Utilizing Advanced Machine Learning Techniques for Predicting Fetal Health Risks

MSc Research Project
Data Analytics

Prakruthi Barthur Prakash
Student ID: 23109742

School of Computing
National College of Ireland

Supervisor: John Kelly

National College of Ireland

MSc Project Submission Sheet

School of Computing

Student Name:	……………Prakruthi Barthur Prakash………………………………………
Student ID:	………………………X23109742…………………………………………………
Programme:	…………Data Analytics…………………… **Year:** 2023-2024.
Module:	………………Research Project…………………………………………………
Supervisor:	…………………John Kelly……………………………………………………
Submission Due Date:	……………………………12/08/2024………………………………………………
Project Title:	Utilizing Advanced Machine Learning Techniques for Predicting Fetal Health Risks
Word Count:	………………8816…………… **Page Count**……25……………..……..

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.
ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature: ………………**PRAKRUTHI BARTHUR PRAKASH** ……………………

Date: ………………………12/08/2024…………………………………………………………

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST

Attach a completed copy of this sheet to each project (including multiple copies)	□
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	□
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	□

Assignments that are submitted to the Programme Coordinator Office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Table of Contents

Abstract

The research study focuses on the application of machine learning algorithms to predict fetal health in the context of antenatal care. A predictive model is developed from such a dataset, including the baseline value, accelerations, fetal movement, uterine contractions, decelerations, without accelerations, and variability measures. Predictive models are developed, and the Boruta feature selection technique is utilized to identify the most critical features for the models. To address class imbalance, the SMOTE technique is used to increase the ability of the models to make reliable predictions across classes. Different machine learning models, such as Random Forests, GBM, Decision Trees, and K-Nearest Neighbors, are implemented on the dataset. Accuracy is combined with precision, among other performance metrics to inform the validity of the models and assess their predictive power. Among all the models implemented GBM performed well with 98% accuracy. The implications of such findings can change antenatal practices, reducing the risks associated with birthing and improving women and newborn health outcomes. The need for models capable of predicting abnormal fetal health is the research objective.

1. Introduction

Prenatal care of the fetus is also an essential part of pregnancy and ongoing research is searching for ways to better diagnose possible fetal pathology (Li and Liu, 2021). As evidenced by the foregoing analysis, proper risk assessment and intervention at the appropriate time is highly useful in preventing adverse effects of nasty external factors on the well-being of mother and child and enhancing the possibility of delivering a healthy baby. Standard antenatal care relies on either the expertise of healthcare practitioners or diagnostic tools to determine the condition of the unborn child (Salini *et al.*, 2024). However, there has been progress in applying various algorithms and architectures related to machine learning which have unveiled new prospects for a finer assessment of the risk and a better prognosis of the baby's condition. This research attempts to establish the justification and the reason why there is a need for further research in this area while reviewing why the machine-learning approach in the assessment of fetal health risks was examined. It also speaks to the possibility that the addition of this research project may contribute to the existing literature.

Integrated machine learning algorithms hold several advantages in this regard. Due to their successful ability to map the connections between data points, these algorithms can find more subtle patterns and connections in the dataset than analysts can. Through exposure to a large volume of historical data, it becomes easier for algorithms to learn different signs and relations that may prove helpful in the detection of fetal distress or possible complications. This predictive capability may be very helpful for healthcare professionals to take timely action and to take necessary protective measures for the mother and the embryo's health.

Several challenges must be addressed, however, to make the predictive models reliable and sound. It may be a problem, for example, since not every variable within the studied dataset can be significant for fetal health prediction equal to others. Selecting the most informative features, in other words, the features that carry a lot of predictive capability is not easily done. Furthermore, it was indicated that class imbalance where some classes dominate the others can influence the effectiveness of the models. Some measures must be undertaken to solve this problem and ensure the reliability of the forecast in all the studied aspects of fetal condition.

Thus, the predictive models under consideration will be assessed by various performance indicators encompassed in accuracy indicators, such as accuracy, recall, precision, F1-score, and confusion metrics. All these measures offer an extensive assessment of the ability of the models to predict the health status of the fetus correctly. In this way, we can get a measure of the relative performance of the models and compare it to existing clinical strategies.

This study's findings can potentially alter the nature of antenatal care as well as risk assessment services. The up-to-date status of an unborn child's health determines whether a specific pregnancy is high risk, or not, and therefore, if it requires closer supervision and interventions that can make it more successful both in terms of the mother's and child's health.

1.1 Research Question

How well can machine learning algorithms predict fetal abnormalities?

1.2 Motivation

Proper fetal care should be taken to note and control any potential risks hence increasing the chances of a successful birth. Conventional fetal care puts the responsibility of caring for the fetus on health practitioners and the conventional tools for diagnosing fetal health. However, advances in several of the identified ML approaches have provided new ways of predicting the health problems of a fetus.

In the line of fetal health abnormalities, this research will in any way assist in the prediction of any change in the state of the fetus, especially in cases of fetal abnormalities. However, there is still much more that needs to be done, despite these fantastic achievements in attributes of the fetal age and medical advancement. The current approaches sometimes entail missing out on diagnoses or delayed interventions due to clinical judgment or inadequate data. Besides, one should mention the great complexity of fetal growth and the numerous possibilities of maternal–newborn interactions during pregnancy.

In this research work, after the introduction, Chapter 2 will contain a literature review aimed at situating the research within the academic literature. The type of approach that has been used and the strategies applied in data pre-processing as well as model development are described in Chapter 3. Chapter 4 is on research design while Chapter 5 is on the practical implementation of the research. The theory of the research approaches and the procedure of the predictive models are discussed. Chapter 6 contains a critical view of the experiments that have been made and the conclusion on the efficiency of the models used in the work. Last, Chapter 7 reviews the conclusion of the research, studies the encountered limitations and outlines the prospects of the research.

2. Related Work

In this review, an attempt has been made to summaries various research papers that are related to the development of predictive models and techniques/measures employed to predict and understand the state of fetal health in prenatal diagnosis. Highlighting various research datasets and methods provides a general view of the current state of the research in the categorization of fetal health.

2.1 An overview of the techniques of machine learning in the context of fetal health tracking.

Daniela Mennickent et al. (Mennickent *et al.,* 2023) explores many of the machine learning techniques that are gradually enhancing prenatal analysis. They explain the existing strengths and weaknesses of the research area, the issues that are yet to be solved for instance how to handle unstructured data and enhancing the current real-time prediction. It assesses the merits and pitfalls of various approaches and integrates the outcomes of multiple studies to give a detailed assessment of every algorithm's contribution. This analysis helps in the understanding of how the process of machine learning can be further improved in relation to the health of mothers and their babies.

2.2 The modern developments in the application of machine learning in fetal health classification

Sujith K Mandala (Mandala *et al.,* 2023) presents LightGBM classifiers in ML for the first time in the context of fetal health classification. This text builds on the basic machine learning approaches described by Daniela Mennickent et al. (Mennickent *et al.,* 2023). This approach contributes to the development of the mentioned issues, since it directly addresses the issues related to the complexity of data and the small amount of available labelled examples that were mentioned by Daniela Mennickent et al. (Mennickent *et al.,* 2023). Mandala's practice is mainly based on obstetrics, especially on the diagnosis and treatment of fetus complications. This not only solves the methodological problems but also increases the diagnostic possibilities. It presents a detailed overview of the traditional approaches to assessment with a focus on a highly effective machine learning-based solution that can handle such problems as data imbalance and interpretability. This assessment not only reveals that the LightGBM classifier has the highest accuracy compared to all the other models discussed in this study but also stresses the importance of further analysis on other models that are scalable and understandable. It presents a significant contribution to machine learning for obstetrics, future works could be done to enhance the current methods and solve the existing issues.

2.3 Technological developments that have been made in the estimation of gestational age of a fetus.

In a study by Lok Hin Lee et al. (Lee *et al.,* 2023) the authors use the most advanced approaches of machine learning to predict the gestational age from the ultrasound images and achieve higher accuracy than traditional biometry estimates. The discrepancies in the analysis of ultrasound findings are being solved; this is in line with Mandala's attempts to enhance the diagnostic performance using modern technologies. The combination of the neural networks

and the regression algorithms enhances the prenatal care techniques as they put forward a new machine learning technique that solves the problems that are present with the current estimates made from the ultrasounds.
The validity of the model has been proven by applying it to different data sets and in different conditions such as cases of intrauterine growth restriction. The necessity of the ability of ML models to work with different image quality and enhance interpretability is discussed. The relevance of using standardized ultrasound planes is recognized and some recommendations for further investigations are made. The current work is most relevant to the study as it improves fetal health evaluations by providing a more accurate estimate of the gestational age. The approaches and findings discussed in this work provide a solid foundation for future research on the application of ML in prenatal care. The integration of these finding highlights the potential of machine learning in revolutionizing obstetric science.

2.4 The Use of Hybrid Machine Learning Methods in Obstetrics.

Following the same logic and with a focus on the topic of the paper "Intelligent diagnosis of obstetric diseases using HGS-AOA based extreme learning machine" Ramesh Vatambeti and Vijay Kumar Damera (Vatambeti and Damera, 2023) introduced a new work that embodies the use of Hunger Games Search and Arithmetic Optimisation Algorithms to improve Extreme Learning Machines' ability in diagnosing obstetric complications. This work also aligns with Lee's approach of employing state-of-the-art machine learning strategies while also dealing with the problem of class imbalance, which has been discussed in prior research. Application of a hybrid model shows the best practice in machine learning applications, especially in the aspect of diagnosis that is very vital in prenatal health.

2.5 Imperative to analyze and categorise ways of assessing the health condition of the foetus.

In the paper "Using machine learning to classify human fetal health and analyze feature importance" from the year 2023, Yiqiao Yin and Yash Bingi (Yin and Bingi, 2023) focus on the analysis of cardiotocography data for fetal health classification using the machine learning models like Support Vector Machines (SVM) and Extreme Gradient Boosting (XGBoost). The present work is devoted to the transition from the hybrid machine learning models to the more specific classification methods. This work enhances the model's decision-making process by using explanation techniques like SHAP and LIME. This is a crucial issue in the present debate on model interpretability and decision making in machine learning systems to be employed in healthcare scenarios.

2.6 Optimizing the machine learning models for better predictive accuracy through comparison.

Kabir Singh et al. (Singh *et al.*, 2023) and Jiaming Li and Xiaoxiang Liu (Li and Liu, 2021) have contributed significantly to the advancement of the field through the development of new methods for assessing the health of fetus and the incorporation of various machine learning models into a 'Blender Model'. This is evident in Singh's work in integrating image processing with ML techniques for use in medical diagnosis. This article provides a detailed discussion of these techniques and presents a nuanced understanding of their strengths and weaknesses in a clinical context. This approach gives a clear understanding of the application of ML in the monitoring of the health of the foetuses, an important aspect of prenatal care. The use of ensemble learning techniques by Li is effective in capturing the strengths of combining multiple models to enhance predictive performance as opposed to relying on a single model. The effectiveness and reliability of these integrated approaches are systematically evaluated according to metrics that demonstrate their advantages over single-model applications and provide a solid foundation for further research on the usage of machine learning in the prediction of fetal health.

2.7 Current Trends and Possible Future Trends in the Classification of Fetal Health in Newborns.

In the research paper "Extreme Gradient Boosting based Fetal Health Classification" by, S. Neelakandan et al., (Neelakandan *et al.,* 2023) apply XGBoost to predict fetal health. Their findings indicate that accuracy, precision, and recall are the highest in the XGBoost model than the other machine learning models as it performs optimally. This evaluation not only proves that XGBoost can handle complex data while being precise but also shows a clear example of how machine learning can be applied in the healthcare industry. To build on this subject, Yalamanchili Salini et al., (Salini *et al.,* 2024) apply a concept of machine learning to enhance the capabilities of fetal health classification to achieve a fair and reliable assessment than the conventional ones. Their work properly compares several ML models and got a high precision of 93%. This shows that there has been a distinct improvement in the objectivity and credibility of fetal health assessments. Nurul Fathia Mohamand Noor et al. (Noor *et al.,* 2021) in their paper "Fetal health classification using supervised learning approach" use multiple supervised learning classifiers and highlights that the combination of Ada Boost and Random Forest gives the best diagnostic results. This analysis of a vast number of supervised learning algorithms improves the existing practices of the field and provides a strong background for the present and future development of machine learning applications for medical diagnostics. In general, the studies at hand are related to the mentioned gaps, for example, the need for better methods for the prediction of complex systems with large and unstructured or incomplete data sets, thereby helping to understand how these technologies can be further improved.

2.8 Synopsis and Integration to the Previous Research Work.

This literature review has systematically analyzed and compared the development of machine-learning approaches in fetuses health monitoring. All the works contribute to the field of research in one way or another, either by introducing new methods, developing the existing ones, or integrating diverse strategies to increase the specificity and predictive value of the diagnostic techniques. Altogether, these works provide a solid ground for future research on the application of machine learning (ML) for the early identification of fetal abnormalities. They particularly concentrate on some concerns like, Data quality, Model interpretability and Real-time analysis skills in prenatal care. Thus, including these observations, the context and justification for further research on the use of machine learning algorithms for predicting fetal anomalies will be further improved.

Sl No.	Author/Year	Approach	Type/Method	Accuracy	Limitations
1.	Mennickent, Rodríguez, Opazo, Riedel, Castro, Eriz-Salinas, Appel-Rubio, Aguayo, Damiano, and Araya (2023)	A brief of the machine learning techniques used in prenatal analysis	SVM , RF , NN	-	Challenges with the unstructured data and enhancing the real time prediction
2.	Mandala (2023)	Advancing Fetal Health Classification through Machine Learning	LightBGM , CatBoost	94%	Large amount of data and limited amount of labeled data.
3.	Lee, Bradburn, Craik, Yaqub, Norris, Ismail, Ohuma, Barros, Lambert, Carvalho, and Jaffer (2023)	Estimation of fetal gestational age based on ultrasound images	Neural Network and regression algorithm	-	-
4.	Vatambeti, and Damera (2023)	Hybrid machine learning methods	Hunger Games Search (HGS) and the Arithmetic Optimization Algorithm (AOA)	91%	Class imbalance issue
5.	Yin, and Bingi (2023)	Classify fetal health	SVM	94%	Limited condition to compare
6.	Singh, Shyry, and Franklin (2023)	Image processing approach	Neural Network	-	Not well noted approach

7.	Li, and Liu (2021)	Ensemble methods	CatBoost , LightGBM, KNN, DT	95%	-
8.	Neelakandan, Brinda, Dheekshitha, and Priya (2023)	XGB based classification	RF , SVM , LR , XGB	92%	Lack of good feature selection and Class balancing
9.	Salini, Mohanty, Ramesh, Yang, and Chalapathi (2024)	Analysis of CTG Data	RF, LR, DT, SVM	93%	No class balancing technique
10.	Noor, Ahmad, and Noor (2021)	Supervised learning approach	KNN, SVM, AdaBoost, DT	94.7%	Not effective feature selection method
11.	Chandrika, V. & Surendran, S. (2022)	Incremental Machine Learning Model for Fetal Health Risk Prediction	Random Forest, Adaboost Classifier	94%	Difficulties with incremental learning in applications where data is constantly being updated
12.	Parvataneni, K., Zaidi, S.H., Kazmi, F. & Kazmi, S.H.A.(2023)	AI Framework for Fetal Health Risk Prediction	RF,XGB, Catboost	95%	Extent of framework development and its compatibility with the existing healthcare management systems
13.	Reddy, A.I. & Priya, W.D.(2023)	XGB Classifier Compared Over KNearest Neighbor Classifier	XGB, KNN	-	Certain enhancements require to be made with regard to the accuracy in order for it to be used in other areas.

14.	Shifa, H.A., Mojumdar, M.U., Rahman, M.M., Chakraborty, N.R. & Gupta, V(2024)	Machine Learning Models for Maternal Health Risk Prediction	SVM, Naïve Bayes, XGB , DT	95%	Concerns about the model's reliability and its ability to provide interpretable results in realistic clinical environments.
15.	Rahman, A. & Rabiul Alam, M.G.(2023)	Maternal Health risk using ML and Deep Learning	Adaboost , SVM, RF, DT	91%	-

Table 1: Summary of the Related Work

3. Research Methodology

In this section, I would like to present the detailed pre-processing steps along with the procedures that will be applied to improve the quality of the dataset obtained from the Kaggle repository. There are 2126 rows and 22 numbers of columns in the given dataset. The method is to balance representation construction and achieve, analyze data's efficiency and coding achievement, and understand feature significance, compile high-quality information. Each of them is crucial to producing the desired model with high training and fine-tuning the network to explore the dataset. In the present study, CRISP-DM methodology was employed, the details about that procedure and the steps involved in doing so are depicted in the following Figure 1.

Fig.1 CRISP-DM Methodology

3.1 Business Understanding

The intended goal of this Machine Learning-based fetal health classification project is to increase the better diagnosis accuracy and less timeframe, trying to reduce the total healthcare costs, be helpful for the doctors and nurses and, better health results for patients. This specifically relates to healthcare providers and medical device manufacturers of which the activities of the project provide data acquisition and preparation of superior quality medical data. In this instance, stakeholders can be recognized as pregnant women, midwives, managers of healthcare facilities, manufacturers of the devices, and policymakers. The expected benefits include early detection of health issues, effective targeting of resources, increased patient trust, reduction in costs, and increased market competitiveness. The ML-based system will be thoroughly evaluated where concerns about data quality, explaining the model's decisions, legal requirements when implementing the new system, integration of the system into the existing system, and ethical issues will be tackled. The steps will assist in achieving the timely and effective implementation and, in turn, the management of the system that is set to revolutionize prenatal care.

3.2 Understanding of Data

When an organization or individual is to employ a technique of Machine Learning as in the fetal health classification project, it is significant to have an adequate understanding of the available data. The data consists of patient data and medical records consisting of attributes such as the fetal heart rate, basal value, gestational age, fetal movements, contractions of the uterus and other related physiological characteristics. The data needs to be very selective, and high quality and include several populations to ensure that the performance of the model is exceedingly strong and reliable. The analytics includes data preprocessing activities like handling missing values, normalizing the data, and transforming the inputs are essential in transforming the raw inputs into forms that the model can use. One factor that should be ideally present in the acquisition of the dataset is that the size must be large enough to contain a variety of conditions relating to the health of the fetus.

3.3 Data preparation

Cleaning the data is a process which is essential when examining data samples in which errors or any missing data are corrected. Thus, when it comes to my specific case, it was observed that there are no null values in the data set. Consequently, it was not necessary to apply techniques of imputation in the work with missing data. Since I did not have null values in my dataset, I did not need to clean the data before continuing with the analysis. It also enabled me to focus on other aspects of the data preparation and analysis procedures.

3.4 Exploratory Data Analysis (EDA) and Statistical Information

Exploratory Data Analysis (EDA) libraries (Matplotlib, seaborn, Plotly) and methods (head, summary and describe function) are used to first gain a good understanding of the provided data. Descriptive statistics consists of frequency distribution mean, standard deviation, minimum value, maximum value, and quartiles to study the distribution, variability, and midpoint of numerical variables. To check if the data collected is comprehensive, the function "info" is used to review the data types and several non-missing values. Non-parametric approaches like visualization of the distributions of the class variables are used to determine the extent of the class imbalance problem and to familiarize with the characteristics of the dataset.

3.5 Feature Selection using the Boruta model

Feature selection is applied to enhance the performance of the model in reducing over-fitting and for easier model interpretation. The Boruta model is one of the most used strategies in conducting feature selections.

Boruta is a random forest-based method, thus it is used for tree models such as Random Forest or XGBoost. The Boruta procedure eliminates attributes one by one starting with features that are significantly less important than a random probe, which are noise variables introduced by the Boruta algorithm. Each time around, the values rejected must be purged to be out of contention in the next run of the loop. It especially culminates in a fair global optimization for feature selection, which is the reason I appreciate it. With a probability, Boruta generates ‘‘shadow’’ versions of features in the data (noise) and runs the feature against the shadow to check if the feature is superior to the noise and thus worthy of a place within the feature set. Running Boruta will give the attributes which are confirmed, tentative, and rejected variables per run.

3.6 Data Balancing using SMOTE

When looking at the dataset EDA, after analysis, it was observed that the dataset has a class imbalance issue, to overcome this problem I have used the Synthetic Minority Oversampling Technique (SMOTE). It is explicitly focused on facing the issue of imbalance by creating new samples for the minority class only. This looks at the importance of SMOTE in handling class imbalance problems especially considering enhancing the efficiency of classifiers. Thus, SMOTE minimizes the effect of bias and aids the model in identifying vital features of the minority class, resulting in improved predictive models and overall performance.

3.7 Data Normalization or Feature Scaling

As for the feature scaling of the data, in my dataset, I used the StandardScaler to normalize the values. StandardScaler can be used to standardize a feature, it has a scaler to do this, first, it subtracts the mean of the features it is standardizing, and then divides the result by the standard deviation to bring it to unit variance. Unit variance equals dividing by the standard deviation of all the numbers. Standardization can help where the data has a Gaussian (or Normal) distribution is present. Through such a good scaling method, all the features of the dataset are evenly put into the process of training a model.

3.8 Model building

After all the pre-processing phase involved data transformation, handling missing values and other qualities, dealing with class imbalance problems, and finally segregating the data into two distinctive sets. To elaborate, during the gathering of the data, 80% of the data was assigned to the training phase while 20% of the data was assigned to the test phase. Such separation was important when testing the abilities of the models on new data.

3.9 Testing and Evaluation

Evaluation is a process of making a critical or against decision about the merits of the research. In other words, assessment is a general process of judging the worthwhile, and appraisal is the specific activity of arriving at the judgment. Another essential component that speaks of the efficiency of the model is an evaluation exercise carried out on the testing dataset. Accuracy is the broadly used measure to evaluate the model's predictive capability and the same is adopted to determine the efficiency of the proposed model by using accuracy metric which is frequently used in the domain of classification tasks. Along with accuracy, there are other parameters such as confusion metrics as well.

4. Design Specification

The four models chosen are impacted by the fetal health risks prediction research that were reviewed: Decision Tree, Gradient Boosting Machine, KNN, and Random Forest.

4.1 Decision Tree

Decision Tree (Li and Liu, 2021) is a nonparametric learning model that is applied in the area of supervised learning for classifications. The basic decision rules related to the target variable are collected from the data characteristics and the model generates the predictions of the target variable. Tree structure contain nodes that represent the characteristics, arcs that represent rules of decision and leaf nodes that represent the result. The primary advantage of the Decision Trees is understandable, and the model is rather simple. They can analyze numeric and categorical data and usually, there is little need to clean the data first. However, Decision Trees are prone to overfitting, and this is especially if the decision tree has become highly complex. Measures like pruning and the constraints on model complexity, which can be some restrictions like the maximum depth of the tree, and minimum samples per split, can greatly minimize this risk.

4.2 GBM

Gradient Boosting Machine or GBM (Mandala *et al.,* 2023) is one of the strong ensembles learning algorithms designed for carrying out both classification and regression tasks. It is an iterative procedure including the building of a series of models where each successive model attempts to make the corrections of the previous models' errors. The basic idea behind GBM is that it aims at accumulating several weak learners together to build a strong learner, typically the decision trees used.

The process in GBM starts with utilizing a primary model, which is commonly a decision tree, in the data set. After that, the new remaining errors that are left out from the first model, the next model is then developed to predict these residuals. This procedure is repeated for a set number of cycles, at the end of which the subsequent model improves the weakness of an ensemble of all earlier models. The final analysis is then obtained when the prediction from each model is summed up and normally, a weighted procedure is used.

Some of the significant features of GBM which can be underlined are the following: The GBM algorithm is capable of handling various types of input data and the ability to fit complex non-linear relationships between features in the data with the help of the iteration boosting step. However, it can be seen that the Gradient Boosting Machine (GBM) overfits since it is much sensitive to the number of iterations when this runs too high. To solve this problem, it is necessary to use certain variables that can affect the contribution of a new model, namely the learning rate which regulates the effective contribution of the new model, and shrinkage and subsampling techniques. Additionally, parameters like the depth of each decision tree, the number of trees (iterations), and minimum samples that should be reached to split a node are also significantly influential for the model's effectiveness and difficulty level.

4.3 Random Forest

Another technique of the ensemble learning model is Random Forest (Chandrika and Surendran, 2022), which is like a decision tree, but its training involves building thousands of trees and the final classification is given by the classification that most of the trees arrived at. While bagging helps in reducing 'enhanced generalization' and over-individuation of the particular decision tree, it aids in increasing the competency of the same. Bagging on the other hand entails the creation of several subsamples of the data whereby a sample is selected randomly from the data with replacement. Moreover, every one of these subsets is applied to build an individual decision tree for classification. Random Forests are random in a way that it adds random features to each tree in it and in each split of a tree, it randomly selects features. This makes the degree of tree types diverse hence enhancing the ability of the model to perform. Random Forests are the least overfitting structures, perform well on massive amounts of information, and have comparatively less variance for noisy info. But they can be computationally heavy and less interpretable in comparison with Shapley trees but its individual decision trees are.

4.4 KNN

K-nearest neighbors (KNN) (Noor *et al.*, 2021) is one of the simplest forms of learning that involves instances for classification problems. It works based on saving all the previous cases and categorizing the new case with the help of a similarity measure, often distance measures like Euclidean distance. The process of classification of a new instance in the case of KNN is done based on the majority or an average of K nearest neighbors in the feature space. The choice of K is crucial as if K is low the model will be sensitive to noise and on the other hand, if K is high, it will oversmooth the prediction. KNN lies in the non-parametric and lazy learning category it does not impose any constraint on the data distribution as well as does not include a training phase. Nevertheless, it also means that no assumptions are made and that no training is performed during the prediction process which may lead to computational costs.

5. Implementation

First, several works were carried out to pave the way for the evaluation. To get and read the dataset I used Visual Studio (version – 1.91.0) and made the correct import of packages for later analysis. Before performing the analysis, the data was imported, and I went through the data to check for empty or null values, but I didn't find any empty or null values. This step helps to ensure consistency in the results.

Next, I continued analyzing the info that is included in each of the columns in the offered dataset. This included checking on the types of data on the columns, the values they have as well as any other peculiarities of the columns. This knowledge made when pre-processing the data as well as when performing feature selection much easier and more informed.

Finally, in the basic statistical analysis of the columns, I was able to get insights into the disposition of the data, trends, and other peculiarities that might be present in the column. It was the basis for subsequent stages of EDA in which the model to be developed and analyzed would be decided with statistical considerations in mind.

5.1 Exploratory Data Analysis (EDA)

Regarding data analysis, the first step that I pay attention to is the EDA which entails a quick analysis and exploration of the given data to notice any features that may be present in the set of data as well as highlighting any unusual features existing in the data and shows how data elements relate to one another. This information is very helpful in making decisions when modelling has advanced to the later stages. In this section, I consider several analyses like fetal moment, histogram mode, abnormal short-term, and class distribution. In this way, I can add better benchmarks of reliability and dependability to each of our predictive models and their fundamental structures are therefore solid.

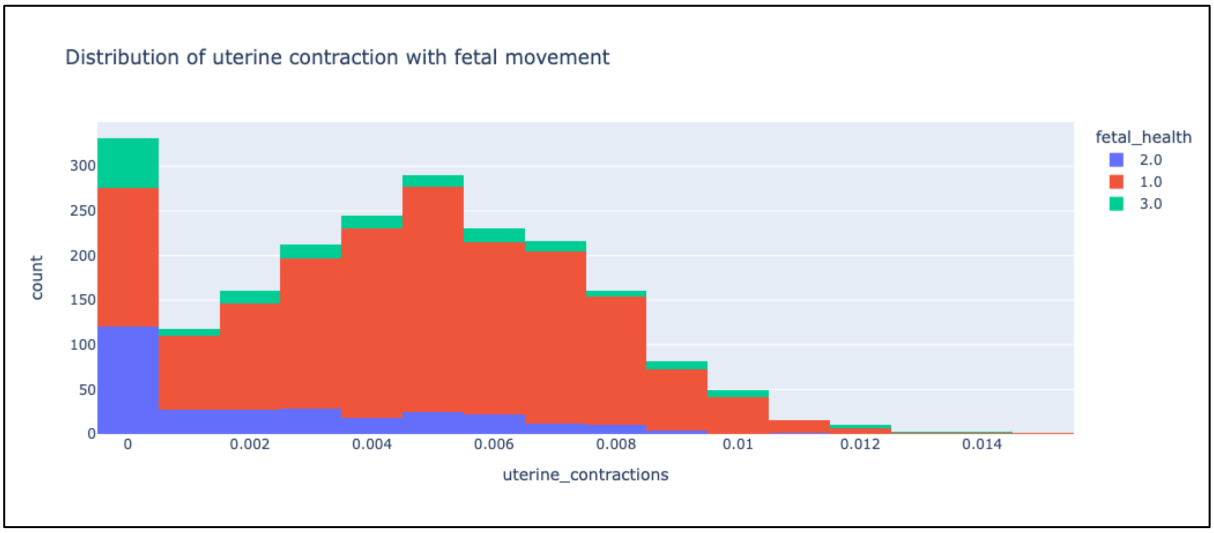

Fig.2 Uterine Contraction and Fetal Movement EDA

Figure 2. show shows the distribution of uterine contractions concerning the fetal health condition, it helps to at what contraction level the health risk would be most.

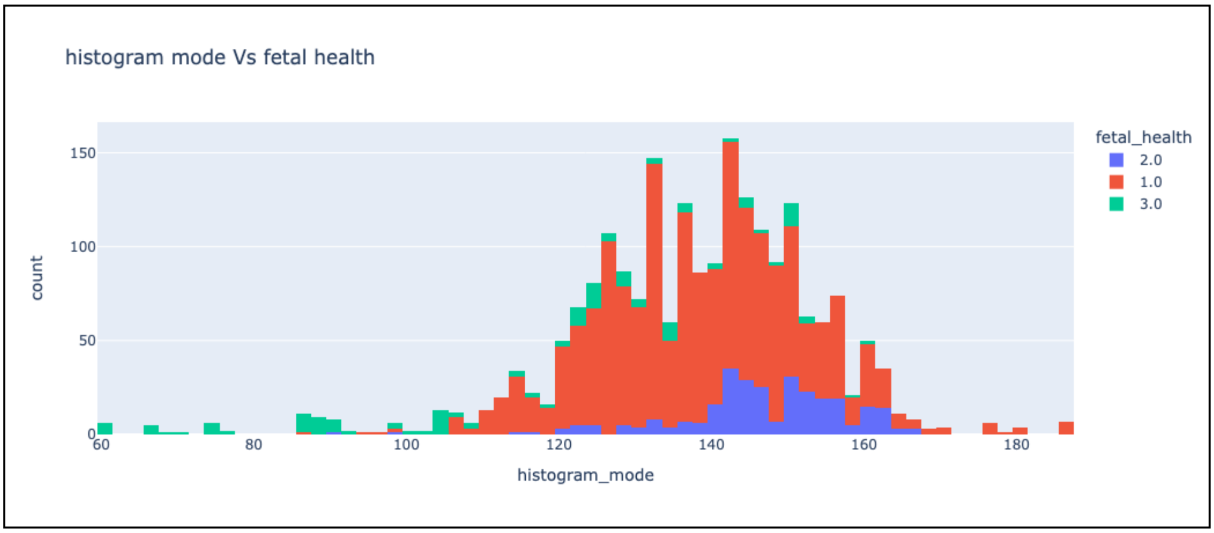

Fig.3 Histogram Mode and Fetal Health EDA

Figure 3. shows the values of histogram mode concerning the fetal health state, and provides the deviation range where the health risk major when the histogram value is high which indicates from the values 120 to 160.

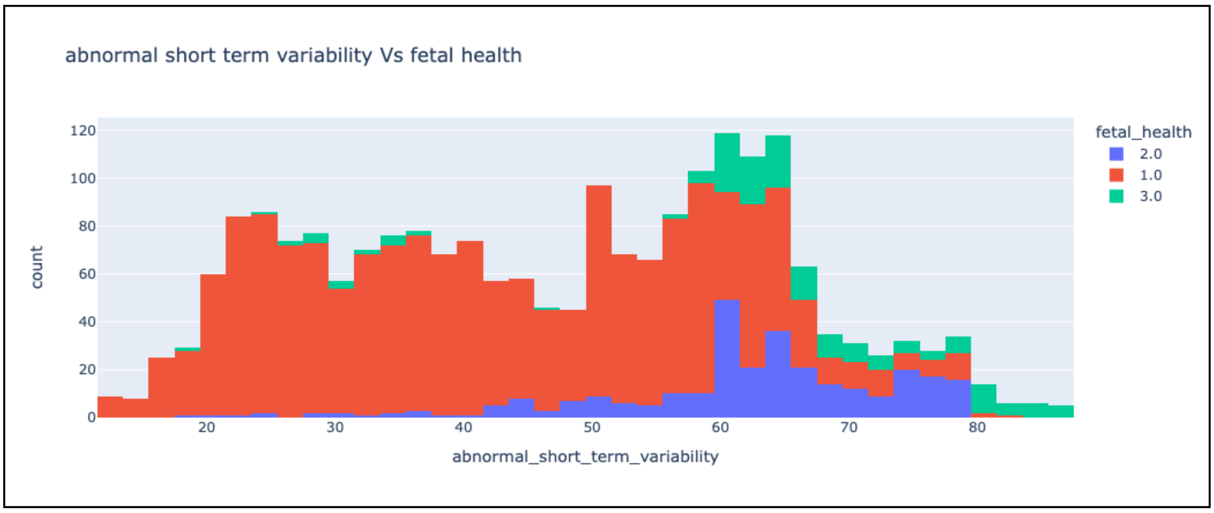

Fig.4 Abnormal Short-Term Variability and Health of Fetal EDA

Figure 4. represents the graph of abnormalities in a short term with rest to a fetal health condition. The high chances of abnormal short-term variability are when the values reach 50 or above.

5.2 DATA BALANCING

Data balancing is the procedure that resolves issues connected to the disproportionate distribution of classes in classification problems. The model may get a bit biased and give a bad result to the underrepresented classes when some classes have much fewer samples than the other classes. The distribution of classes can be made to be fairer by carrying out oversampling (creating new samples for the minority classes). This ensures that the model gets acquainted with instances that are uniformly distributed across the classes, increasing its accuracy in each class.

To address this problem, I applied the Synthetic Minority Over-sampling Technique more commonly referred to as SMOTE. This helps in a way to address the issue of class imbalance and hence the quality of the sample data set is also enhanced. This is vital in creating models that can generalize and give better predictions and at the same time, be fair.

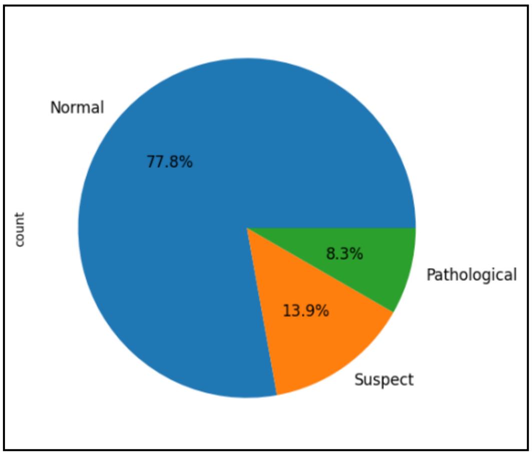

Fig.6 Data Imbalance without SMOTE

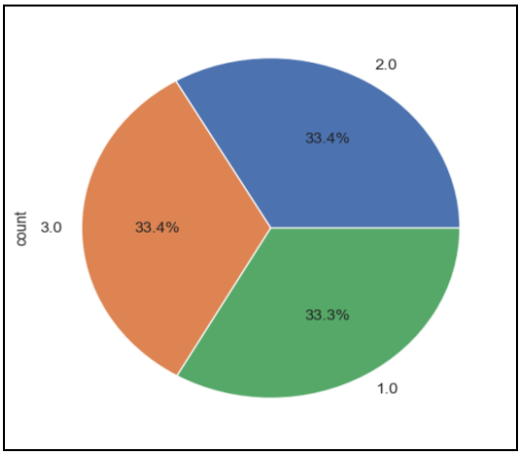

Fig.7 Data Balanced after using SMOTE

Before employing SMOTE, the dataset exhibited an obvious class imbalance as shown in Figure 6. The ‘Normal’ class, representing the larger portion, comprised 77.8% of the data, while the ‘Pathological’ and ‘Suspect’ classes represent a smaller portion, comprising 8.3% and 13.9% respectively. After the implementation of SMOTE the class distribution was balanced, with all the classes accounting for 33.4% of the data each, as illustrated in Figure 7. Achieving a balanced class distribution is essential for improving the performance of the classification algorithm.

5.3 MODELS IMPLEMENTED

Several steps were taken to prepare the dataset that is to be used for training the model and performing feature selection as well as handling class imbalance and then the dataset was split into two sections. The preprocessed data was divided into two sets where 80% was used for training, and 20% used for testing. This division was important for assessing the model’s ability to predict outcomes on new data.

Later, processing of dataset and performing all the tasks, I used StandaredScalar to scale the numerical variables which brought it to the desired range. To get rid of the problem of scale difference each feature should not dominate the model and for this normalization was performed. This helps with an equal chance of influencing the model training hence improving the accuracy of the model.

5.3.1 Experiment 1: Basic Model without Data Balancing or Feature Selection

The first experiment was performed with no feature selection or data balancing with the utmost aim of having a baseline for the following reasons. The hyperparameters for the algorithms were set as follows:

Decision Tree - A sequence of assessments was also done to seek the most appropriate maximum depth for the decision trees. The model was trained and checked using the mentioned dataset with the depth of the tree increasing to different levels at the stages. In foreseeing the real-time scenario of handling the larger database, the accuracy score is analyzed to test the performance of the model constructed at different levels of max_depth value in achieving Goodness of Fit. This strategy was practical in minimizing overfitting and made sure that the degree of complexity of the model was sufficiently high to learn the characteristics of the data. The selection of max_depth is very important as when the model builds the tree, I need to tell how many branches it must be divided, and whereby which it can achieve the highest accuracy. The selected max_depth value for the experiment is 6 as at this node number 6, it has achieved the highest accuracy.

GBM - In this model, for the n_estimators parameter, I have set it to 500. The model builds 300 sequential trees, the number established that is the most effective for the given dataset, not requiring a lot of time. Furthermore, from the mask parameters set in the previous section, I have given the learning_rate as 0. 05. This parameter determines how the individual tree affects the final prediction and helps to avoid overfitting since the model learns in stages. In successive iterations, the same logic was used to ensure some level of consistency, therefore, the random_state parameter is set to 42. Finally, I have set the max_features parameter to 5, which means that the number of features to be considered while searching for the best split has been limited. This increases the usability of the training process and reduces issues of overfitting to a large extent. The selection of these parameters helps in better performance of the model and avoids the chances of overfitting of the model.

KNN – In this model, the value of n_neighbors have been fixed to 1. This was arrived at from the least mean error graph where from the figure with the least number of points we get the highest accuracy for the model. Apart from the neighbors, I have also given that the metric must be Minkowski and the distance ‘p’ is 2, which relates to the use of Euclidian distance.

5.3.2 Experiment 2: Model Built with applying Data Balancing Technique (SMOTE) and Feature Selection (Boruta model)

As for the second experiment to solve the class imbalance problem, the SMOTE model was used and to have the most significant attribute feature, the Boruta model was applied. As for other hyperparameters, all experiment’s defaults were preserved as in the first experiment except the max depth value for the decision tree is changed to be 9. Reducing class imbalance, and selecting relevant features also helps to eliminate or reduce the impact of noisy or irrelevant features.

6. Evaluation

6.1 Basic Model without Data Balancing or Feature Selection

Algorithm	Accuracy	Precision	F1-Score	Recall
Decision Tree	0.93	0.93	0.93	0.93
Random Forest	0.95	0.94	0.94	0.95
KNN	0.92	0.92	0.91	0.92
GBM	0.95	0.95	0.95	0.95

Table 2: Results of Experiment 1

Experiment 1 was carried out as a basis model run on the raw dataset, without using data balancing and feature selection. Results and the value of the experiment are provided and shown in Table 2. Looking at the values of the model provided by the evaluation metrics scores, I have made out that the Random Forest and GBM performed better than the other two algorithms. But as performed with the raw dataset without any balancing the output would be not that perfect as the class which has more class would be classified better than the class variable which has less class.

6.2 Model Built with applying Data Balancing Technique (SMOTE) and Feature Selection (Boruta model)

Algorithm	Accuracy	Precision	F1-Score	Recall
Decision Tree	0.94	0.95	0.95	0.95
Random Forest	0.97	0.97	0.97	0.97
KNN	0.97	0.96	0.97	0.97
GBM	0.98	0.98	0.98	0.98

Table 3: Results of Experiment 2

The second experiment was data balanced by the SMOTE (Synthetic Minority Over-sampling Technique) and feature selected with the Boruta model. This increased in the performance across all the models which is quite evident from these results. The accuracy of the Decision Tree model increased incrementally from 0.93 to 0.94 with these modifications The accuracy went up from 0.95 to 0.97 for the Random Forest and the greatest improvement was realised in the GBM model with an accuracy level of 0.98. Moreover, the precision, F1-score and recall for all models also increased to near 0.97 - 0.98 with some values reaching or touching above that as well. This test confirmed that when we apply the Data Balancing, and Feature Selection Process on top of it we can get a boost in our model performance; Random Forest Model & GBM showed to gain the most from this.

Hence, it is possible to conclude that the second experiment, where the data balancing and the feature selection were incorporated, was more successful in this regard than the first one, where these procedures were neglected. It is also apparent that the implementation of the SMOTE method and the Boruta model considerably increased the performance of all the algorithms, and GBM posted the highest increase. As a result, the stages of the pre-processing process in the machine learning domain are more promising in terms of the accuracy and reliability of the model.

6.3 Discussion

For Experiment 1, I used the original dataset without any balancing of data or feature selection to train the models. It was observed that Random Forest and GBM outperformed all other techniques with an evaluation criterion of 0.95. Moreover, Decision Tree and KNN models achieved a little bit lower accuracy from 0.92. Therefore, we can say that modern ensemble methods like Random Forest and GBM can handle raw data. But there is still some improvement to done especially among the easy methods like Decision Tree and KNN. The data even could be inconsistent or there might have been noise in the features and these things were hurting the performance of many models although some have taken such into account.

The second experiment applied a different approach by applying SMOTE to balance the data and Boruta to choose the most relevant features. These two techniques have not been combined in any of the previous studies to the best of the author's knowledge. All the models' performance was boosted by the preprocessing steps that were conducted. The accuracy of the Decision Tree model rose to 0.94, Random Forest to 0.97, KNN to 0.97, and GBM to an impressive 0.98. This means that if the class imbalance is handled and the most significant features are selected, then the performance of the machine learning models can be enhanced. Through this fresh approach, I was able to exclude the insignificant features while working with the Boruta model, thus lowering the noise and increasing the model's readability. This approach which is considered as the state of art has not been applied in most of the previous research.

Examining the two experiments emphasizes the importance of pre-processing the data in a machine learning process. For the estimation of the risks of foetal health, this means that the medical professions can rely on these improved models to arrive at more accurate and reliable predictions. The new models have the potential of identifying further possible health complications, thus allowing early interventions to be made thus improving the health status of the mother and the fetus.

At last, the integration of data balancing methods such as SMOTE and the novel approach of a variable importance method by Boruta is completely supported in improving model prediction properties. The work shows the need for preprocessing to build accurate predictive models. Future studies may dive deeper into the further improvement of pre-processing techniques and model architectures to improve more on predictability and reliability. These results contribute to the burgeoning literature on machine learning in healthcare, offering applied insights for improving model performance outside of experimental settings and into practical clinical environments.

7. Conclusion and Future Work

The findings of the study suggested that using advanced data preprocessing techniques like SMOTE to balance the dataset and the Boruta method for feature pruning can drastically enhance model performance when it comes to predicting foetal health risks with machine learning. In the experiments, we show that our techniques undeniably improve model accuracy sensitivity on all these scales. Decision Tree, Random Forest, KNN and Boosting models (i.e. GBM) had a positive impact due to these preprocessing steps with the most improvement for the GBM model. This stresses the importance of class imbalance removal and feature selection for getting a good-performing predictive model.

This research is novel because the Boruta feature selection method has not been attempted before by many researchers for fetal health prediction. This feature subset selection helped the models developed in this study to decrease noise and increase interpretability, consequently providing more reliable predictions. This makes a valued impact in health care machine learning especially prenatal where the availability of timely and accurate risk assessment can result in effective intervention.

There are some limitations of this research and some ideas for future work. One limitation is the scope of this dataset, and additional studies using larger datasets in more diverse populations could examine a wider array of fetal health outcomes. In addition, other novel machine learning algorithms and deep learning methods might be better suited for discovering more fetal health features. This extension would provide not just deeper insights into fetal health, but also more refined predictive healthcare. It is due to the endless evolution of machine learning technologies which forms a base for an overall increasing accuracy and validity of this fetus health prediction models.

References

Mennickent, D. *et al.* (2023) ‘Machine learning applied in maternal and fetal health: a narrative review focused on pregnancy diseases and complications’, *Frontiers in Endocrinology*, 14, p.1130139. Available at: https://pubmed.ncbi.nlm.nih.gov/37274341/

Mandala, S.K. (2023) ‘Unveiling the Unborn: Advancing Fetal Health Classification through Machine Learning’, *arXiv preprint*, arXiv:2310.00505. Available at: https://doi.org/10.48550/arXiv.2310.00505

Lee, L.H. *et al.* (2023) ‘Machine learning for accurate estimation of fetal gestational age based on ultrasound images’, *NPJ Digital Medicine*, 6(1), p.36. Available at: https://pubmed.ncbi.nlm.nih.gov/36894653/

Vatambeti, R. and Damera, V.K. (2023) ‘Intelligent diagnosis of obstetric diseases using HGS-AOA based extreme learning machine’, *Acadlore Transactions on AI and Machine Learning*, 2(1), pp.21-32. Available at: https://doi.org/10.56578/ataiml020103

Yin, Y. and Bingi, Y. (2023) ‘Using machine learning to classify human fetal health and analyze feature importance’, *BioMedInformatics*, 3(2), pp.280-298. Available at: https://doi.org/10.3390/biomedinformatics3020019

Singh, K. *et al.* (2023) ‘Efficient Fetal Health Monitoring and Classification with Machine Learning’, *2023 IEEE 7th International Conference on Intelligent Computing and Control Systems (ICICCS)*, pp. 355-359. Available at: https://doi.org/10.1109/ICICCS56967.2023.10142312

Li, J. and Liu, X. (2021) ‘Fetal health classification based on machine learning’, *2021 IEEE 2nd International Conference on Big Data, Artificial Intelligence and Internet of Things Engineering (ICBAIE)*, pp. 899-902. Available at: https://doi.org/10.1109/ICBAIE52039.2021.9389902

Neelakandan, S. *et al.* (2023) ‘Extreme Gradient Boosting based Fetal Health Classification’, *2023 IEEE International Conference on Applied Intelligence and Sustainable Computing (ICAISC)*, pp. 1-6. Available at: https://doi.org/10.1109/ICAISC58445.2023.10200526

Salini, Y. *et al.* (2024) ‘Cardiotocography Data Analysis for Fetal Health Classification Using Machine Learning Models’, vol. 12, pp. 26005-26022. Available at: https://doi.org/10.1109/ACCESS.2024.3364755

Noor, N.F.M. *et al.* (2021) ‘Fetal health classification using supervised learning approach’, *2021 IEEE National Biomedical Engineering Conference (NBEC)*, pp. 36-41. Available at: https://doi.org/10.1109/NBEC53282.2021.9618748

Chandrika, V. and Surendran, S. (2022) ‘Incremental Machine Learning Model for Fetal Health Risk Prediction’, *2022 IEE International Conference on Smart Generation Computing, Communication and Networking (SMART GENCON)*, pp. 1-6. Available at: https://doi.org/10.1109/SMARTGENCON56628.2022.10084232

Parvataneni, K. *et al.* (2023) ‘AI Framework for Fetal Health Risk Prediction’, *2023 IEEE 5th International Conference on Bio-engineering for Smart Technologies (BioSMART)*, pp. 1-8. Available at: https://doi.org/10.1109/BioSMART58455.2023.10162061

Reddy, A.I. and Priya, W.D. (2023) ‘An Efficient Fetal Health Prediction System Using Enhanced XGB Classifier Compared Over K-Nearest Neighbor Classifier with Less Complexity and Improved Accuracy’, *2023 Intelligent Computing and Control for Engineering and Business Systems (ICCEBS)*, pp. 1-5. Available at: https://doi.org/10.1109/ICCEBS58601.2023.10448547

Shifa, H.A. *et al.* (2024) ‘Machine Learning Models for Maternal Health Risk Prediction based on Clinical Data’, *2024 11th International Conference on Computing for Sustainable Global*

Development (INDIACom), pp. 1312-1318. Available at: https://doi.org/10.23919/INDIACom61295.2024.10498822

Rahman, A. and Rabiul Alam, M.G. (2023) ‘Explainable AI based Maternal Health Risk Prediction using Machine Learning and Deep Learning’, *2023 IEEE World AI IoT Congress (AIIoT)*, pp. 0013-0018. Available at: https://doi.org/10.1109/AIIoT58121.2023.10174540

Gill, K. S. *et al.* (2023) ‘Foetal Health Classification Using Input Dataset and Fine Tuning it using K-nearest Neighbour, Naïve Bayes and Decision Tree Classifier’, *2023 First International Conference on Advances in Electrical, Electronics and Computational Intelligence (ICAEECI)*, pp. 1-5. Available at: https://doi.org/10.1109/ICAEECI58247.2023.10370853

Gopakumar, D. *et al.* (2023) ‘An Artificial Intelligence Approach to Fetal Health Risk Prediction’, *2023 IEEE Integrated STEM Education Conference (ISEC)*, pp. 206-207. Available at: https://doi.org/10.1109/ISEC57711.2023.10402221

Shetty, S. P. *et al.* (2024) ‘FetalCare: Enhancing Prenatal Care through a Machine Learning Approach for Fetal Health Classification and Birth Weight Prediction’, *2024 3rd International Conference on Applied Artificial Intelligence and Computing (ICAAIC)*, pp. 700-704. Available at: https://doi.org/10.1109/ICAAIC60222.2024.10575277

Marandi, A. K. and Shah, H. (2022) ‘A Convolution Neural Network-based Prediction of Fetal Health’, *2022 11th International Conference on System Modeling & Advancement in Research Trends (SMART)*, pp. 1369-1374. Available at: https://doi.org/10.1109/SMART55829.2022.10047304

Dang, T. H. *et al.* (2023) ‘Improve Classification Quality of Fetal Status from Cardiotocogram Data by Using Machine Learning’, *2023 1st International Conference on Health Science and Technology (ICHST)*, pp. 1-6. Available at: https://doi.org/10.1109/ICHST59286.2023.10565357

Saini, P. *et al.* (2023) ‘Analysis of High-Risk Pregnancies in India Using Artificial Intelligence’, *2023 2nd International Conference on Ambient Intelligence in Health Care (ICAIHC)*, pp. 01-06. Available at: https://doi.org/10.1109/ICAIHC59020.2023.10431479

Dang, T. H. *et al.* (2023) ‘Classification of Fetal Status from Cardiotocogram Data by Using Machine Learning’, 2023 12th International Conference on Control, Automation and Information Sciences (ICCAIS), pp. 352-357. Available at: https://doi.org/10.1109/ICCAIS59597.2023.10382381