Integrated PreNATAL Care: *Mitigating Risks and Improving Maternal-Child Health Outcomes*

Aamleen Ahmed
aamleen20002@iiitd.ac.in

Sufyan Azam
sufyan20312@iiitd.ac.in

Mohammad Shariq
shariq20220@iiitd.ac.in

Indraprastha Institute of Information Technology, Delhi

Abstract

The ***Integrated PreNATAL Care*** *Model employs Machine Learning to predict maternal and fetal health risks, empowering healthcare teams to make proactive decisions. By leveraging diverse pre-natal data sources, the project aims to reduce stillbirths and neonatal complications through informed interventions.*

1. Introduction

The journey to parenthood is an extraordinary passage marked by joy and challenges. However, stillbirths, neonatal deaths, and preventable complications overshadow this miraculous process. Our project is driven by an unwavering motivation to address this heart-wrenching issue. By leveraging the power of Machine Learning and analyzing a wealth of pre-natal data sources, we aim to provide healthcare professionals with a tool to predict potential health risks for mothers and fetuses. With these insights, medical teams can make informed decisions, enabling timely interventions and a more coordinated approach to maternal and child health. Through the fusion of data science and compassionate care, we envision a future where this integrated model significantly reduces the occurrence of tragic outcomes, paving the way for healthier beginnings and brighter tomorrows.

2. Literature Review

1. **A Machine Learning Approach for the Prediction of Fetal Health using CTG [5]**, *Pradhan A.K., et.al*: The authors discuss machine learning techniques for the prediction of fetal health based on cardiotocography (CTG) data. CTG is a method employed for monitoring fetal heart rate and uterine contractions during pregnancy. Specific focus on its application is in classifying fetal heart rate signals and diagnosing medical conditions. This paper uses classifiers such as Logistic Regression, K-Nearest Neighbors (KNN), Random Forest, and Gradient Boosting Machine (GBM) for this task. Model performance evaluation is conducted using metrics such as accuracy, precision, recall, and F1-score. The Random Forest model emerges as the most effective, achieving the highest accuracy rate of 0.99, particularly in the context of predicting CTG class labels.

   In terms of future research directions, the paper underscores the potential for further enhancement in predictive model performance through the incorporation of additional preprocessing steps, such as linear discriminant analysis or principal component analysis. The paper firmly underscores the clinical significance of early detection and timely intervention within the domain of prenatal care. It posits that machine learning models, with particular emphasis on Random Forest, offer promising prospects for more precise and objective assessments of fetal well-being, thereby contributing to improved healthcare outcomes.

2. **Machine Learning Approaches for Early Diagnosis and Prediction of Fetal Abnormalities**[2] *R. Chinnaiyan and Stalin Alex*: In this paper, the authors explored and reviewed machine learning approaches for early detection of prenatal anomalies. They discuss the importance of early diagnosis of fetal anomalies, particularly in the first trimester, and explore the significance of applying these models in improving accuracy for ultrasound fetal imaging. This paper emphasizes on fetal abnormalities during the first trimester of pregnancy, aiming to examine machine learning mechanisms for improved diagnosis and prognosis of abdominal abnormalities, with the goal of lowering their occurrence rates. It involves four key steps: segmentation, image enhancement, feature extraction, and image classification. It also confers that the best model for our classification task is a Neuro Fuzzy Based Genetic Algorithm and the accuracy achieved through it was around 0.98.

3. **Evaluation of support vector machines and random forest classifiers in a real-time fetal monitoring system based on cardiotocography data**[4] *Vinayaka Nagendra et.al.*: In this paper, evaluation techniques for the foetal state prediction based on Cardiotocography (CTG) data are compared. The authors extract other features that potentially offer more details about the foetal status and assess the performance of these predictions in a real-time clinical decision support system. This study differs from earlier research in that it takes into ac-

count all three foetal states (normal, suspicious, and abnormal). The effectiveness of Support Vector Machines (SVM) and Random Forests (RF) in forecasting foetal outcomes is compared by the authors. SVM performed marginally better for suspicious instances, although both SVM and RF had above 96% accuracy in predicting foetal outcomes.

The significance of machine learning in aiding obstetricians in questionable circumstances is also discussed in the paper. The importance of CTG data in forecasting the condition of the foetus during labour to show the danger of foetal acidosis (low blood pH from low oxygen levels) is also highlighted by the authors.

4. **Machine learning to predict pregnancy outcomes: a systematic review, synthesizing framework and future research agenda, *Muhammad Nazrul Islam et. al.*:**[3] This study reviews 26 articles from 2000-2020 that focuses on the current research and development perspectives utilizing ML techniques for predicting optimal childbirth modes and detecting complications. It explores algorithms, features, data sources, and their performance in pregnancy outcomes. It also highlights future research opportunities for reducing maternal complications and mortality rates, including unsupervised and deep learning algorithms, ML-based clinical decision support systems, dataset enhancement, and surgical robotic tools.

3. DataSet[1]

3.1. Description

The *dataset* comprises of 2126 records of human fetal heart rate and mother's uterine contraction characteristics obtained using a cardiotocogram device which helps doctors trace the heart rate of the fetus using a technique called cardiotocography. The characteristics comprises of various statistics of the heart rate and contraction data, such as the baseline, accelerations, decelerations, variability, histogram parameters, and tendency. For each record in the dataset, there are two additional class labels: one for the fetal state (normal, suspect, or pathological), and the other for the fetal pattern (A, B, C, D, E, AD, DE, LD, FS, or SUSP). Based on the cardiotocographic properties of the dataset, the health condition of the fetus can be predicted. Currently, we are predicting the fetal state using the three classes (normal, suspect and pathological) and later on we'll work on the fetal pattern classification problem to create a comprehensive model for fetal health prediction and classification.

3.2. Pre-Processing Steps

3.2.1 Column Headers, Null Values and Outliers

The dataset was analyzed for any missing or null values present, and the records found with any such value were dropped. After that, the column headers were renamed for the dataset for easier interpretation purposes.

Next, we analyzed the outliers to remove them for an accurate model training. For outlier removal, there are two main techniques: ***elimination*** and ***substitution***. Elimination is the process of deleting the records that have outlier values, while substitution replaces the outliers with more reasonable values. In this project, we used both techniques to compare their effects on the data quality and the machine learning model performance. For visualising the outliers we made a boxplot of the dataset which gave us a fair idea of the number of outliers present in each column. We then tested the following outlier removal techniques-

- **Z-score normalization:** This method converts the data to normal distribution and replaces values having extreme Z-scores with appropriate values.
- **KNN imputer:** This method replaces outliers with the mean of their nearest neighbors.
- **IQR:** This method finds the difference between the quartiles and removes values outside a certain range.
- **Robust covariance:** This method uses a robust estimator for the covariance matrix and removes values above a threshold.

Out of these the Z-score normalisation method was used as it removed most of the outliers accurate to the real-world data.

3.2.2 Feature Engineering

In our feature engineering process, we made strategic decisions to enhance model efficiency and interpretability. We identified certain columns that exhibited high correlation among each other, indicating redundancy in our feature set. Consequently, we removed these redundant columns, recognizing that a single representative feature could effectively capture the information conveyed by the entire group.

Furthermore, we also identified columns with extremely low correlation with the output data, suggesting minimal relevance in predicting fetal health and prenatal complications. In the interest of simplifying our models and improving their performance, we opted to remove these less informative columns. This streamlined feature set not only reduced dimensionality but also contributed to more focused and accurate predictive models.

3.2.3 Output Class Imbalance Problem

Output class imbalance arises when classes in the output label are unevenly distributed, leading to bias in machine learning algorithms towards the dominant class. Examining the pie chart (Fig 1) for fetal states (normal, suspect pathological) reveals a significant imbalance, heavily favoring the normal class. To tackle this, resampling techniques like oversampling and undersampling were applied.

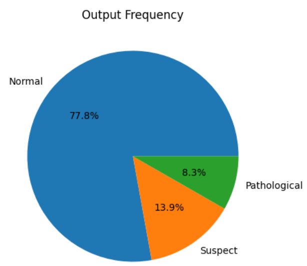

Figure 1. Fetal State Classes

3.2.4 UnderSampling

Undersampling involves randomly removing data from the majority class to achieve a more balanced distribution. The near miss algorithm selectively removes samples from the majority class, enhancing class separation. However, this led to a significant reduction in total records to 1000, risking underfitting during ML model training.

3.2.5 OverSampling

Oversampling aims to balance the dataset by adding data to the minority class through duplication or synthetic record generation. SMOTE, employed here, creates synthetic samples by randomly selecting a minority instance and its nearest neighbors, mitigating majority class bias and boosting minority class records.

3.3. Data Visualisation

Exploratory Data analysis and visualization was done on the preprocessed dataset from the above methods and scatter plots, distribution plots, pie chart, boxplots, pairplots, heatmap etc were drawn to get a better understanding of the each attribute present in the dataset. From visualising data, following insights were obtained: The baseline value attributes data, and tendency attribute can be approximated to a normal distribution. Fetal state is imbalanced with more normal cases and fetal pattern contains majority of the A cases. The accelerations, fetal movement, decelerations, and abnormal-long-term-variable attributes have highly skewed data. Attributes such as calm-sleep, REM-sleep, active-vigilance etc contain binary classes which is evident from scatter plots. For pathological class the REM-sleep, vigilance, and decelerative pattern have values that belong to the same negative class.

Using these insights we can see the relation between attributes in a more effective manner and possess a better understanding of the dataset for building our ML model.

4. Methodology

4.1. Preparing Train-Test Data

In our study focused on predicting fetal health and prenatal complications, we implemented a robust evaluation methodology to ensure the credibility of our machine learning models. We adopted three different data splitting ratios (80:20, 70:30, and 75:25) while carefully maintaining class distribution balance in each subset.

To enhance the reliability of our results, we conducted a 10-fold stratified cross-validation, to maintain the class-distribution. This approach involved partitioning the dataset into ten subsets, preserving class distribution, and iterating through each fold. During each fold, nine subsets were used for training, and the remaining subset was employed for testing. This comprehensive process allowed us to assess the models' generalization performance, stability, and effectiveness.

By combining these data splitting techniques and cross-validation, our research methodology ensures a thorough and credible evaluation of our machine learning models' ability to predict fetal health and prenatal complications. This approach establishes the foundation for reliable and robust predictive models.

4.1.1 Principal Component Analysis (PCA):

To mitigate overfitting in our 2000-record dataset, we employed Principal Component Analysis (PCA) for dimensionality reduction. Experimentation revealed that projecting features onto approximately 5 to 10 principal components effectively reduced data dimensionality, enhancing the generalization capability of our models.

4.1.2 Regularization (L2 and L1):

Addressing overfitting, we implemented L2 (Ridge) and L1 (Lasso) regularization with the "liblinear" solver. L2 regularization encouraged smaller non-zero feature values, preventing extreme weights. Simultaneously, L1 regularization emphasized feature selection by driving some coefficients to zero, reducing the impact of less relevant features. This dual regularization approach significantly enhanced model robustness.

5. Models

In our study, we implemented two classification models to address distinct facets of prenatal healthcare—one focused on predicting fetal states (normal, suspect, pathological), and the other dedicated to classifying morphologic patterns (A, B, C, ...).

5.1. Foetal Risk Classification *3-Class Classification*

In the prenatal healthcare, we addressed a 3-class classification challenge to predict fetal risk states (normal, suspect, pathological) based on the mentioned dataset. Employing machine learning models such as SVM, XGBoost, and ANN, we aimed to optimize predictive accuracy. Hyperparameter tuning, including GridSearchCV and Keras

Tuner, was performed to enhance model performance, contributing to proactive risk assessment and improved outcomes for both mothers and babies during childbirth.

1. **Logistic Regression (F1 Score: 0.72):**
Logistic Regression provides a decent F1 score of 0.72 for our 3-class classification model. However, given the complexity of our prenatal health dataset with 40 features, its simplicity may limit its ability to capture intricate relationships effectively.

2. **K-Nearest Neighbors (KNN) (F1 Score: 0.82):**
KNN performs well with an F1 score of 0.82, suggesting its effectiveness in capturing patterns. However, the dimensionality of our dataset with 40 features might impact KNN's performance, given its sensitivity to high-dimensional data.

3. **Support Vector Machines (SVM) (F1 Score: 0.87):**
SVM stands out with a high F1 score of 0.87, indicating strong performance in distinguishing fetal states. Its ability to handle complex relationships makes it a suitable choice, although careful tuning may be necessary due to the dataset's size.

4. **Naive-Bayes (F1 Score: 0.66):**
Naive-Bayes, with an F1 score of 0.66, suggests limitations in capturing dependencies among the 40 features in our prenatal health dataset. Its assumption of feature independence may not align well with the complex relationships inherent in health parameters.

5. **Decision Trees (Overfitting):**
Decision Trees, prone to overfitting, might struggle with our dataset. The overfitting issue could stem from the 40 features, leading to a model that performs well on training data but poorly on unseen data.

6. **Random Forests (F1 Score: 0.85):**
Random Forests offer a robust solution, mitigating overfitting and achieving a high F1 score of 0.85. This ensemble method leverages multiple decision trees, proving effective in capturing the complexities of prenatal health features.

7. **XGBoost (F1 Score: 0.79):**
XGBoost, with an F1 score of 0.79, demonstrates solid performance in our prenatal health context. Its ability to handle complex relationships makes it a valuable choice, though it might require careful parameter tuning to optimize performance.

8. **Artificial Neural Network (ANN) (F1 Score: 0.92):**
ANN shines with the highest F1 score of 0.92, indicating its capability to learn intricate patterns in our prenatal health dataset. Given the richness of our data, the deep learning architecture of ANN proves effective, but it might demand more computational resources and data for optimal training.

5.2. Morphological Predictions *10-Class Classification*

Expanding our focus, we addressed a 10-class classification challenge in prenatal healthcare, aiming to predict morphologic patterns (A, B, ...) using the same dataset. This classification provides a detailed understanding of diverse fetal morphologies, enabling more nuanced health assessments.

1. **Logistic Regression (LR) (F1 Score: 0.68):**
LR was applied to predict morphologic patterns, yielding an F1 score of 0.68. This model provides a foundational understanding of the data, though its performance may be further improved.

2. **K-Nearest Neighbors (KNN) (F1 Score: 0.78):**
KNN demonstrated a solid F1 score of 0.78, showcasing its effectiveness in capturing complex relationships within the dataset.

3. **Decision Trees (Overfitting):**
Decision Trees, while prone to overfitting, were utilized to understand morphologic patterns. However, their performance may be limited due to the complexity of the task.

4. **Random Forests (F1 Score: 0.8):**
Random Forests, an ensemble of decision trees, achieved an F1 score of 0.8, proving effective in capturing diverse morphologic patterns.

5. **XGBoost & CatBoost (F1 Score: 0.8):**
XGBoost, while demonstrating solid performance with an F1 score around 0.76, led us to explore CatBoost—an algorithm designed to handle categorical features more effectively. Given the categorical nature of our health parameters, CatBoost emerged as a better fit for our problem. Subsequently, its incorporation resulted in an improved F1 score (0.8 +), showcasing its enhanced suitability for our prenatal health dataset. classification problem.

6. **Artificial Neural Network (ANN) (F1 Score: 0.88):**
ANN, with the highest F1 score of 0.88, proved to be a powerful model for predicting diverse morphologic patterns. The Keras Tuner was used to optimize its architecture.

5.3. Hyperparameter Tuning

In addition to model evaluation, we performed hyperparameter tuning using GridSearchCV for models like SVM, XGBoost, etc., to obtain the best configurations for improved performance.

For the Artificial Neural Network (ANN), we utilized Keras Tuner to determine the optimal layer size and number, striking a balance between computational complexity and achieving a high F1 score in our 3-class classification model.

6. Refinement Strategies and Overcoming Challenges

Our research project encountered significant challenges that required strategic interventions to ensure the reliability of our predictive models for fetal health and prenatal complications. Key challenges included initial overfitting, where models demonstrated excessively high accuracy scores of around 99%, even when the training dataset was reduced to 50

To overcome these challenges and enhance model robustness, we employed a focused approach:

- **Dimensionality Reduction (PCA):** Implementing Principal Component Analysis (PCA) addressed overfitting by reducing the number of features while preserving essential information. This mitigated overfitting tendencies, promoting more generalized model performance.

- **Regularization Techniques:** The application of L2 (Ridge) and L1 (Lasso) regularization techniques with the "liblinear" solver introduced penalty terms to model cost functions. This facilitated feature selection and smaller feature coefficients, contributing to reduced overfitting.

- **Cross-Validation Assurance:** Utilizing 10-fold stratified cross-validation ensured stable and robust model performance, assessing their ability to generalize to unseen data while preventing overfitting issues.

- **Resampling Strategies:** Addressing class imbalance involved both oversampling and undersampling. Oversampling increased the representation of the minority class, while undersampling reduced the majority class instances. These strategies balanced class distribution, enhancing model performance.

- **Outlier Imputation Measures:** Minimizing the influence of outliers on model training, we implemented appropriate imputation techniques to handle outlier values in the data effectively.

- **Incorporation of Advanced Algorithms:** In addressing overfitting challenges, we explored and incorporated advanced algorithms such as CatBoost, particularly beneficial for handling categorical features. This resulted in improved F1 scores, showcasing enhanced model suitability for our problem.

The combination of these refined strategies not only mitigated challenges but also contributed to the overall improvement of our predictive models for fetal health and prenatal complications. This comprehensive and adaptive approach strengthened the credibility of our research findings and furthered the models' generalization capabilities.

7. Results and Analysis

7.1. Combined Observations

In our comprehensive prenatal healthcare project encompassing both 3-class and 10-class classification challenges, we draw the following observations:

1. ***Common Model Strengths:*** Across both models, Random Forests demonstrated robustness, while Support Vector Machines (SVM) showcased strength in distinguishing fetal states.

2. ***Deep Learning Dominance:*** Artificial Neural Network (ANN) consistently outperformed other models, affirming its dominance in capturing intricate health patterns.

3. ***Hyperparameter Tuning Consistency:*** The process of hyperparameter tuning, involving techniques such as GridSearchCV for traditional models and Keras Auto Tuner for the Artificial Neural Network (ANN), played a pivotal role in optimizing model performance.
   Notably, the tuned parameters exhibited a consistent pattern across both the 3-class and 10-class models, underscoring the robustness of certain configurations in capturing the intricacies of prenatal health data.

7.2. Comparative Observations

Comparing the results between the 3-class and 10-class classification models yields insightful observations:

1. ***Data Scarcity Impact:*** The 10-class model, with limited records per class, faced more overfitting issues compared to the 3-class model with relatively abundant class-wise distributed data.

2. ***Impact on Regularization:*** The 10-class model necessitated stricter regularization measures due to data scarcity and increased complexity, while the 3-class model had more flexibility.

3. ***Utilization of PCA:*** PCA played a more significant role in the 10-class model, acting as a crucial tool to combat overfitting and reduce dimensionality.

4. ***Clustering Challenges:*** Clustering results were more distinct in the 3-class model, while the 10-class model presented additional challenges due to the complexity introduced by multiple classes.

7.3. Combined Inference

Our combined inference across both classification challenges culminates in the following insights:

- ***Model Adaptability:*** The resilience of Random Forests and SVM across both models underscores their adaptability to diverse prenatal health challenges.

- ***Deep Learning Significance:*** The consistent dominance of ANN highlights the significance of deep learning in capturing intricate health patterns, even in scenarios with limited data per class.

- ***Impact of Data Characteristics:*** The observed differences in regularization, PCA usage, and clustering emphasize the profound impact of data distribution and scarcity on model performance.

- ***Optimized Configurations:*** The consistent patterns observed in tuned hyperparameters across both classification challenges suggest that certain configurations are universally effective in our prenatal healthcare context. This implies that fine-tuning parameters can lead to optimized model configurations, contributing to improved predictive accuracy and generalization.

- ***Transferability of Insights:*** The insights gained from addressing both the 3-class and 10-class classification challenges can be leveraged to enhance the broader understanding of prenatal healthcare. Transferability of findings across diverse models and data scenarios fosters a more comprehensive approach to predictive modeling in this critical domain.

- ***Continuous Contribution:*** Leveraging these insights, our project strives to continuously contribute to a more comprehensive understanding of fetal health states and patterns, facilitating proactive healthcare interventions for improved maternal and neonatal outcomes.

7.4. Discussion of Implications

The enhanced performance and robustness of our models have significant implications for fetal health monitoring and prenatal complication prediction. These models can contribute to more accurate and timely diagnoses, potentially improving patient outcomes and healthcare resource allocation.

7.5. Limitations

While our approach demonstrated promising results, it is essential to acknowledge the limitations of our study. These limitations include potential biases in the dataset, the choice of hyperparameters for PCA and regularization, and the specific challenges of real-world healthcare applications.

8. Conclusion

In summary, our prenatal healthcare project stands as a significant stride in advancing predictive modeling for maternal and fetal health. By strategically employing diverse machine learning models and embracing advanced algorithms like CatBoost, we've elevated our predictive accuracy. The integration of artificial neural networks (ANNs) has proven pivotal, with our ANN model emerging as a potent force in capturing intricate patterns within our prenatal health dataset. Techniques such as PCA, regularization, and cross-validation have further solidified the reliability of our models.

Looking ahead, our dual-model approach—predicting fetal states and classifying morphological patterns—promises a holistic understanding of fetal health. This comprehensive strategy not only enriches diagnostic accuracy but also positions our work at the forefront of innovative solutions in prenatal healthcare. Our project not only contributes to the refinement of predictive capabilities but also holds the potential to empower clinicians with more accurate insights for proactive interventions, ultimately improving outcomes for both mothers and newborns.

References

[1] D. Campos and J. Bernardes. Cardiotocography. UCI Machine Learning Repository, 2010. DOI: https://doi.org/10.24432/C51S4N.

[2] R. Chinnaiyan and Stalin Alex. Machine learning approaches for early diagnosis and prediction of fetal abnormalities. In *2021 International Conference on Computer Communication and Informatics (ICCCI)*, pages 1–3, 2021.

[3] Mustafina S.N. Mahmud T. et al Islam, M.N. Machine learning to predict pregnancy outcomes: a systematic review, synthesizing framework and future research agenda. *BMC Pregnancy Childbirth 22, 348 (2022)*, 2022.

[4] Vinayaka Nagendra, Harikishan Gude, Divya Sampath, Steven Corns, and Suzanna Long. Evaluation of support vector machines and random forest classifiers in a real-time fetal monitoring system based on cardiotocography data. In *2017 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB)*, pages 1–6, 2017.

[5] Astik Kumar Pradhan, Jitendra Kumar Rout, Aurobinda Bharat Maharana, Bunil Kumar Balabantaray, and Niranjan Kumar Ray. A machine learning approach for the prediction of fetal health using ctg. In *2021 19th OITS International Conference on Information Technology (OCIT)*, pages 239–244, 2021.

Appendix

- **Project Code:** Click here to access the code for our project on Google Colab.

- **Presentation Slides:** Click here to refer to the presentation slides for a detailed overview of our project.

- **Dataset:** Click Here to access the dataset used in our project.