest neighbors (ENN), One-sided selection (OSS). The dataset used in this research had a substantial amount of imbalance. The unhealthy fetus class is comprised of 11% of the total 300 samples. An imbalance of this scale wouldn't yield usable results. Before the application of other data transformation techniques the data had been undersampled. Some of the samples belonging to the original dataset were removed randomly. The amount and the samples to be removed were decided imperially through trial and error. In the end 183 samples among the 300 were as the base dataset. After the application of undersampling the majority class had 150 samples while the minority class was kept at 33 samples.

	Samples in class 0	Samples in class 1
Before undersampling	267	33
After undersampling	150	33

Table 3: Comparison of sample before and after undersampling

3.1.12 Standard Scaler

Standard scaler is a data transformation method used for standardizing the data. The range or units of the data can be different based on the feature. Having uneven distribution in the dataset can cause some machine learning models to perform poorly. Standard scaler transformation is applied to achieve a standard normal distribution where the mean is 0 and standard deviation is 1. Standard scaler works by subtracting the mean from sample and dividing it by the standard deviation. The features in our dataset have been standardized using the standarScaler module from Sklearn. If standard scaler score is Z it can be defined as:

$$Z = \frac{sample - mean}{standard\ scaler} \tag{3.1}$$

3.1.13 SMOTE

SMOTE or Synthetic Minority Oversampling Technique is an effective algorithm used for handling imbalanced data. It was proposed in 2002 by Chawla et al. SMOTE augments the minority samples by creating synthetic data. The creation of synthetic data points based on the pattern from the original data points prohibits the addition of new information, it balances the data while preserving its original traits. SMOTE uses K-nearest neighbor algorithm to generate synthetic data for minority class. The number of neighbors considered while using KNN is randomly chosen based on the amount of oversampling needed. Number of neighbors or K value was set to five in the original paper. The size of majority class is used as the stopping point so that, the oversamples are created in an amount that gets the minority class to be equal to the majority class. SMOTE works by calculating the nearest neighbors of the data points from the original data set and compares it to the feature vector generated from the original sample. The difference between the two is then multiplied by a number between 0 and 1 and the resultant number is added to the original feature vector. This causes the selection of a random point along the “line segment” between the features. (Chawla et al) Through this process the selected new data points are more general and similar in pattern.

In the dataset used in this research after undersampling the minority class consists of only 20% of the total samples. Presence of such imbalance makes the performance matrices such as: accuracy, sensitivity, specificity, precision and recall were not being purposeful. For removing the imbalance SMOTE was used to create synthetic samples of the minority class. The library required for SMOTE was imported from scikitlearn. Before applying SMOTE the minority class had 33 samples while the majority class had 150 samples. After smote has been applied both the majority and minority class had 150 samples which creates 300 samples in total.

	Samples in class 0	Samples in class 1
Before applying smote	150	33
After applying smote	150	150

Table 4: Comparison of samples before and after using SMOTE

3.1.14 Train-test split

Machine learning works by learning patterns from a set of known data and applying the rules extracted from the patter to predict the outcome of unknown data. Before the model operates on real world data it is tested on a set of unknown data for which the outcome is known and stored separately. The machine perceives the data similarly to real world data as the actual outcomes are not provided in advance. The prediction made by the machine is compared with the actual value which was previously stored separately to evaluate the performance model. The aim with using such data is to mimic the real world for analyzing the models capability. The known set of data the model is trained on is called training data and the data used for evaluating the model is known as testing data. Testing data is nothing but a part of the actual data for which the outcome is not provided beforehand. The ratio of train-test split can impact the quality of the model. Decrease in training data can

It is crucial data preprocessing step to split the actual data into training and testing set for the model to be trained and evaluated. For the purpose of this research 80% of the original data is used as training data and 20% is used as testing data. For the imbalanced dataset the training data had 146 samples and the testing data had 37 samples. For the balanced dataset the training data consisted of 240 samples and testing data consisted of 60 samples.

	Training data	Testing data
Imbalanced dataset	146	37
Balanced dataset	240	60

Table 5: Comparison of train-test split among balanced and imbalanced dataset

3.1.15 K-fold Cross validation

K-fold cross validation is a method for evaluating machine learning models. When there is a limitation due to the size of the dataset, K-fold cross validation can prove to be an effective method in determining and comparing the performance of various machine learning models. K-fold works in iteration by splitting the dataset in sets depending on the value of K. Each split is known as a fold. Among the folds, 1 part is considered the testing data and the rest of the K-1 parts are considered training data. For each of the iteration, different combination of the folds will be used as training and testing data. This ensures that the distribution of samples is altered among the training and testing dataset to evaluate the efficiency of the model. A model can perform well at one specific split than the other. Training and testing the model at different split and averaging the accuracy provide a better understanding of how the model might perform in real world scenario. The value of K controls the number of folds as well as the number of iterations. There is no specific rule or guide of choosing the value of K. But ideally it is a number within 3-10. The value of K is chosen empirically through trials using various possible values. The larger the K value the more combinations will be made.

As our dataset had limited samples K-fold was applied to evaluate the efficiency of the 9 algorithms compared in this paper. K-fold was applied to the dataset before and after applying SMOTE. The dataset before applying SMOTE is regarded as the imbalanced dataset and the dataset after applying SOME is regarded as imbalanced dataset. The value of K was chosen to be 10 empirically. The imbalanced dataset has 183 samples which were divided into 10 different folds. The first 7 folds contain 18 samples and the last 3 contain 19 samples. During each of the 10 iterations, 1 fold was used for testing while the other 9 folds were used for training. The balanced dataset has 300 samples. Each fold contains 30 samples.

The accuracy was calculated for each of the iteration and averaged to get the final accuracy score. K-fold was applied to both the balanced and imbalanced dataset for 9 different algorithms. Based on the accuracy scores, CNN has the highest accuracy score for the imbalanced dataset with 0.87 accuracy score and SVM has the least accuracy score of 0.75 similarly, also for the balanced dataset CNN had the highest accuracy score of 0.78 and linear regression has the lowest score of 0.6

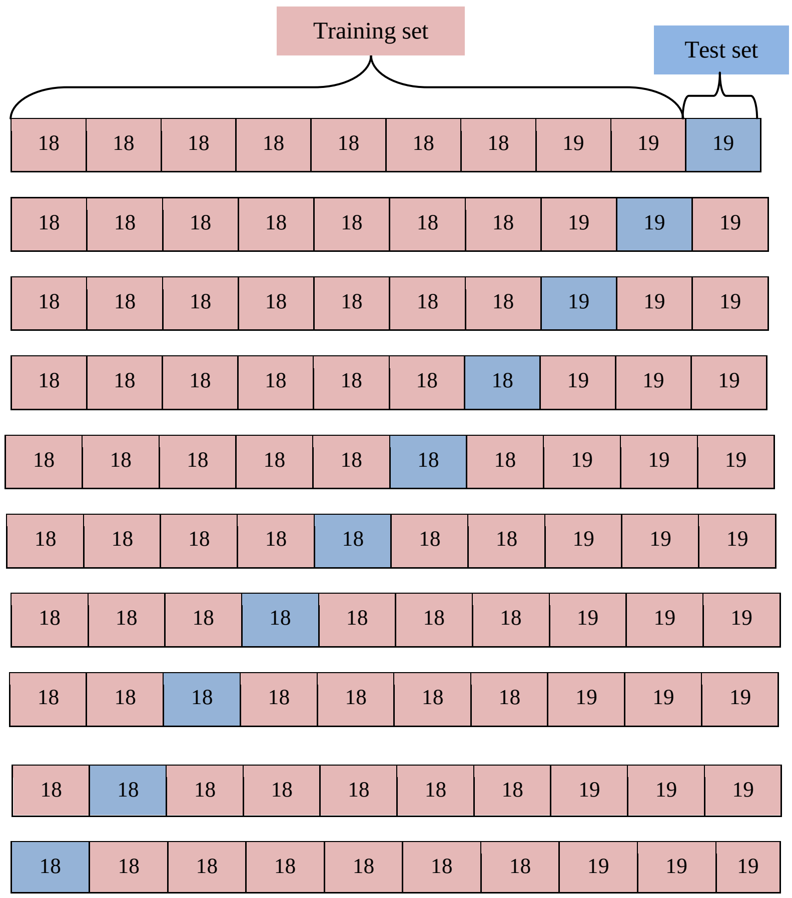


Figure 10: 10-fold cross validation on imbalanced dataset

Algorithm	K-fold score (Imbalanced)	K-fold score (Balanced)
KNN	0.83	0.71
Linear regression	0.76	0.61
Logistic regression	0.78	0.72
Decision tree	0.76	0.76
Random Forest	0.83	0.73
SVM	0.75	0.68
Lgbm	0.80	0.72
XGboost	0.77	0.69
CNN	0.87	0.78

Table 6: 10-fold cross validation score for balanced and imbalanced dataset

Algorithm

A process for fixing a hassle or appearing a computation is referred to as an algorithm. Algorithms are a particular set of commands that perform special moves in hardware and software, mostly on totally based routines. Algorithms are on the coronary heart of computing [21]. Observing our environment exhibits numerous algorithms at paintings to resolve each day problems: Algorithms strength packages together with social media networks, GPS packages, Google search, e-trade platforms, Netflix advice systems, and so on. In arithmetic and laptop science, a set of rules is a truthful method for fixing a repeating problem. Algorithms also are applied as facts processing requirements and play a critical component in computerized systems [18]. A set of guidelines can be used to type records gadgets or do more complex tasks, along with promoting patron content material fabric on social networking platforms. Initial enter and commands that specify a selected computation are popular in algorithms. We used nine algorithms for our implementation: KNN, Linear regression, Logistic regression, Decision tree, Random Forest, SVM, Lgbm, Boost, and CNN. All algorithms are crucial for our implementation, and the result is pretty satisfactory. Some explanations about all algorithms are given below:

K-Nearest Neighbors

The K-Nearest Neighbour set of rules is primarily based totally on the SL method and also based on ML algorithms. Because the KNN set of regulations is non-parametric, it makes no assumptions about the information [33]. "K" represents the number of nearest neighbors of a new unknown variable to be predicted or classified [35]. As further information is generated, the KNN set of rules can speedily classify it into an appropriate class [15]. During the preparing stage, the KNN set of rules stores the dataset, and as unused the data is obtained, it classifies them in an exceptionally comparative type with today's information [17]. The KNN set of rules may be a straightforward, administered contraption acing strategy that can be utilized to bargain with classification and relapse issues [10]. It is fundamental to present and use, but it faces the challenge of becoming significantly slower as the total amount of information in use proliferates. Its goal, like first falls, is to find the closest neighbors among all the previously undiscovered data points so that it can determine which class it corresponds to. As a result, it is an interspace-based approach. In our paper, we used various algorithms, and we found no method suitable for determining the precise value of knn.KNN provides the highest accuracy for our training and testing datasets [38]. If we take a case study from the actual world, Bangladesh is a democratic nation; therefore, many different types of parties participate in elections, albeit winning is difficult. During decision campaigns, political parties compete for voter support. Open votes for the candidate with whom they have the strongest enthusiastic association. When all the candidates' votes have been counted, the candidate with the most votes is declared the election's winner. For regression, knn engages the mean-average method for predicting the value of new data and the value of knn considering all of the closer neighbors

[32]. When forecasting the value of incoming data and the value of knn when taking into account all of the closest neighbors, knn uses the mean-average method for regression [40]. In our thesis paper, KNN achieves an accuracy of 86% in a balanced dataset while it achieves an accuracy of 80% in an unbalanced dataset. Below is provided the knn picture diagram:

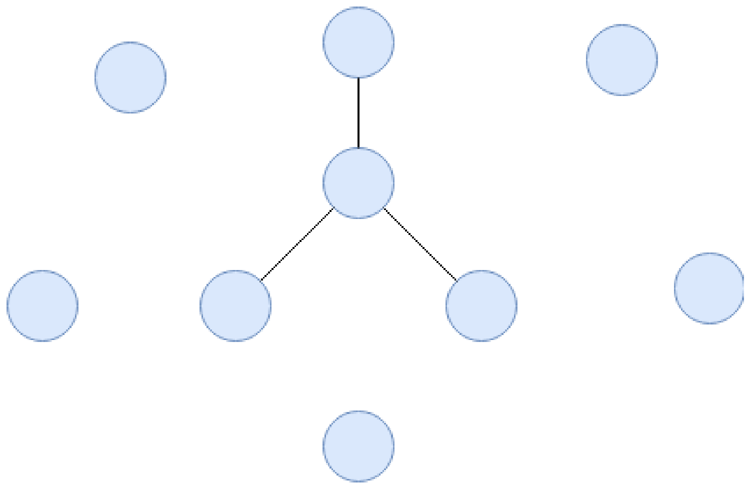


Figure 11: KNN Algorithm

Linear Regression

In the handiest terms, Linear Regression is a supervised Machine Learning version that determines the high-quality healthy linear line among the impartial and structured variables or the linear courting among the two [38]. Regression fashions are available in a lot of shapes and sizes [11].It’s a statistical approach for appearing predictive analysis. Period, Deals, outcome price, earnings and different numeric variables have anticipated the usage of linear regression [10]. The well-known set of linear regression rules shows a unbent courtship between a structured variable and one or more unbiased variables, hence its name. Because LR shows a unbent courting, it determines how the cost of the impartial variable adjustments, and the structured variable adjustments as well [20]. A diagonal instantly line represents the hyperlink among the variables withinside the linear regression version. The correlation coefficient, which ranges in value from -1 to 1, is a useful numerical measure of the link between two variables [33]. It expresses the degree of association between the observed data for the two variables. A mathematical representation of LR is given by the equation $y = a + bx$, where x is the explanatory variable and Y is the dependent variable. A line's intercept (the value of y when $x = 0$) equals the slope of the line, which is b [39]. There are two types of linear regression models:

- SLR (Simple Linear Regression): SLR is a regression model that has one dependent and one independent variable.
- MPR (Multiple Linear Regression): MPR could be a relapse show with more than one free variable and one subordinate variable.
- In our thesis paper, LR achieves an accuracy of 78% in a balanced dataset while it achieves an accuracy of 70% in an unbalanced dataset. Below is provided the LR picture diagram:

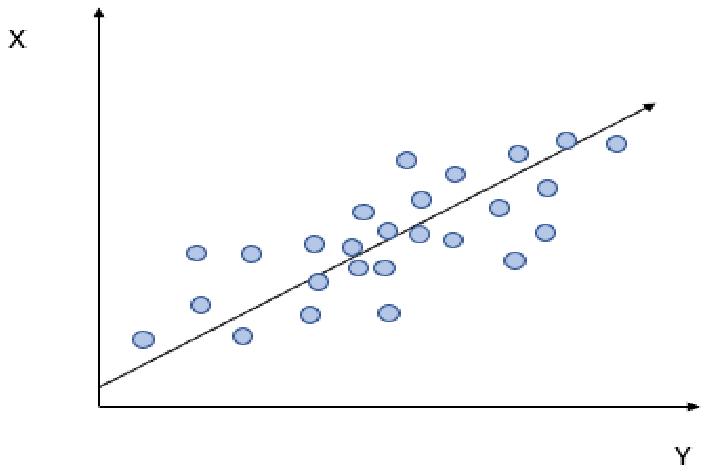


Figure 12: LR Algorithm

Decision Tree

The DT (Decision Tree) set of rules is most of the maximum truthful and extensively used type techniques. The supervised mastering strategies consist of the Decision Tree set of rules. The choice tree approach, not like different supervised mastering algorithms, can resolve regression and type problems as well [23]. We begin from the basis of the tree while the use of DT to expect a history elegance label. The values of the root attributes are in comparison to the record's attribute's values [7]. Based at the comparison, we leap to the following node with the aid of using following the department that corresponds to that fee. The root is first notion to be the whole education set [11]. Implementation of decision tree in Python:

- Utilizing the training set of data to fit a Decision-Tree algorithm
- Step before processing
- depicting the results of the test set
- appraise test result confirm result precision (Ccm)

The choice is available for particular work values [22]. If the statistics is continuous, it should be discretized earlier than the version may be built. The decision tree is based on the highlights of the provided data set. It could be a graphical representation of all conceivable arrangements to a problem/decision based on given conditions [15. The algorithm compares the attribute value with the other sub nodes for the next node and continues. Continue the process until you reach the leaf node of the tree [28]. The Benefits of the Decision Tree is that it requires less data cleaning than other algorithms and also because it uses the same procedure that people do in real life to make decisions, it is simple to comprehend. In our thesis paper, DT achieves an accuracy of 67% in a balanced dataset while it achieves an accuracy of 81% in an unbalanced dataset. Below is provided the LR picture diagram:

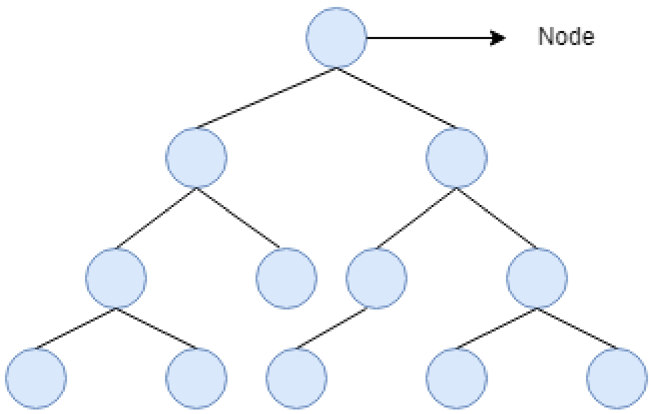


Figure 13: Decision tree algorithm

Random forest

Random Forest may be an administered machine learning calculation that's commonly utilized to illuminate classification and relapse problems [18]. The name "Random Forest" comes from the truth that each tree within the Random Forest is developed by haphazardly selecting a test of the data. It makes choice trees from different tests, utilizing the lion's share vote for classification and the normal for regression. When we have a large data set and interpretability is not a big issue, Arbitrary RF could be a great choice [3]. Decision trees are a parcel less demanding to studied and comprehend. As an unpredictable Random Forest comprises of a few chosen trees, moving it continuously gets to be a challenge. Finance, e-commerce, and health care are examples of random forest uses. Its fundamental impediments are that it is more complex, that it requires more assets, that it could be a time-consuming handle, which its fundamental benefits are that it is more effective [1]. It's simple to figure out which features are most important, and there's less risk of imbalanced datasets. Data scientists use random forest on the job in many industries, stock trading, including banking, medicine, and e-commerce [12]. RF is utilized in hereditary investigate and computational science.

In our thesis paper, RF achieves an accuracy of 89% in a balanced dataset while it achieves an accuracy of 75% in an unbalanced dataset. Below is provided the RF picture diagram:

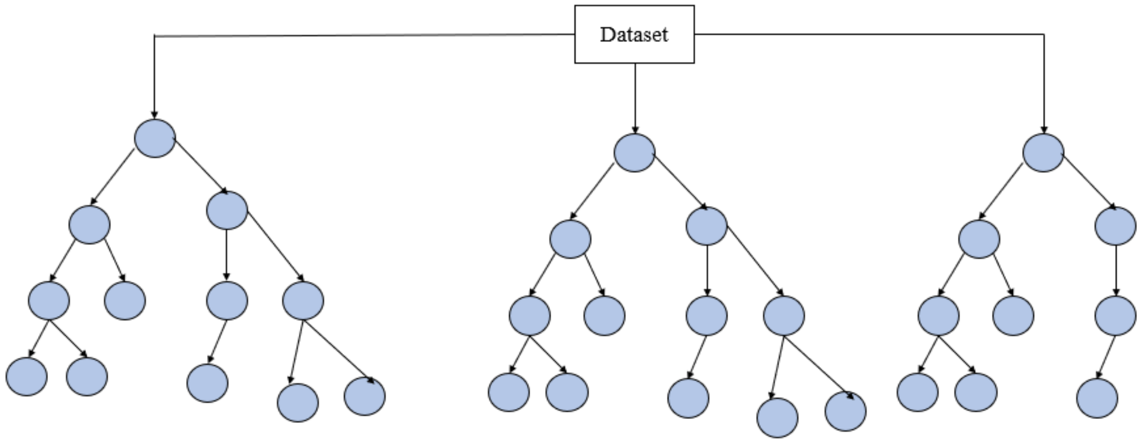


Figure 14: Random Forest Algorithm

Support Vector Machine

The SVM (support vector machine) is a supervised device that applies category techniques to solve issues with -institution category issues [23]. After being given units of classified education statistics for each category, SVM models can be used to categorize new textual content. It has a number of advantages, including the ability to use a neural network to increase speed and overall performance with a small sample set. It was created in 1963. Corinna Cortes and Vapnik proposed the "smooth margin" version in 1993, and it was changed to posted in 1995 [13]. SVMs are used to address a wide range of real-world international concerns, such as SVMs are beneficial in textual content and hypertext classification in both inductive and transudate circumstances since they can dramatically reduce the demand for classified education instances' can also be used to categorize images. According to test statistics, SVMs offer far better seek accuracy than common question [13]. The application of supervised SVM is categorized as satellite statistics, which includes SAR statistics [28].Information that can be directly isolated can be classified using a vector reinforcement machine. If it isn't directly divisible, able to make it work utilizing the bit strategy [33]. To just stick with a linear kernel is preferable for text classification. Benefits of sv machines are When there is an understandable margin of dissociation between classes, the support vector machine performs comparably well and also more effective in large spaces [35]. Furthermore, it is useful when the number of dimensions is greater than the number of specimens [36]. Support vector machine is memory systematic in comparison [30]. The Svms algorithm is sometimes not suitable for large datasets and does not perform well when the dataset contains more noise e.g., the target classes overlap. Since the support vector classifier works by placing data points above and below the

classification hyperplane, there is no probabilistic clarification for the classification [40].
In our thesis paper, SVM achieves an accuracy of 81% in a balanced dataset while it achieves an accuracy of 66% in an unbalanced dataset. Below is provided the SVM picture diagram:

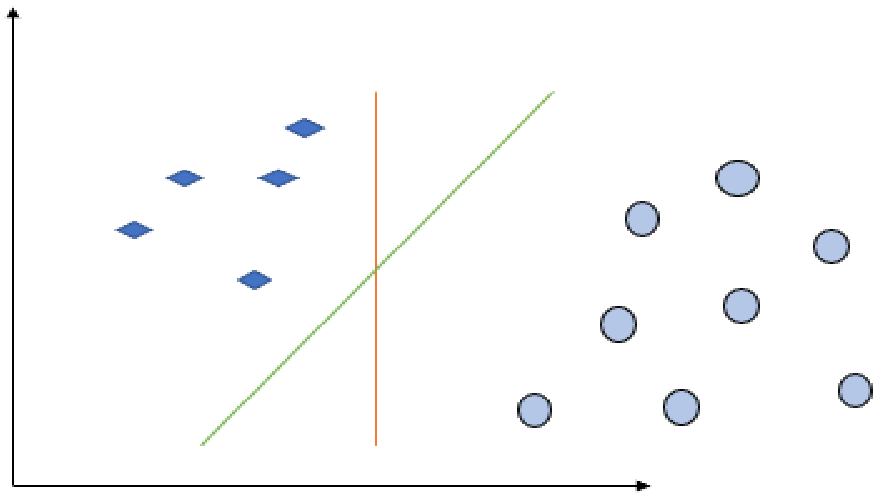


Figure 15: Support Vector Machine Algorithm

Logistic regression

LR (logistic regression) is a fixed of supervised take a look at class guidelines that are expecting the opportunity of a goal variable [33]. The nature of the goal or established variable is dichotomous, this means that there are most effective feasible classes [23]. In different words, the established variable is binary in nature, with facts coded as both 1 (for success/yes) or 0 (for failure/no). A logistic regression version predicts P(Y=1) as a feature of The LR method is quite popular, and because it is based on a basic algorithm, it is employed in a variety of medical prediction and detection applications [12]. Throughout overall, logistic regression refers to binary logistic regression with binary goal variables, however it is able to additionally are expecting extra classes of goal variables.

There are several kinds of Binomial, multinomial and possibly ordinal Before we dive into logistic regression implementation, we ought to be privy to the subsequent assumptions like:

- There need to be no multi-collinearity with inside the version, this means that that the unbiased variables need to be unbiased of 1 another.
- We need to use a massive pattern length for logistic regression.

- The goal variables in binary logistic regression ought to continually be binary, and the preferred final results is represented with the aid of using issue stage 1.

Logistic regression evaluation is beneficial for predicting the chance of an occasion. It aids in figuring out the possibilities of any classes. conclusion, logistic regression can be expecting whether or not or now no longer an occasion will arise primarily based totally on ancient facts [1]. In summary, logistic regression can predict whether, a tumor is fatal, group of users will buy a product, An email is considered spam, Today the weather will be cloudy, the recipient of the promotional email may be a responder or non-responder, an online transaction is fraudulent, an election will be won by one of the candidates, policyholder will die before the term of the policy expires, a person is going to buy a car [40].

In our thesis paper, LR achieves an accuracy of 83% in a balanced dataset while it achieves an accuracy of 71% in an unbalanced dataset. Below is provided the LR picture diagram:

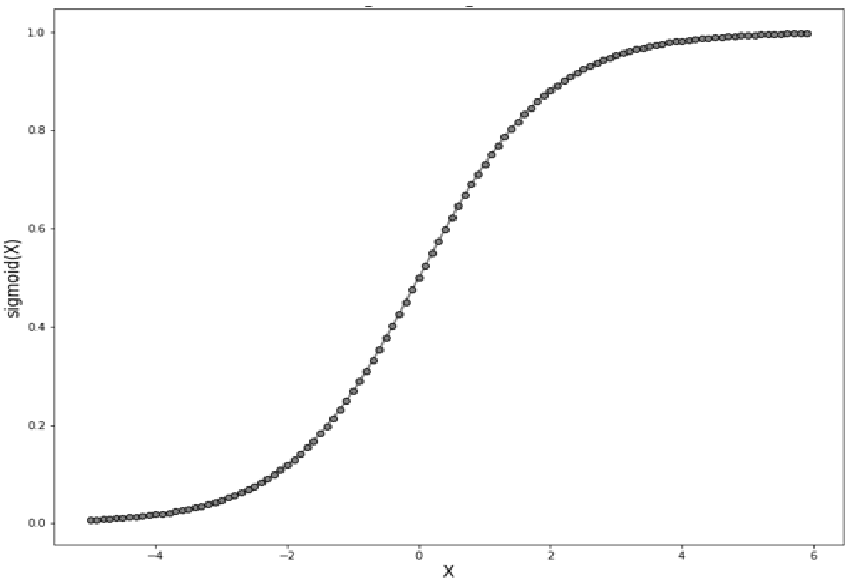


Figure 16: Logistic Regression Algorithm

Convolutional Neural Network

CNN means Convolutional Neural Networks. Convolutional neural systems (CNNs) are neural systems with one or more convolutional layers that are utilized basically for picture planning, classification, division, and other auto-correlated information. Convolution is essentially the sliding strategy of a channel [16].Convolutional Neural Systems (CNN) are a kind of deep innovation in learning that has come to notoriety in computer vision and is examined in an assortment of divisions, counting radiology [23].A convolutional neural arrange is made up of numerous layers, such as convolution layers, pooling layers, and fully associated levels, and learns the spatial progressions of highlights accordingly and adaptively using a backpropagation method [17].Understanding convolutional neural networks' concepts, benefits, and confinements is pivotal for maximizing their potential to progress radiologist execution and, eventually, understanding care. In most modern radiomics examinations, hand-crafted highlight extraction approaches like surface investigation are utilized to begin with, taken after by conventional machine learning classifiers like arbitrary timberlands and back vector machines.

There are many qualifications to be made between such strategies and CNN. To start with, CNN does not necessitate feature extraction by manually [14]. Second, CNN plans don't continuously require human specialists to fragment tumours or organs. Third, since there are millions of learnable parameters to gauge, CNN is altogether more information hungry and computationally costly, requiring the utilize of design handling units (GPUs) for demonstrate preparing. CNNs provide detailed information despite their monstrous control and resource complexity [11]. It's almost all about identifying designs and highlights that are so small and unrelated to the human eye that they go unnoticed [20]. However, when it comes to comprehending an image's substance, it falls short. Convolution is the method of combining two capacities in arrange to deliver the yield of the other. A Include outline is made by convolving the input picture with the utilize of channels in CNNs. Weights and inclinations are randomly delivered vectors within the organize, and they are utilized as filters. CNN utilizes the same weights and inclinations for all neurons instead of having particular weights and inclinations for each [23].

In our thesis paper, CNN achieves an accuracy of 83% in a balanced dataset while it achieves an accuracy of 75% in an unbalanced dataset. Below is provided the CNN picture diagram:

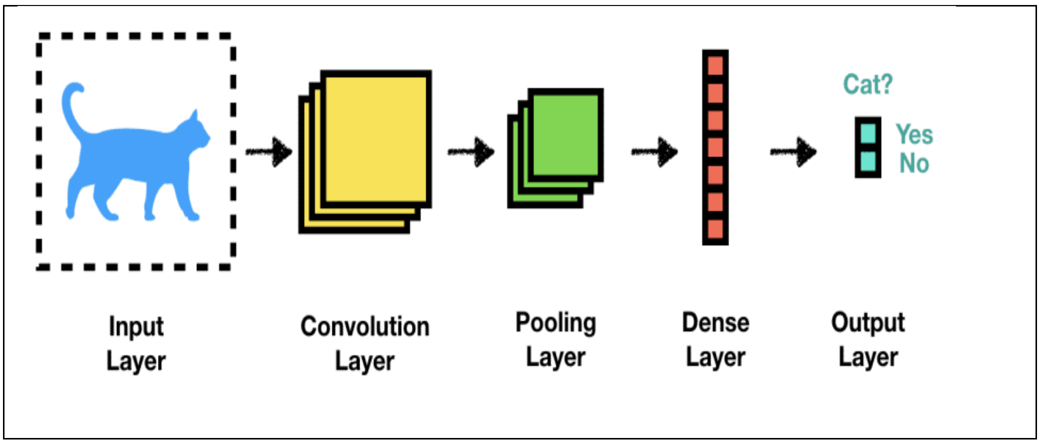


Figure 17: Convolutional Neural Networks Algorithm

Light Gradient Boosting Algorithm

Lgbm means Light Gradient Boosting Machine. Lgbm is critical in the health-care system and is widely used in other industries [21]. Lgbm is a relatively new algorithm that is quite beneficial due to its high processing speed [10]. Lgbm is a gradient augmentation framework that can be used for tree-based learning algorithms to achieve high computational efficiency [15]. The Lgbm method develops sheet by sheet, while other algorithms develop level by level. Lgbm decides to grow the leaf with the greatest loss. Large scale projects are carried out using Lgbm [16].When it comes to obtaining rapid and high-accuracy results, LightGBM is considered to be a really fast algorithm and the most widely used algorithm in machine learning. In the LightGBM manual, there are over 150 parameters to choose from [30]. Lightgbm's speed is due to histogram-based division and gradient-based one-side sampling division (GOSS), as well as exclusive feature bundling (EFB) [40].

In our thesis paper, Lgbm achieves an accuracy of 86% in a balanced dataset while it achieves an accuracy of 71% in an unbalanced dataset. Below is provided the Lgbm picture diagram:

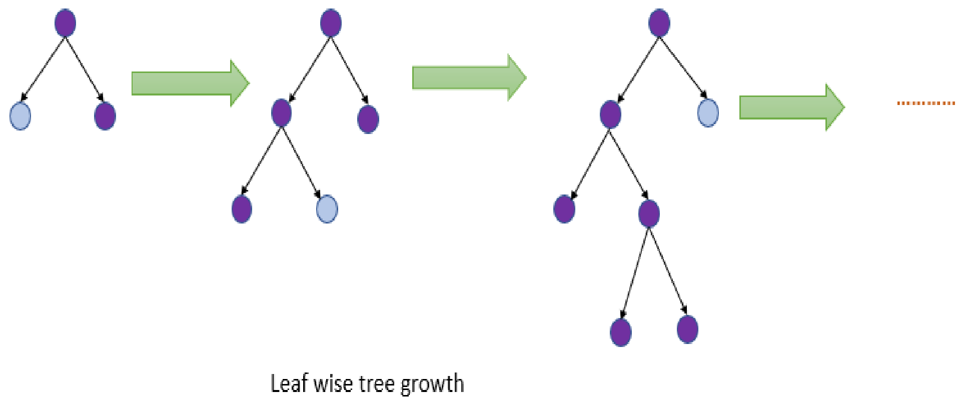


Figure 18: Light Gradient Boosting Machine Algorithm

XGboost

XGboost means Extreme Gradient Boosting. Gradient boosting with portability is known as XGboost [22]. To comprehend Boost, first need understands the ml ideas and methods on which it is based: SML, ensemble learning, decision trees and gradient boosting [2]. The points of interest are various, the foremost imperative of which are:

- Its employments parallel preparing control.
- advances regularization.
- It is faster than gradient enhancement.
- After each emphasis, the client can perform cross-validation.

- He is extremely adaptable.
- It is built such an amazing highlight that can handle lost data

Since of its capacity to deliver profoundly precise comes about, XGBoost is one of the foremost prevalent ML algorithms. On scanty and unstructured information, XGBoost does not continuously perform well [28]. Gradient Boosting is very sensitive to exceptions because each proofreader is required to correct the mistakes of predecessor students [20]. Overall, the strategy isn't scalable [30]. Indeed, the estimators base their precision on past indicators, which makes the rationalization of the method problematic. Following the way that a choice tree takes to form its choice, for case, is minor and self-explanatory, but following the paths of hundreds or thousands of trees is much more difficult." born-again" choice tree that approximates the same choice work to realize both execution and interpretability [15].

In our thesis paper, XGboost achieves an accuracy of 86% in a balanced dataset while it achieves an accuracy of 71% in an unbalanced dataset. Below is provided the XGboost picture diagram:

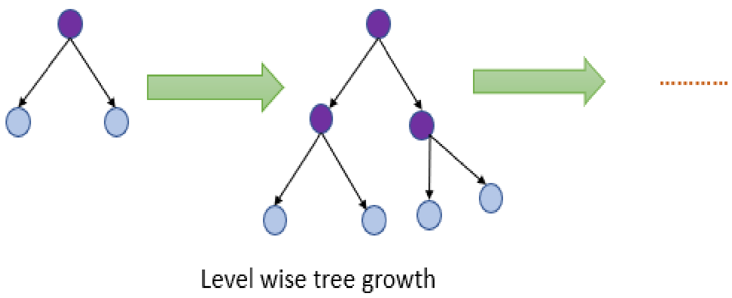


Figure 19: Extreme Gradient Boosting Algorithm

3.1.16 Tools and Libraries Used

All kind of implementation was done by using different kind of tools and libraries. We have used several libraries for different types of cases. These libraries and tools support the study of convolutional neural networks and machine learning [19]. Such technologies additionally involve the development of new procedures and the advancement of effective implementation. Some essential features are Scikiti-learn, TensorFlow, Imblearn, Matplotlib, and OpenCV. We used all parts for our execution.

Scikitlearn

Scikitlearn is versatile and powerful Python gadget mastery package in machine learning [20]. It makes use of a Python consistency interface to offer a large number of green machines learning competencies and statistical modeling, including classification, regression, clustering, and dimensionality reduction. NumPy, SciPy and Matplotlib are the idea of this package, written specially in Python [21]. It is easy to use and, above all, effective. Install the library with the scikitlearn id (eg.pip introduce scikit-learn), but consequence it with the sklearn identifier in your source code (i.e., purport sklearn) [4]. At JPMorgan, Scikitlearn is a critical aspect of the Python gadget mastering toolset. Applied for different purposes such as the bank employs classification, predictive analysis, and different gadget mastering techniques [22]. It gives usefulness for both administered and unsupervised learning. Scikitlearn started out as scikit. learn, a Google Summer of Code extend started by French information researcher David Cournapeau. Other designers modified the first codebase later. Fabian Pedregosa, Gael Viroqua, Alexandre Gramfort, and Vincent Michel, all from the French Organized for Inquire about in Computer Science and Mechanization in Rocquencourt, France, took over as extend pioneers in 2010 and discharged the primary open adaptation on February 1st, 2010. One of the best-known machine learning libraries is Scikitlearn.It is one of the foremost prevalent machine learning libraries on GitHub.

TensorFlow

TF (TensorFlow) is a Google experiment that is deepened with the open supply package. It additionally works with common gadget getting to know methods. It was designed for large-scale statistical processing rather than gimmicky knowledge [23]. However, it proved treasured for gadget getting to know, so Google made it open-supply. When coping with significant quantities of statistics, non-linear and non-arrays are helpful. TensorFlow is primarily based on graphs describing statistical flows with nodes and edges [20]. TensorFlow code is extensively simpler to execute in an allotted manner throughout a cluster of Since the method of execution is in the form of graphics, computer systems using GPUs can run faster. The coding system for gadget getting to know and deep getting to know turned into appreciably greater complicated earlier than the invention of libraries [21]. This library provides a high-level API that eliminates the need for elegant coding to create a neural network, install a neuron, or apply a neuron. All of those obligations are finished through the library [15]. TensorFlow additionally gives Java and R integration. Dataflow graphs describe how statistics actions thru a chart, and TensorFlow lets in you to layout them. The nodes withinside the diagram constitutes a systematic system. Multiple database arrays are a connection or facet among elements. TensorFlow structure is made of 3 great steps:

- Data pre-processing - prepare the statistics and restriction it to an unmarried value.
- The statistics version is created by the model advent.
- Model estimation and training - use the statistics to educate and take a look at the version with unknown statistics.

Imblearn

Imbalanced-learn (additionally called learning) is an MIT-certified open-supply bundle that uses scikiti-learn [19]. The degree of class imbalance varies, but severe imbalances are more difficult to observe and may require specialized methods and techniques. Imblearn techniques are used for various strategies in order to gather an equal number of classes [36].This type of data collection would allow the predictive model to generalize well. The basic approach for trading with an unbalanced data set is up sampling and down sampling [37]. Data can be created by creating simulated data for the minority class to match the ratio to the majority class [23]. In contrast, Down sampling is the process of reducing the majority class data points to compare them to the minority class.

Matplotlib

Matplotlib is a widely used python package for data visualization. It is an open-source alternative to MATLAB. Pyplot is the API of matplotlib which enables it to use as an alternative. For basic to advanced, static or dynamic data visualization matplotlib is the optimal choice of library. It is an extremely convenient tool for making plots, graph, and charts as it requires very few lines of code to design these high level graphical representations of data.

OpenCV

Open-Source Computer Vision Library (OpenCV) is a free, open-deliver software program application library for pc vision and machine learning. We used a variety of languages, including

Java, Python, and C++, all of which are widely used [18]. OpenCV is written in C++ and has a first-rate interface this is additionally written in C++. of the contemporary breakthroughs and algorithms were displayed through manner of way of the c++ interface [20]. Agreements are available for Python, Java, and MATLAB/OCTAVE. \ OpenCV end up released in version 3.4 and protected JavaScript bindings for a subset of OpenCV abilities [21]. Windows, Linux, macOS, FreeBSD, NetBSD, and OpenBSD are all supported through manner of way of OpenCV on pc computers. Android, iOS, Maemo, and Windows Phone are all supported through manner of way of OpenCV. BlackBerry 10 (version 17) The customer can get respectable SourceForge releases or the most contemporary GitHub sources. Make is the compiler used by OpenCV. There are different Features of OpenCV library:

- Please make a video recording and save it.
- Images ought to be take a look at and written.
- 3.Do a feature search.
- Image processing (filtering and transformation)
- Identify specific items in films or photos, consisting of people, eyes, and cars.
- 6.Evaluate the clip by predicting its movement, removing the background, and monitoring the objects within it.

PC cognitive problems were solved by OpenCV Intel created OpenCV in 1999, despite the fact that Willow Garage later sponsored it. OpenCV allows C++, Python, Java, and specific programming languages [22].

Seaborn

Seaborn is a roaring Python library for statistical graphs. It aids in the analysis and exploration of data. Seaborn is a library that plots graphs with Matplotlib. It will be used to visualize random distributions. Anyone can easily represent their data on a plot by using this library. This library is used to visualize data; no need to worry about internal details; simply pass the data set or data inside the replot () function, and it will calculate and place the value appropriately. It aids in the visualization of information from matrices and data frames, as well as the display of distributions in an easy and flexible manner. For Creating visually appealing statistical plots and custom color palettes, this library is useful. It enables the creation of enhanced data visuals. This aids our understanding of the data by presenting it in a visual context, allowing us to uncover any hidden correlations between variables or trends that may not be obvious at first. Seaborn has a higher-level interface than Matplotlib. Seaborn enhances the visual appeal of our charts and plots while also meeting some of the most common data visualization requirements (like mapping color to a variable or using faceting). Essentially, it facilitates data visualization and exploration. There are a few limitations in matplotlib that Seaborn addresses, such as the fact that matplotlib functions do not work well with data frames, whereas Seaborn does, and it also includes a large number of high-level interfaces and customized themes that matplotlib lacks because it is difficult to figure out the settings that make plots appealing.

Chapter 4

Results or findings

4.1.1 Evaluation of Algorithm performance

Evaluation metrics

Accuracy

Accuracy is one of most popular and simplest method for evaluating a machine learning model's performance. Although, it is not the most effective evaluation metrics, the simplicity for it makes it one of the staple choices. Accuracy is simply the percentage of correct prediction made by the model. It compares the predicted labels to the actual labels to calculate the percentage of correct predictions. Accuracy can be defined as:

$$Accuracy = \frac{Number\ of\ correct\ predictions}{Total\ number\ of\ predictions} \qquad 4.1$$

For a simple use case scenario accuracy can be the only metrics used for comparison among various models. However, in complex scenarios accuracy of a model cannot be relied on completely for evaluation of the model. In case of imbalanced data, accuracy cannot show the actual efficacy of the model because of not considering the imbalance while calculating the correct predictions. The accuracy score will be higher even without the model being able to predict most of the samples from the minority class. In our research, the dataset has a significant amount of imbalance. So, accuracy score was not able to provide a proper overview of the model's performance. But, for the balanced dataset achieved by using SMOTE accuracy score was providing a good estimation of the performance. We compared accuracy scores for both balanced and imbalanced dataset for a comparative analysis.

Sensitivity

Sensitivity is the measure of the ability of a machine learning model to correctly classify the positive instances. It is also known as the True positive rate or Recall. For instance, if healthy fetus is considered positive class and unhealthy fetus is considered as the negative class, a healthy fetus being predicted healthy by the model will be considered as true positive. So, true positive is correctly classifying an instance as positive. Similarly, if an instance is positive but it is predicted to be negative, it is considered as false negative. False negative is classifying an instance as negative when the actual class is positive.

True Positive: Being predicted positive when the actual class is positive or truly identified as positive.

False Negative: Being predicted negative when the actual class is positive or falsely identified as negative.
Sensitivity is the measure of true positives. It can be defined as:

$$Sensitivity = \frac{True\ positives}{True\ positives + False\ negatives} \qquad 4.2$$

If a model has high sensitivity that means, the model is able to truly classify the positive labels more than negatively classifying them. A low sensitivity means, the model is falsely classifying positive classes as negative more than truly classifying them. This provided a way to evaluate the model's ability to distinguish among the classes.

Specificity

Specificity is the measure of the ability of a machine learning model to correctly classify the negative instances. Specificity measures the proportion of true negatives. Specificity is also known as the True negative rate. For example: If healthy fetus is considered positive class and unhealthy fetus is considered the negative class and if a fetus is unhealthy and the model predicts it correctly is considered a true negative. This means, correctly classifying the negative class. Similar to this, if a model predicts positive and the actual class is negative, it is considered false positive.

True Negative: Being predicted negative when the actual class is negative or truly identifying as negative.

False Positive: Being predicted positive when the actual class is negative or falsely identifying as positive.

$$Specificity = \frac{True\ negatives}{True\ negatives + False\ positives} \qquad 4.3$$

A higher specificity tells that the model is predicting the negative class correctly more than predicting the class incorrectly. A low sensitivity tells that the model is predicting the negative classes incorrectly more than predicting them correctly. Similar to sensitivity, specificity also provides the evaluation of the model based on its ability to correctly differentiate among the classes. For our dataset, specificity is really low for the imbalanced dataset due to the lower amount of data present for unhealthy fetus.

AUC

AUC or Area under the ROC curve is an evaluation metrics for machine learning model which measures the model's ability to distinguish between various classes. AUC score ranges from 0 to 1 where 0 represents 100% incorrect predictions and 1 represents 100% correct predictions. So, the closer the AUC score is to 1 the better.

ROC

ROC or Receiver operator characteristic is a graph used for evaluating the performance of a machine learning model. It evaluates the model's performance at all possible classification thresholds. It is based on the true positive rate and the false positive rate.

True positive rate: TPR also known as sensitivity is the rate of correctly classified positive instances.

$$TPR = \frac{True\ positive}{True\ positive + False\ negative} \tag{4.4}$$

False positive rate: FPR is known as the rate of incorrectly classified negative instances as positive.

$$FPR = \frac{False\ positive}{False\ positive + true\ negative} \tag{4.5}$$

An ROC curve is a plot of TPR vs FPR at all possible classification threshold. The threshold has to be chosen to maintain a balance between the classifications for different classes. Increasing or decreasing the threshold impacts the TPR and FPR significantly. We have generated ROC for each of the 9 algorithms for both the balanced and imbalanced dataset.

Confusion Matrix

Confusion matrix is a table providing overview of the performance of a machine learning model. Unlike accuracy, confusion matrix gives result based on the performance of the model considering all the classes. It is simple and easy to understand table displaying the number of correctly and incorrectly classified instances from the test data. For a binary class the confusion matrix is a 4×4 table.

		Predicted class	
		Class 0	Class 1
Actual class	Class 0	True positives	False negatives
	Class 1	False positives	True negatives

Figure 20: Confusion matrix

Precision

Precision is the ratio of correctly predicted instances and the total number of instances. Precision value tells us the rate of correctly predicted instances from a specific class within the total number of instances belonging to that class. High precision shows the model's ability to correctly classify the instances of a class while keeping the false predictions at a minimum. For our dataset, precision for the healthy class is the ratio of correctly predicted healthy fetus among the total healthy predictions that has been. It has been calculated for both the classes separately. Precision can be defined as:

$$Precision\ (positive\ class) = \frac{TP}{TP + FP} \tag{4.6}$$

$$Precision\ (Negative\ class) = \frac{TN}{TN + FN} \tag{4.7}$$

$$Precision = \frac{Correctly\ predicted\ instances\ of\ a\ class}{Total\ predicted\ instances\ for\ that\ class} \tag{4.8}$$

Recall

Recall is another name for sensitivity. For our dataset, recall for the healthy class will be correctly predicted healthy fetus among actual total number of healthy fetus. Recall for the positive class can be defined as:

$$Recall = \frac{True\ positives}{Total\ positives} \tag{4.9}$$

F1 score

F1 score is a measurement based on both precision and recall. It uses the harmonic mean of the two metrics to calculate a combined score. As it uses both precision and recall for calculation, it provides a good measure for achieving a balance between the two metrics. F1 score provides a better estimation when there is imbalance in the dataset. F1 score can be defined as:

$$F1\ score = \frac{2(recall \times precision)}{recall + precision} \tag{4.10}$$

Result for imbalanced vs. balanced dataset

Based on accuracy

By observation of the accuracy for the 9 different algorithms for both imbalanced and balanced data we see that, accuracy has dropped to some extent for the balanced dataset. The classifiers build using imbalanced dataset were able to classify the majority classes much better than the balanced dataset. Among all the algorithms Random Forest has the highest accuracy of 89% for imbalanced dataset and Decision tree has the highest accuracy of 83% for the balanced data. The lowest accuracy for the balance data is for decision tree with an accuracy of 68% where for balanced data Linear regression has the lowest accuracy of 63%

Algorithms	**Accuracy**	**AUC**	**Sensitivity**	**Specificity**
KNN	86	0.83	1.00	0.28
Linear regression	0.85	78	0.76	0.85
Logistic regression	0.57	83	0.90	0.57
Decision tree	0.28	67	0.76	0.28
Random Forest	0.42	89	1.00	0.42
SVM	0.85	81	0.80	0.86
Lgbm	1.00	86	0.83	1.00
XGboost	0.71	81	0.86	0.71
CNN	0.42	83	0.93	0.57

Table 7: Accuracy, AUC, sensitivity and specificity for imbalanced data

Algorithms	**Accuracy**	**AUC**	**Sensitivity**	**Specificity**
KNN	80	0.86	0.78	0.71
Linear regression	70	0.80	0.25	0.96
Logistic regression	71	0.79	0.64	0.84
Decision tree	81	0.83	0.75	0.87
Random Forest	75	0.84	0.75	0.78
SVM	66	0.76	0.35	0.93
Lgbm	71	0.86	0.32	1.00
XGboost	76	0.81	0.50	0.84
CNN	75	0.82	0.67	0.87

Table 8: Accuracy, AUC, sensitivity and specificity for balanced data

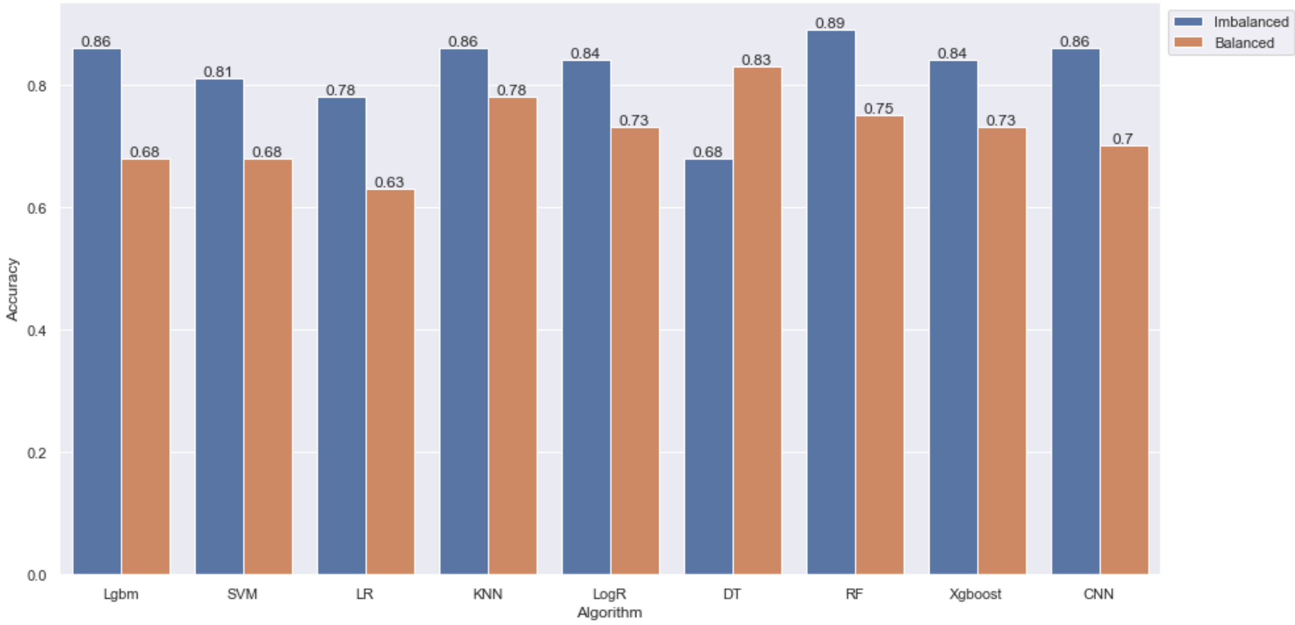


Figure 21: Accuracy of various algorithms

Based on Sensitivity and specificity

From the graph-bar-2 we can observe the difference between the sensitivity value for each of the 9 algorithms using both balanced and imbalanced data. The sensitivity value is highest for K nearest neighbors for both the balanced and imbalanced data. For the imbalanced data the value is 1.00 and for the balanced data the value is 0.86. Similarly, from the graph-bar-3 we can observe the specificity values. Specificity value is highest for Random forest using imbalanced data with a value of 1.00 while Lgbm has the highest specificity score of 1.00 for balanced data. Using both imbalanced and balanced data KNN has the lowest specificity of 0.28 using the imbalanced data and 0.73 using the balanced data. If both of these metrics are considered Lgbm has the most balanced score for imbalanced data with a sensitivity score of 0.92 and specificity score of 0.83. Similarly, for the balanced dataset decision tree performed the best with a sensitivity score of 0.83 and specificity score of 0.80

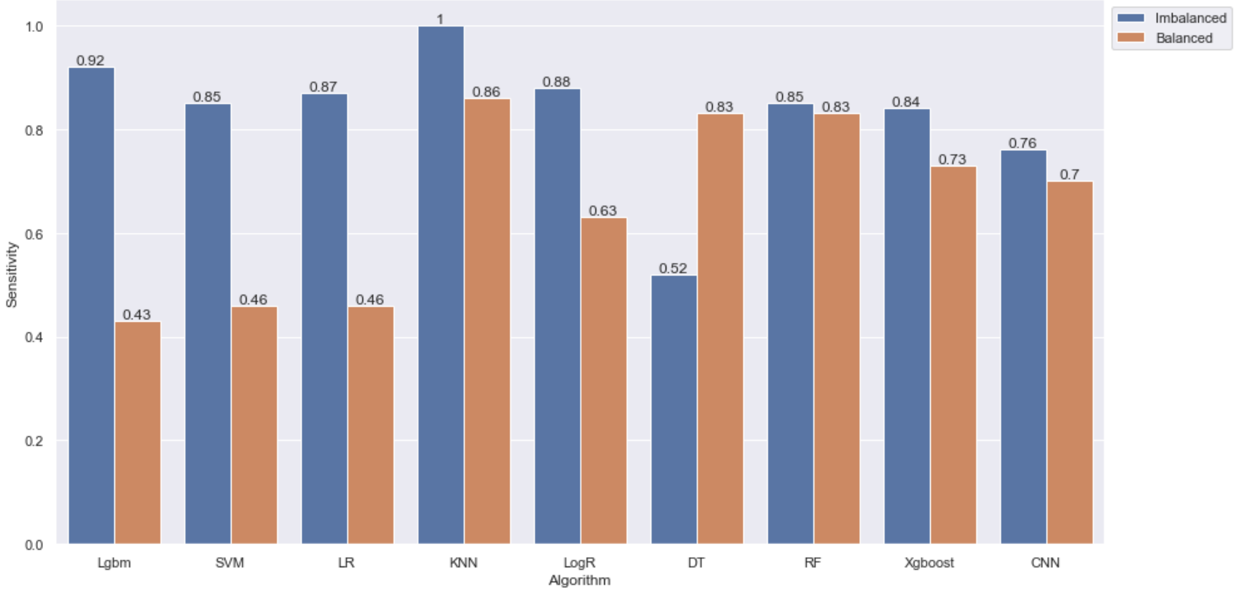
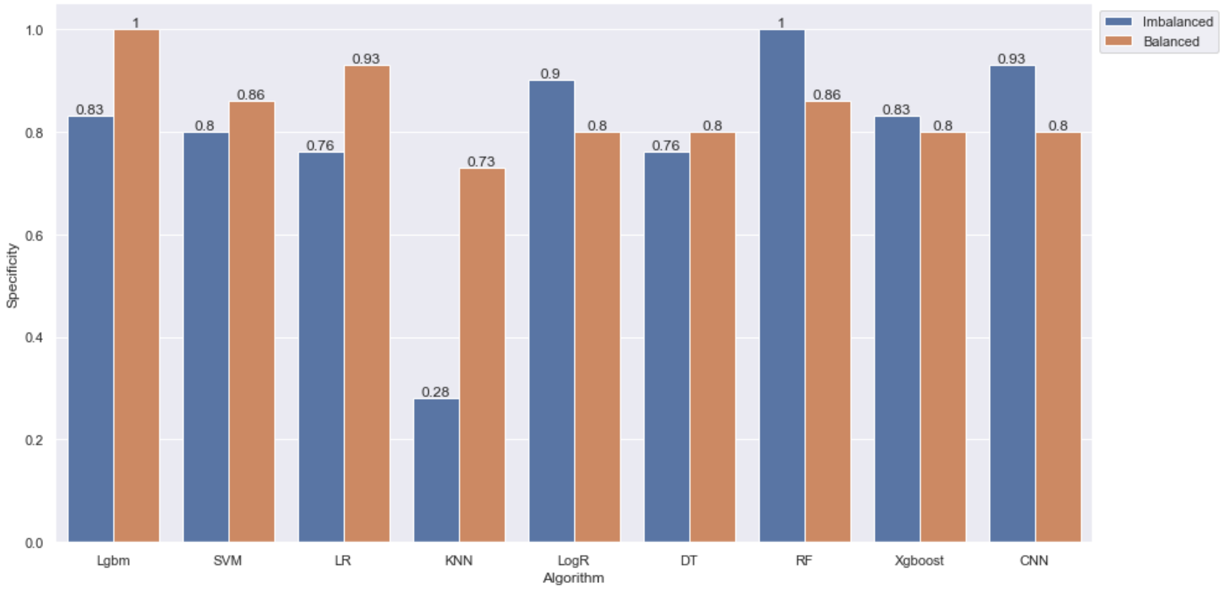


Figure 22: Sensitivity score for balanced and imbalanced data



Graph- Specificity score for balanced and imbalanced data

Based on ROC and AUC

Based on the AUC score derived from the ROC curve we can see that, Lgbm has the highest AUC score for the imbalanced dataset as well as for the balanced dataset with an AUC score of 0.92 and 0.87 respectively. Considering the AUC score for both imbalanced and balanced data, the AUC scores are reduced significantly due to the application of SMOTE. For each of the algorithms except decision tree the AUC score has decreased. The only significant improvement can be seen for decision tree. Decision tree has an AUC score of 0.52 which has drastically improved for the balanced dataset to a score of 0.82. SMOTE has positively impacted only the results for decision tree, the other algorithms has been impacted negatively. If we look at the ROC curves in Fig- for imbalanced data and in Fig- for balanced data we can observe the same phenomenon of the curves having more coverage for the imbalanced dataset than the balanced dataset.

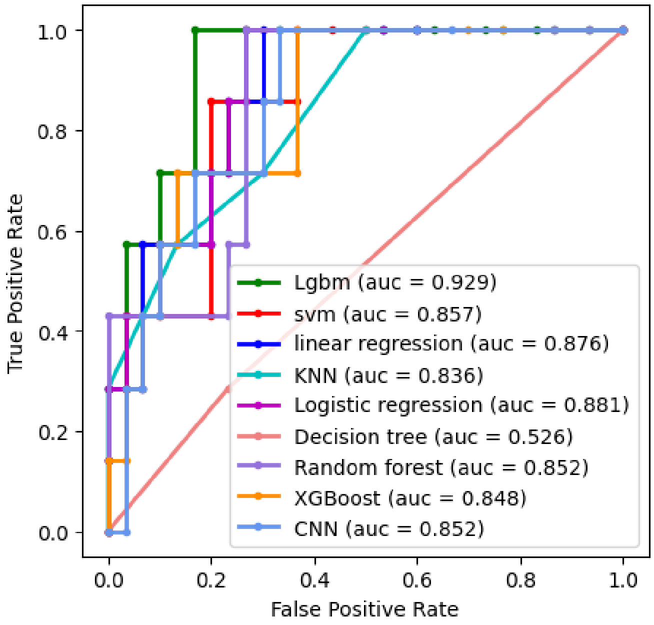


Figure 23: ROC curve for imbalanced data

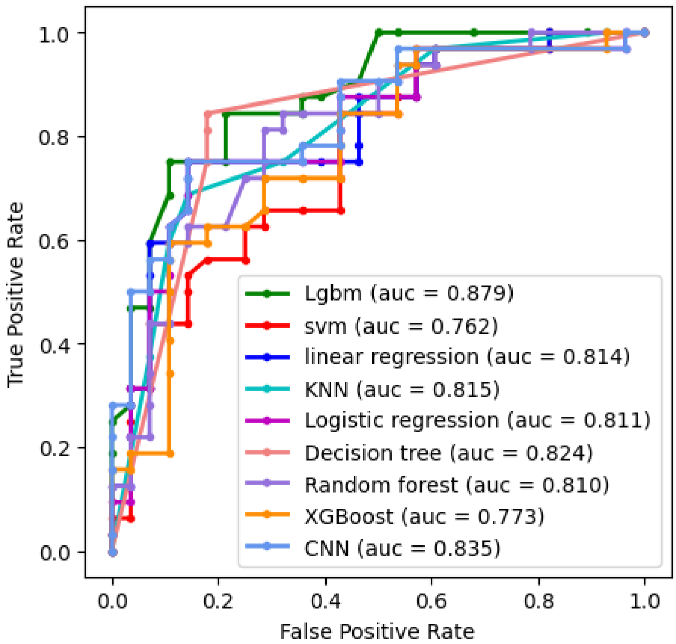


Figure 24: ROC curve for balanced data

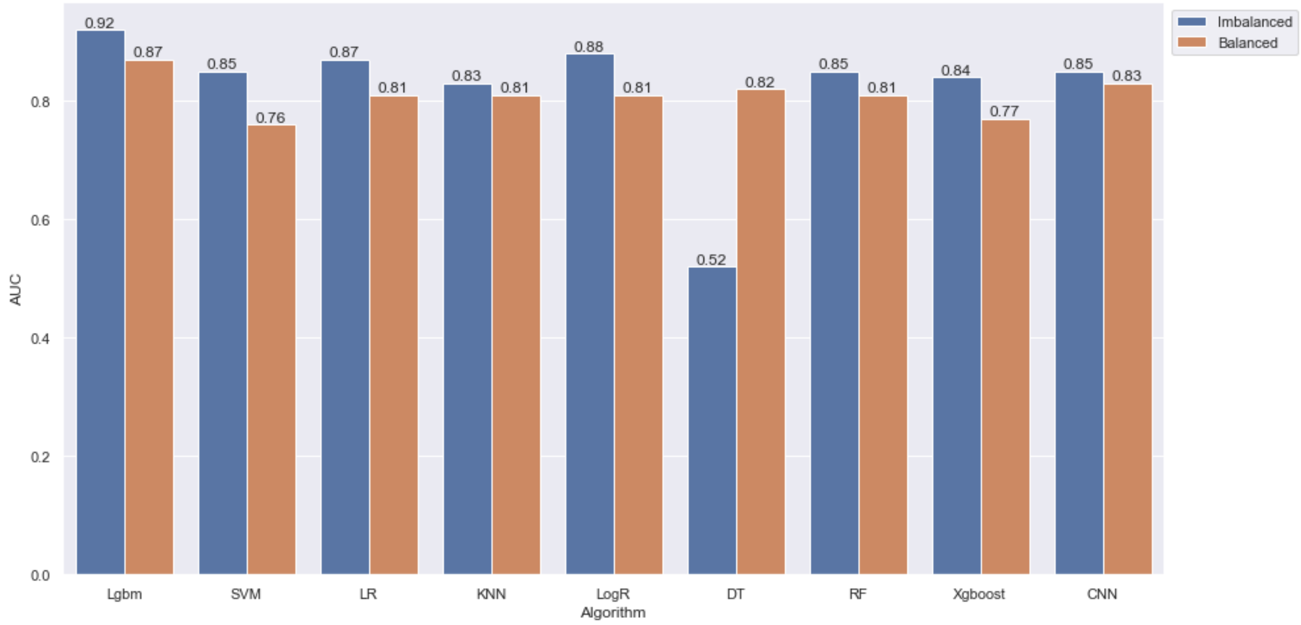


Figure 25: Comparison of AUC score among various algorithms

Based on precision recall and F1 score

As the aim of this study is to be able to predict any risk of fetus being born with birth defects before the birth of the fetus, the focus is mostly on increasing the number of correctly predicted instances belonging to the unhealthy class and decreasing incorrectly predicted instances of the unhealthy class. For our research, healthy class has been considered the positive class and unhealthy class has been considered the negative class. The goal is to increase the number of “True negatives” and decrease the number of “False positives”

Precision for the healthy class or positive class increases with an increase in TP and decrease in FP. Similarly, precision for the negative class increases with an increase in TN and a decrease in FN. As we are aiming to increase TN and decrease FP the precision score for both the class needs to be maximized.

For recall score for the positive class, it increases with an increase in TP and for negative class it increases with an increase in TN. For achieving better result of TN the recall for the unhealthy or negative class needs to be maximized.

F1 score considers both precision and recall. Maximizing the F1 score gives an overall balanced representation of both precision and recall.

By analyzing the precision, recall, F1 score for both the classes separately from Table-2-1 for the imbalanced dataset, Light Gradient Boosting (Lgbm) has the most balanced precision, recall and F1 score. For the balanced dataset analyzing Table-2-2 shows that, decision tree has the most balanced scores for all three metrics.

	Precision		Recall		F1-score	
	0	1	0	1	0	1
KNN	0.86	1.00	1.00	0.29	0.92	0.44
Linear regression	0.96	0.46	0.77	0.86	0.85	0.60
Logistic regression	0.90	0.57	0.90	0.57	0.90	0.57
Decision tree	0.82	0.22	0.77	0.29	0.79	0.25
Random Forest	0.88	1.00	1.00	0.43	0.94	0.63
SVM	0.96	0.50	0.80	0.86	0.87	0.63
Lgbm	1.00	0.58	0.83	1.00	0.91	0.74
XGboost	0.93	0.50	0.87	0.71	0.88	0.59
CNN	0.87	0.50	0.93	0.57	0.89	0.46

Table 9: Precision, recall, F1 score for imbalanced data

	Precision		Recall		F1-score	
	H	U	H	U	H	U
KNN	0.71	0.79	0.79	0.72	0.75	0.75
Linear regression	0.88	0.64	0.25	0.97	0.61	0.76
Logistic regression	0.76	0.69	0.64	0.84	0.69	0.74
Decision tree	0.81	0.83	0.75	0.88	0.82	0.81
Random Forest	0.72	0.80	0.75	0.78	0.77	0.73
SVM	0.78	0.62	0.36	0.94	0.58	0.72
Lgbm	1.00	0.64	0.32	1.00	0.60	0.78
XGboost	0.79	0.75	0.50	0.84	0.76	0.77
CNN	0.78	0.73	`0.68	0.88	0.74	0.76

Table 10: Precision, recall, F1 score for balanced data

Result based on the algorithm

K-Nearest Neighbors

For K-nearest neighbors the accuracy is 86% for the imbalanced dataset and 78% for the balanced dataset. Being trained on the imbalanced dataset knn performed better at predicting the healthy fetuses. KNN has sensitivity score of 1 for the imbalanced dataset 0.86 for the balanced dataset and specificity score of 0.28 for imbalanced dataset and 0.73 for the balanced dataset. The AUC score is 0.83 for imbalanced dataset and 0.81 for balanced dataset. Overall, introducing synthetic data has not been effective in improving the performance for knn algorithm. Rather, a significant reduction in performance is being noticed throughout each of the evaluation metrics. Below is the confusion matrix for both balanced and imbalanced dataset using knn:

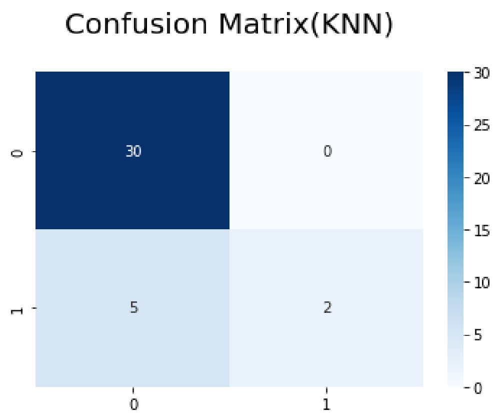


Figure 26: Confusion matrix knn for imbalanced dataset

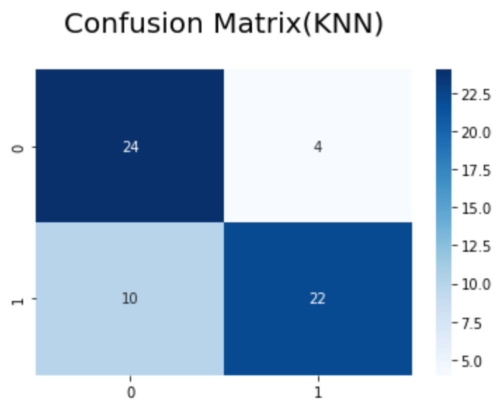


Figure 27: Confusion matrix knn for balanced dataset

Linear regression

Linear regression has an accuracy of 84% for balanced dataset and 73% for the imbalanced dataset. The sensitivity and specificity scores are 0.87 for balanced dataset 0.46 for imbalanced dataset and 0.76 for balanced dataset, 0.93 for imbalanced dataset respectively. The AUC score using linear regression is 0.87 for imbalanced dataset and 0.81 for the balanced dataset. Similar to knn, the balancing has also decreased the ability of the classifier to detect healthy fetuses which has brought down the overall score for each of the metrics. Below is the confusing metrics generated for linear regression using bother imbalanced and balanced datasets:

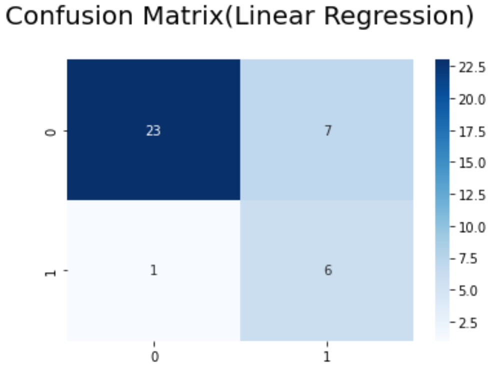


Figure 28: Confusion matrix Linear Regression for imbalanced dataset

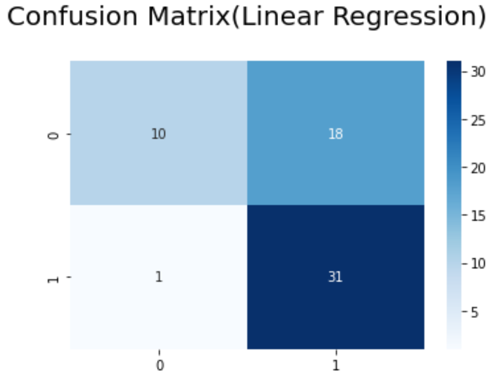


Figure 29: Confusion matrix Linear Regression for imbalanced dataset

Logistic regression

Logistic regression has an accuracy of 78% for balanced dataset and 63% for the imbalanced dataset. The sensitivity and specificity scores are 0.88 for balanced dataset 0.63 for imbalanced dataset and 0.90 for balanced dataset, 0.80 for imbalanced dataset respectively. The AUC score using logistic regression is 0.84 for imbalanced dataset and 0.81 for the balanced dataset. Similar to the previously mentioned algorithms, using the balanced dataset decreases the overall performance of the classifier in its ability to detect instances of the healthy fetus class.

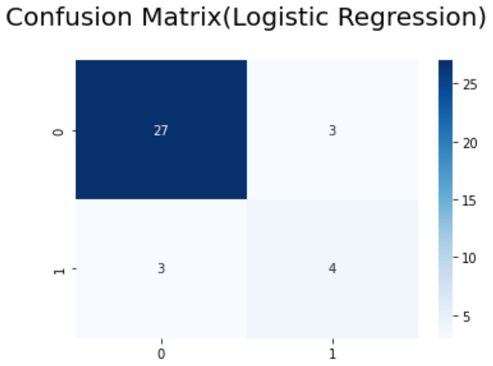


Figure 30: Confusion matrix Logistics regression for imbalanced dataset

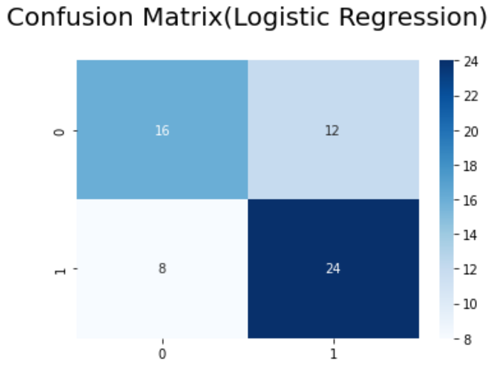


Figure 31: Confusion matrix Logistic regression for balanced dataset

Decision tree

Decision tree has an accuracy of 68% for balanced dataset and 83% for the imbalanced dataset. The sensitivity and specificity scores are 0.52 for balanced dataset 0.83 for imbalanced dataset and 0.76

for balanced dataset, 0.80 for imbalanced dataset respectively. The AUC score using decision tree is 0.52 for imbalanced dataset and 0.82 for the balanced dataset. Unlike previously mentioned algorithms, decision tree has significant improvement in its performance. Both healthy and unhealthy fetuses are getting detected better through the use of balanced dataset for decision tree. Using balanced data has improved the accuracy, sensitivity, specificity and AUC score. If we look at the confusion matrix, we can observe the same improvements. For the balanced dataset, 5 instances from each classes has been misclassified but for the imbalanced dataset only 2 instances out of 7 instances of unhealthy fetus was correctly classified by the classifier. Below is the confusing metrics generated for decision tree using bother imbalanced and balanced datasets:

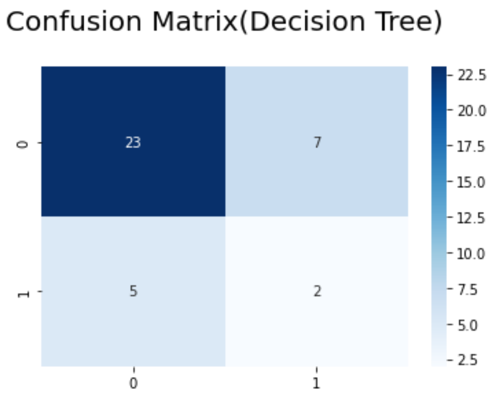


Figure 32: Confusion matrix decision tree for imbalanced dataset

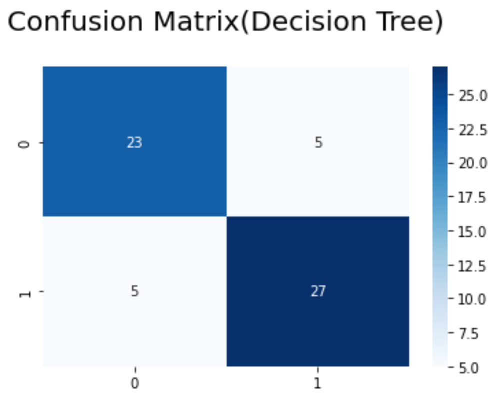


Figure 33: Confusion matrix decision tree for balanced dataset

Random Forest

Random forest has an accuracy of 89% for balanced dataset and 75% for the imbalanced dataset. The sensitivity and specificity scores are 0.85 for balanced dataset 0.83 for imbalanced dataset and 1.00 for balanced dataset, 0.86 for imbalanced dataset respectively. The AUC score using random forest is 0.85 for imbalanced dataset and 0.81 for the balanced dataset. Similar to the previously mentioned algorithms, using the balanced dataset decreases the overall performance of the classifier in its ability to detect instances of the healthy fetus class. But, the ability to detect unhealthy fetuses has increased after using balanced dataset. Below is the confusing metrics generated for random forest using bother imbalanced and balanced datasets:

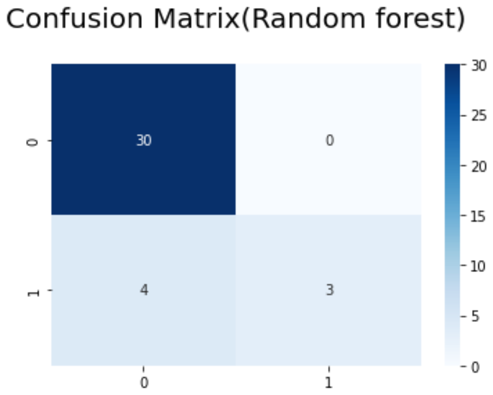


Figure 34: Confusion matrix random forest for imbalanced dataset

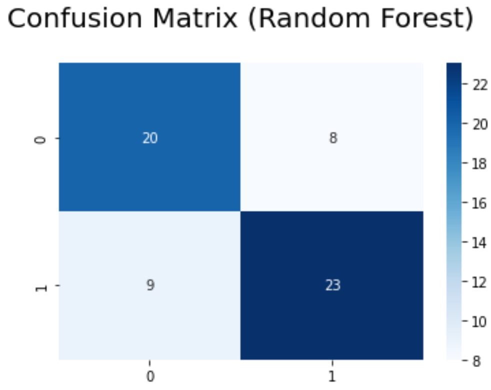


Figure 35: Confusion matrix random forest for balanced dataset

SVM

SVM has an accuracy of 81% for balanced dataset and 68% for the imbalanced dataset. The sensitivity and specificity scores are 0.85 for balanced dataset 0.46 for imbalanced dataset and 0.80 for balanced dataset, 0.86 for imbalanced dataset respectively. The AUC score using SVM is 0.85 for imbalanced dataset and 0.76 for the balanced dataset. Balancing the data has improved the specificity for SVM but it has reduced the sensitivity significantly. This implies that the rate of detecting instances from the negative class has increased the rate of detecting the positive class has reduced. The reduction in sensitivity is much more significant than the improvement in specificity which has decreased the AUC. The overall performance has deteriorated even with increase in specificity.

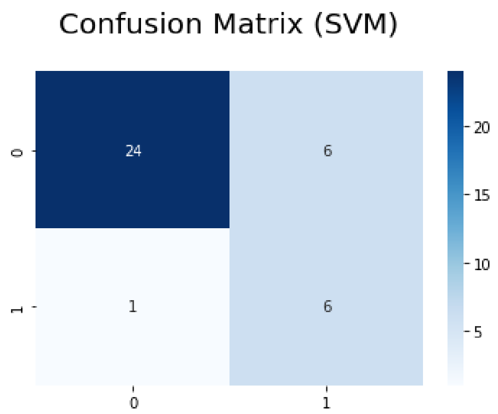


Figure 36: Confusion matrix svm for imbalanced dataset

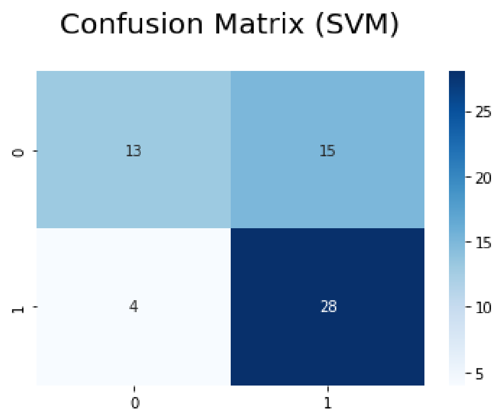


Figure 37: Confusion matrix svm for balanced dataset

Lgbm

Light gradient boosting algorithm (Lgbm) has an accuracy of 86% for balanced dataset and 68% for the imbalanced dataset. The sensitivity and specificity scores are 0.92 for balanced dataset 0.43 for imbalanced dataset and 0.93for balanced dataset, 1.00 for imbalanced dataset respectively. The AUC score using lgbm is 0.92 for imbalanced dataset and 0.87 for the balanced dataset. In case of Lgbm, the comparison between the performance of imbalanced and balanced data yielded a similar result as the previously discussed algorithms. The overall performance has decreased due to the decrease in detection rate of healthy fetuses. All the instances from the unhealthy fetuses are being detected for both imbalanced and balanced dataset but, the detection rate for healthy samples is extremely poor. Lgbm is the best performing algorithm for the imbalanced dataset which is performing poorly after balanced dataset is being supplied. Below is the confusing metrics generated for lgbm using bother imbalanced and balanced datasets:

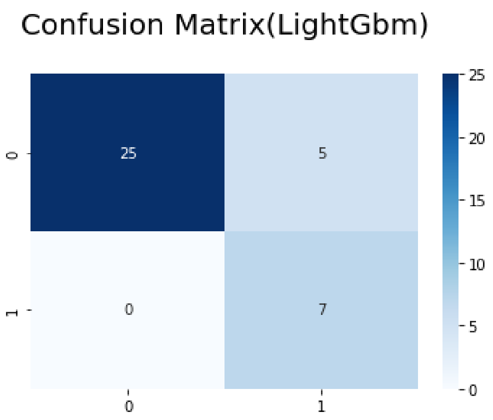


Figure 38:Confusion matrix lgbm for imbalanced dataset

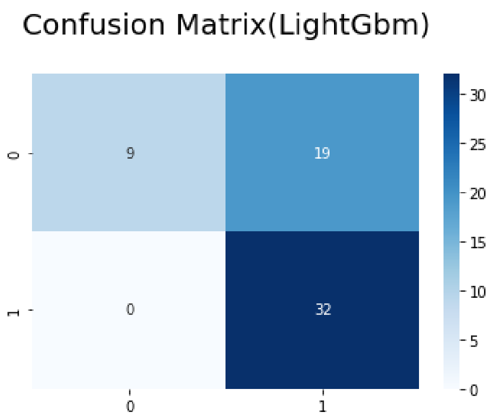


Figure 39: Confusion matrix lgbm for balanced dataset

XGboost

Extreme gradient boosting algorithm (XGBoost) has an accuracy of 84% for balanced dataset and 73% for the imbalanced dataset. The sensitivity and specificity scores are 0.84 for balanced dataset 0.73 for imbalanced dataset and 0.83 for balanced dataset, 0.80 for imbalanced dataset respectively. The AUC score using XGBoost is 0.84 for imbalanced dataset and 0.77 for the balanced dataset. Similar to the previously mentioned algorithms, using the balanced dataset decreases the overall performance of the classifier in its ability to detect instances of the healthy fetus class. Below is the confusing metrics generated for XGBoost using bother imbalanced and balanced datasets:

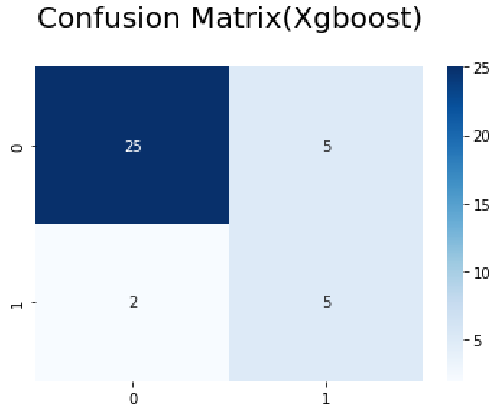


Figure 40: Confusion matrix xgboost for imbalanced dataset

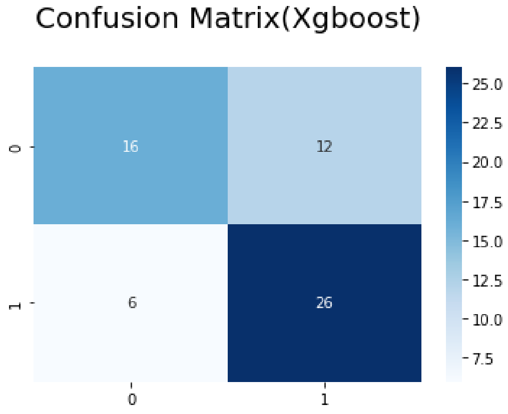


Figure 41: Confusion matrix xgboost for balanced dataset

Convolutional neural network

Convolutional neural network (CNN) has an accuracy of 86% for balanced dataset and 70% for the imbalanced dataset. The sensitivity and specificity scores are 0.76 for balanced dataset 0.70 for imbalanced dataset and 0.93 for balanced dataset, 0.80 for imbalanced dataset respectively. The AUC score using cnn is 0.85 for imbalanced dataset and 0.83 for the balanced dataset. Similar to the previously mentioned algorithms, using the balanced dataset decreases the overall performance of the classifier in its ability to detect instances of the healthy fetus class.

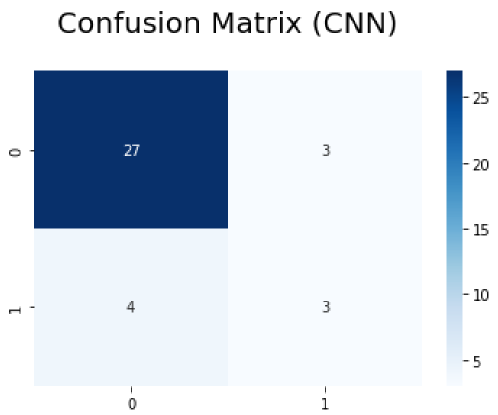


Figure 42:Confusion matrix cnn for imbalanced dataset

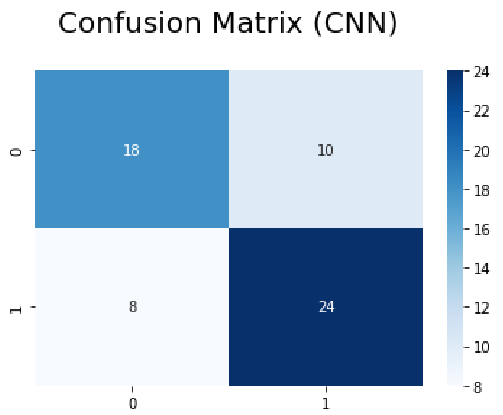


Figure 43: Confusion matrix svm for imbalanced dataset

Selected final algorithm

After overall analysis of the performance evaluation metrics for each of the 9 algorithms, it can be observed that for the imbalanced dataset the machine is learning more about the healthy fetus than the unhealthy fetus. For the balanced data using SMOTE the opposite is true. The unhealthy fetuses are getting detected more than the healthy fetuses. By looking at the overall performance of each of the algorithms for both imbalanced and balanced data Light gradient boosting algorithm or Lgbm trained and tested on the imbalanced dataset is the most preferable choice. Lgbm has a highest AUC score of all the algorithms in tested in both scenarios. Along with highest AUC score, the precision, recall and F1 score is the most balanced out of any other algorithms. From the balanced dataset scenario, decision tree is the one with best overall performance. Yet, Lgbm is more reliable option for real world scenario due to its ability to detect unhealthy fetuses accurately for all the instances provided in the test data for both imbalanced and balanced dataset. As the main aim is to detect unhealthy fetuses to ensure that preventative are being taken to mitigate the risks, the algorithm detecting the unhealthy fetuses correctly for majority of the times is the most optimal choice.

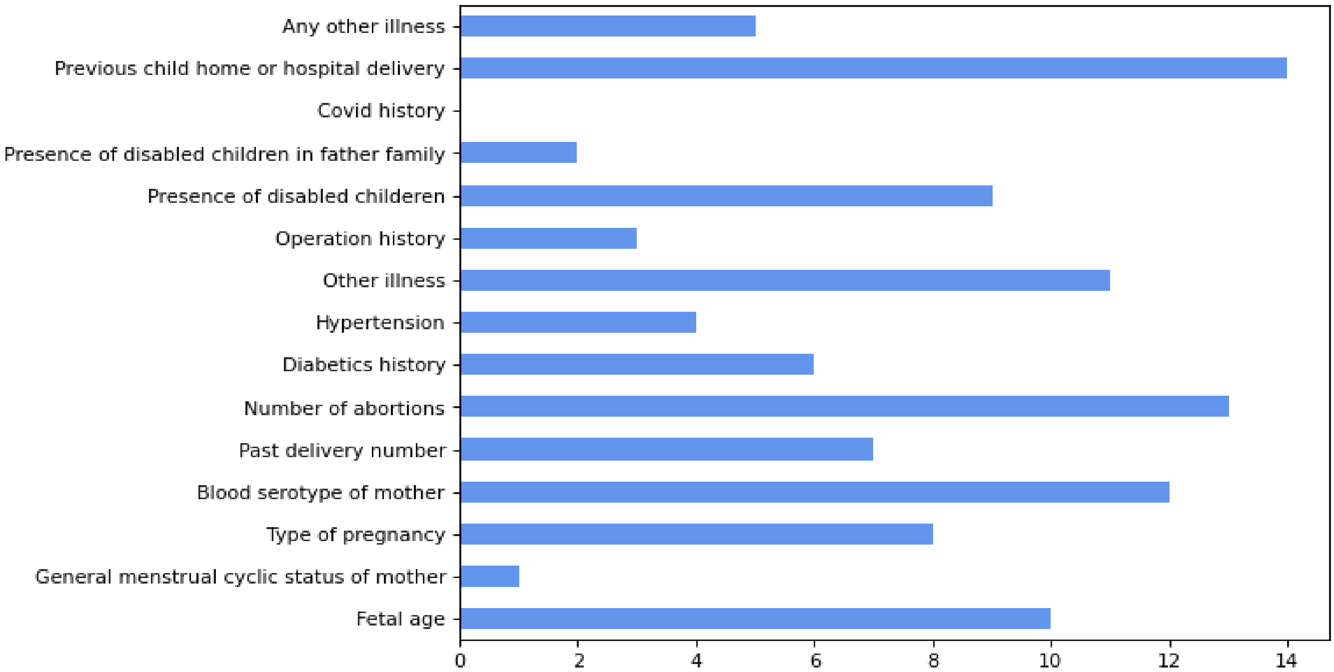


Figure 44: Feature ranking imbalanced data

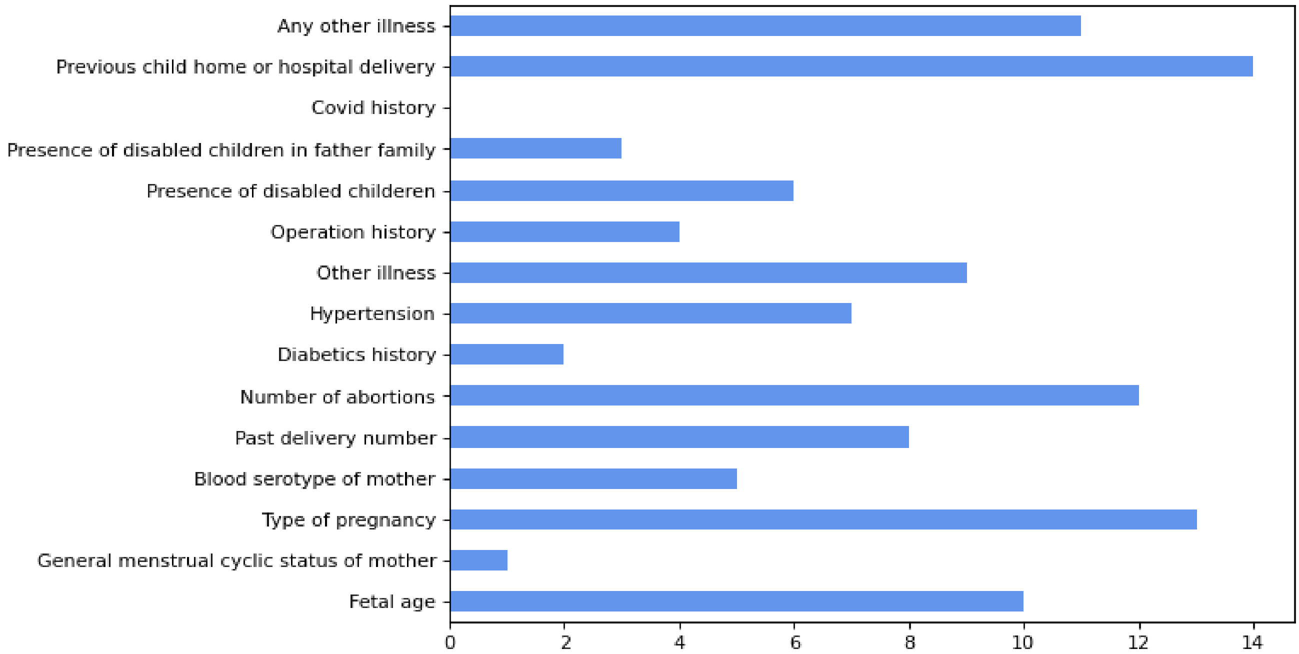


Figure 45: Feature ranking balanced data

Chapter 5

Discussion

The well-being and development of the fetus and the visiting contacts inside the womb of pregnant women are referred to as fetal health. Maximum pregnancy period complexities cause fetal to have a serious trouble that limits right development, coming about in lack or death. Harmless pregnancy period anticipating risk levels not long-ago challenges event increases legitimate fetal growth [23]. Experiments show that, when compared to classic machine learning models, Random Forests achieve the highest accuracy of 89 percent in a balanced dataset, while KNN achieves the highest accuracy of 80 percent in unbalanced datasets. Additionally, Decision Tree achieves a minimum accuracy of 67% in a balanced dataset, while SVM achieves a minimum accuracy of 66% in an unbalanced dataset. We have studied two approaches to obtain the possible results. This article showed the different approaches and inquiries about done so distant for determining fetal wellbeing and development state from a set of pre-classified designs information and its accuracy [23]. Predicting fetal well-being is essential when building a precognitive classifier with machine learning calculations [1].

This thesis, on the other hand, improved our understanding of how to use precision and accuracy on multiclass datasets. We utilize nine machine learning calculations in our work that each has their own distinct strengths and performs admirably but there were a few impediments, which is why we'll include a few more functionalities to make strides the precision of our framework [22]. Due to the limitations of fetal health records, this paper can be improved in many ways. We will work on it in the future [2]. The dataset in this document could have been improved by including more anomaly data. Feature extraction was not used in this task due to the impact of various features on the model.in order to comprehensively evaluate the performance of different machine learning models on fetal health data set. Further feature extraction will be performed in the future to improve the model's performance [15]. We are going be able to utilize much more advantageous or cross breed calculations in our work interior long term. Since machine learning calculations work in degree to the entirety of data, there's room for alter in our work [23]. In the future, we plan to make our website more powerful, user-friendly, and visually appealing. So that people can easily access that website and obtain critical information [1]. The quality of the think about was the application of nine distinctive machine learning approaches to the fetal wellbeing dataset and suggested the two with the most noteworthy precision on both the adjust and awkwardness fetal information sets [24]. It also used the SMOTE balancing strategy to avoid the model's bias toward skewed data, which improved the prediction accuracy of the machine learning algorithm. However, one significant disadvantage of this study is that the dataset was obtained from a developed hospital's repository [24]. The accuracy of the machine learning system may alter as the socio-demographic characteristics of pregnant women change. Besides, no data on participants' socio-demographic information or other pertinent clinical characteristics, such as primiparity, maternal wholesome status and frailty, gestational This dataset included information such as age, fetal well-being, and so on, which could potentially help refine the AI model [23].

Chapter 6

Conclusion

Restorative conclusion and forecast are inseparably tied to machine learning. Throughout machine learning, a dataset can be instructed with an appropriate machine learning calculation to predict a specific characteristic. This ability can help specialists take precautions or conduct a more thorough investigation of a situation [6]. In spite of the truth that the human body is amazingly complicated and therapeutic choices cannot be made completely on the premise of estimates, the concept of machine learning in pharmaceutical has the potential to be a valuable device for specialists and clinicians [9]. Postural wellness complications in pregnancy are a fundamental challenge for women around the world [8]. Poor maternal health, poor dietary habits, drug use, and birth defects are all common wellness issues that affect fetal enlargement. This research will assist pregnant women in making timely decisions to save both her and her child's lives [23].

In this research, we used nine machine learning computations to build a web-based resource that allows pregnant women to foresee almost negative fetal outcomes. A few reasons for changing the wellbeing of pregnant ladies have been anticipated employing a machine learning calculation [21]. But the 183 pregnant women dataset is not enough immense, the current form gives satisfactory classification results. We select a total of 16 features. All of them healthy was 150 and unhealthy was 33. For training purposes, we use a total of 146 training datasets and 37 for testing. We have done many experiments to find the most accurate classifier of all top SVM, LightGBM, XgBoost, CNN, Decision Tree, Random Forest, K Nearest Neighbors, Logistic Regression and Linear Regression to achieve the required performance [19]. Experiments demonstrate that, when compared to classic machine learning models, Random Forests achieve the maximum accuracy of 89 percent in a balanced dataset, whereas KNN achieves the highest accuracy of 80 percent in imbalanced datasets. Furthermore, Decision Tree achieve the minimum accuracy of 67 percent in a balanced dataset, whereas SVM achieves the lowest accuracy of 66 percent in imbalanced datasets. Smooth has been used to increase the results. By applying smoothing, we removed noise from the dataset and also kept crucial patterns from standing out. We looked at two ways to get the possible results. The primary is to augment the dataset to urge more precise comes about with that information. After receiving preliminary information about the women’s fetal health, taking exercises was encouraged for the pregnant women [6]. Although the proposed system does not replace a medical procedure, please provide customers with preparatory information and treatment proposals for future periods [1].

This thesis article teaches us how to collect data and how to use algorithms to get better results. It broadens our understanding of algorithms and procedures, as well as how to predict them [2]. In the end, we tried our best to find the best classifier that could work accurately for fetal health. We intend to improve our line of work by managing high-performance websites and adding new administrations and some additional highlights have been included in our framework to improve accuracy [15]. Because deep learning models perform proportionally to the amount of data, there is room for improvement in our work [17].

Bibliography

[1] Achiron, Reuwen, Anat Achiron, and Simcha Yagel. "First trimester transvaginal sonographic diagnosis of Dandy-Walker malformation." *Journal of clinical ultrasound* 21 (1993): 62-62.

[2] Christianson, Arnold, and Bernadette Modell. "Medical genetics in developing countries." *Annu. Rev. Genomics Hum. Genet.* 5 (2004): 219-265.

[3] Al Jazeera. “Bangladesh: One in Five People Live below Poverty Line.” *News | Al Jazeera*, Al Jazeera, 26 Jan. 2020, https://www.aljazeera.com/news/2020/01/bangladesh-people-live-poverty-line-200126100532869.html.

[4] Roy, Ashim, and Lela Shengelia. "A review on situation of congenital disorders and access to community genetics services in Bangladesh." *Annals of clinical and laboratory research* 4.2 (2016): 0-0.

[5] Fatemaq, K., et al. "Major congenital malformations among the newborns in BSMMU hospital." *Bangladesh Medical Journal* 40.1 (2011): 7-12.

[6] Islam, M. N., et al. "Pattern of Congenital Anomalies in Newborns in a Tertiary Level Hospital in Bangladesh." *Journal of Surgery Pakistan (International)* 18.1 (2013): 32-36.

[7] Quadery, Sheikh Elhum Uddin, Mehedi Hasan, and Mohammad Monirujjaman Khan. "Consumer side economic perception of telemedicine during COVID-19 era: A survey on Bangladesh's perspective." *Informatics in Medicine Unlocked* 27 (2021): 100797.

[8] Akbulut, A., Ertugrul, E. and Topcu, V., 2018. Fetal health status prediction based on maternal clinical history using machine learning techniques. *Computer methods and programs in biomedicine, 163,* pp.87-100.

[9] Cheng, W., Fang, L., Yang, L., Zhao, H., Wang, P. and Yan, J., 2014, December. Varying coefficient models for analyzing the effects of risk factors on pregnant women's blood pressure. In *2014 13th International Conference on Machine Learning and Applications* (pp.

55-60). IEEE.

[10] Lakshmi, B.N., Indumathi, T.S. and Ravi, N., 2015, October. Prediction based health monitoring in pregnant women. In *2015 International Conference on Applied and Theoretical Computing and Communication Technology (iCATccT)* (pp. 594-598). IEEE.

[11] Lakshmi, B.N., Indumathi, T.S. and Ravi, N., 2015, November. A comparative study of classification algorithms for predicting gestational risks in pregnant women. In *2015 International Conference on Computers, Communications, and Systems (ICCCS)* (pp. 42-46). IEEE.

[12] Sahin, H. and Subasi, A., 2015. Classification of the cardiotocogram data for anticipation of fetal risks using machine learning techniques. *Applied Soft Computing*, *33*, pp.231-238.

[13] Mu, Y., Feng, K., Yang, Y. and Wang, J., 2018. Applying deep learning for adverse pregnancy outcome detection with pre-pregnancy health data. In *MATEC Web of Conferences* (Vol. 189, p. 10014). EDP Sciences.

[14] Giger, Maryellen L. "Machine learning in medical imaging." *Journal of the American College of Radiology* 15.3 (2018): 512-520.

[15] Kononenko, Igor. "Machine learning for medical diagnosis: history, state of the art and perspective." *Artificial Intelligence in medicine* 23.1 (2001): 89-109.

[16] Li, Jiaming, and Xiaoxiang Liu. "Fetal Health Classification Based on Machine Learning." *2021 IEEE 2nd International Conference on Big Data, Artificial Intelligence and Internet of Things Engineering (ICBAIE)*. IEEE, 2021.

[17] Bhowmik, Pankaj, et al. "Cardiotocography Data Analysis to Predict Fetal Health Risks with Tree-Based Ensemble Learning." (2021).

[18] Yarlapati, Anisha R., Sudeepa Roy Dey, and Snehanshu Saha. "Early prediction of LBW cases via minimum error rate classifier: A statistical machine learning approach." *2017 IEEE International Conference on Smart Computing (SMARTCOMP)*. IEEE, 2017.

[19] Li, Jianqiang, et al. "Automatic classification of fetal heart rate based on convolutional neural

[20] Yu, Shuang, et al. "Lumbar ultrasound image feature extraction and classification with support vector machine." *Ultrasound in medicine & biology* 41.10 (2015): 2677-2689.

[21] Sahin, Hakan, and Abdulhamit Subasi. "Classification of the cardiotocogram data for anticipation of fetal risks using machine learning techniques." *Applied Soft Computing* 33 (2015): 231-238.

[22] Ball, Gareth, et al. "Machine-learning to characterise neonatal functional connectivity in the preterm brain." *Neuroimage* 124 (2016): 267-275.

[23] Ocak, Hasan. "A medical decision support system based on support vector machines and the genetic algorithm for the evaluation of fetal well-being." *Journal of medical systems* 37.2 (2013): 1-9.

[24] Jadhav, Shivajirao M., Sanjay L. Nalbalwar, and Ashok A. Ghatol. "Modular neural network based arrhythmia classification system using ECG signal data." *International Journal of Information Technology and Knowledge Management* 4.1 (2011): 205-209.

[25] A. C. of Obstetricians , Gynecologists , et al. , Physical activity and exercise dur- ing pregnancy and the postpartum period, Obstet. Gynecol. 126 (6) (2015) e135–142 . Committee opinion no. 650.

[26] Pradhan, A.K., Rout, J.K., Maharana, A.B., Balabantaray, B.K. and Ray, N.K., 2021, December. A Machine Learning Approach for the Prediction of Fetal Health using CTG. In 2021 19th OITS International Conference on Information Technology (OCIT) (pp. 239-244). IEEE.

[27] Mehbodniya, A., Lazar, A.J.P., Webber, J., Sharma, D.K., Jayagopalan, S., Singh, P., Rajan, R., Pandya, S. and Sengan, S., 2021. Fetal health classification from cardiotocographic data using machine learning. Expert Systems, p.e12899.

[28] Jezewski, M., Czabanski, R., Wróbel, J. and Horoba, K., 2010. Analysis of extracted cardiotocographic signal features to improve automated prediction of fetal outcome. Biocybernetics and Biomedical Engineering, 30(4), pp.29-47.

[29] Raj, P. and David, P.E., 2020. *The Digital Twin Paradigm for Smarter Systems and Environments: The Industry Use Cases*. Academic Press.

[30] Chinnaiyan, R. and Alex, S., 2021, January. Machine Learning Approaches for Early Diagnosis and Prediction of Fetal Abnormalities. In *2021 International Conference on Computer Communication and Informatics (ICCCI)* (pp. 1-3). IEEE.

[31] Hoodhood, Z., Noman, M., Shafique, A., Nasim, A., Chowdhury, D. and 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), p.226.

[32] Teng, L.Y., Mattar, C.N.Z., Biswas, A., Hoo, W.L. and Saw, S.N., 2022. Interpreting the role of nuchal fold for fetal growth restriction prediction using machine learning. *Scientific Reports, 12*(1), pp.1-10.

[33] Fung, R., Villar, J., Dashti, A., Ismail, L.C., Staines-Urias, E., Ohuma, E.O., Salomon, L.J., Victora, C.G., Barros, F.C., Lambert, A. and Carvalho, M., 2020. Achieving accurate estimates of fetal gestational age and personalised predictions of fetal growth based on data from an international prospective cohort study: a population-based machine learning study. *The Lancet Digital Health*, *2*(7), pp.e368-e375.

[34] Garcia-Canadilla, P., Sanchez-Martinez, S., Crispi, F. and Bijnens, B., 2020. Machine learning in fetal cardiology: what to expect. *Fetal diagnosis and therapy*, *47*(5), pp.363-372.

[35] Pisapia, J.M., Akbari, H., Rozycki, M., Goldstein, H., Bakas, S., Rathore, S., Moldenhauer, J.S., Storm, P.B., Zarnow, D.M., Anderson, R.C. and Heuer, G.G., 2018. Use of fetal magnetic resonance image analysis and machine learning to predict the need for postnatal cerebrospinal fluid diversion in fetal ventriculomegaly. *JAMA pediatrics*, *172*(2), pp.128-135.

[36] Fergus, P., Hussain, A., Al-Jumeily, D., Huang, D.S. and Bouguila, N., 2017. Classification

of caesarean section and normal vaginal deliveries using foetal heart rate signals and advanced machine learning algorithms. *Biomedical engineering online*, *16*(1), pp.1-26.

[37] Rahmayanti, N., Pradani, H., Pahlawan, M. and Vinarti, R., 2022. Comparison of machine learning algorithms to classify fetal health using cardiotocogram data. *Procedia Computer Science*, *197*, pp.162-171.

[38] Liu, L., Jiao, Y., Li, X., Ouyang, Y. and Shi, D., 2020. Machine learning algorithms to predict early pregnancy loss after in vitro fertilization-embryo transfer with fetal heart rate as a strong predictor. *Computer Methods and Programs in Biomedicine*, *196*, p.105624.

[39] Subasi, A., Kadasa, B. and Kremic, E., 2020. Classification of the cardiotocogram data for anticipation of fetal risks using bagging ensemble classifier. *Procedia Computer Science*, *168*, pp.34-39.

[40] Huang, M.L. and Hsu, Y.Y., 2012. Fetal distress prediction using discriminant analysis, decision tree, and artificial neural network.

[41] Rauschert, S., Melton, P.E., Heiskala, A., Karhunen, V., Burdge, G., Craig, J.M., Godfrey, K.M., Lillycrop, K., Mori, T.A., Beilin, L.J. and Oddy, W.H., 2020. Machine learning-based DNA methylation score for fetal exposure to maternal smoking: development and validation in samples collected from adolescents and adults. *Environmental health perspectives*, *128*(9), p.097003.

[42] Fergus, P., Huang, D.S. and Hamdan, H., 2016. Prediction of intrapartum hypoxia from cardiotocography data using machine learning. In *Applied computing in medicine and health* (pp. 125-146). Morgan Kaufmann.

[43] Batra, A., Chandra, A. and Matoria, V., 2017, February. Cardiotocography analysis using conjunction of machine learning algorithms. In *2017 International Conference on Machine Vision and Information Technology (CMVIT)* (pp. 1-6). IEEE.

[44] Kaur, P., Singh, G. and Kaur, P., 2018. A review of denoising medical images using machine learning approaches. *Current medical imaging*, *14*(5), pp.675-685.

[45] Reddy, G.T., Reddy, M.P.K., Lakshmanna, K., Kaluri, R., Rajput, D.S., Srivastava, G. and Baker, T., 2020. Analysis of dimensionality reduction techniques on big data. *IEEE Access*, *8,* pp.54776-54788.

[46] Devi, M.S., Sridevi, S., Bonala, K.K., Dadi, R.H. and Reddy, K.V.K., 2021. Oversampling Response Stretch based Fetal Health Prediction using Cardiotocographic Data. *Annals of the Romanian Society for Cell Biology*, pp.1448-1464.

[47] Singh, V., Agrawal, R., Gourisaria, M.K., Singh, P.K. and Das, H., 2022, April. Comparative Analysis of Machine Learning Models For Early Detection of Fetal Disease using Feature Extraction. In *2022 IEEE 11th International Conference on Communication Systems and Network Technologies (CSNT)* (pp. 169-175). IEEE.