Journal of Advanced Analytics in Healthcare Management

Predicting Fetal Health using Cardiotocograms: A Machine Learning Approach

Rohit R Dixit
The University of Texas at Tyler

Submitted: 13 June 2021
Accepted after revision: 30 November 2021
Published online: 22 January 2022

Abstract

Cardiotocograms (CTGs) are widely used in assessing fetal health and preventing child and maternal mortality. This study aims to develop machine learning models for predicting fetal health based on features extracted from CTG exams. The dataset used contains 2126 records, classified by expert obstetricians into three classes: Normal, Suspect, and Pathological. To address the challenges associated with feature selection and class imbalance, univariate feature selection (SelectKBest) and SMOTE RandomOverSampler techniques were employed. Various machine learning algorithms including Logistic Regression, k-Nearest Neighbors (kNN), Support Vector Machines (SVM), Decision Tree, Extra Trees, Random Forest, Gradient Boosting, and Neural Network Multi-Layer Perceptron (NN MLP) were evaluated. The results demonstrate promising performance across multiple metrics. The Random Forest model achieved the highest accuracy (95.77%), recall (95.77%), precision (95.87%), F1-score (95.81%), and MCC score (88.58%). The Extra Trees, Gradient Boosting, and NN MLP models also exhibited strong performance. Furthermore, key factors contributing to the prediction of fetal well-being were identified. These included the rate of accelerations, abnormal short-term and long-term variability of fetal heart rate, as well as the characteristics of the histogram constructed from fetal heart rate values (width, mean, and variance). Additionally, the trend or histogram tendency over time emerged as a significant predictor, capturing changes and patterns associated with potential complications.

Keywords: *Cardiotocograms (CTGs), Fetal health, Machine learning, Feature selection, Prediction*

Introduction

Fetal mortality, although a significant public health concern, is often overlooked in discussions surrounding global health issues. The loss of a fetus before birth not only devastates families emotionally, but it also has far-reaching implications for maternal health and the overall well-being of communities [1]. The consequences of fetal mortality extend beyond the immediate loss, affecting societies at both individual and societal levels.

Firstly, fetal mortality has a profound impact on families and communities. The loss of a fetus is a traumatic event for expectant parents, who had eagerly anticipated the arrival of their child. The emotional toll of such a loss can be long-lasting, leading to feelings of grief, depression, and a sense of emptiness. Families may struggle to cope with the loss and require support from healthcare providers and social networks [2]. Furthermore, the social fabric of communities can be disrupted as families grapple with the aftermath of fetal mortality, leading to a decrease in social cohesion and well-being.

Secondly, fetal mortality is closely intertwined with maternal health. The risk factors associated with fetal mortality often overlap with those affecting maternal health, including inadequate access to prenatal care, chronic health conditions, substance abuse, and socioeconomic disparities. Addressing fetal mortality requires a comprehensive approach that prioritizes maternal well-being, ensuring early and regular prenatal care, promoting healthy behaviors during pregnancy, and providing support systems for expectant mothers. By focusing on improving maternal health, we can significantly reduce the risk of fetal mortality and improve overall community health outcomes [3].

Lastly, the impact of fetal mortality extends beyond individual families and maternal health to broader societal consequences. High rates of fetal mortality are indicative of deeper health system challenges, including limited access to healthcare services, insufficient infrastructure, and inadequate training of healthcare professionals [4]–[7].

Cardiotocograms (CTGs) have emerged as an invaluable tool in the field of obstetrics, playing a crucial role in assessing fetal health and ensuring the well-being of both the child and the mother.

These devices measure various parameters, including fetal heart rate, fetal movement, and uterine contractions, to ensure the well-being of both the unborn child and the mother. CTGs are a simple and cost-accessible option that has greatly improved the quality of prenatal care and reduced the risks associated with childbirth [8].

Fetal heart rate monitoring is one of the key aspects of CTGs. By continuously tracking the heart rate, healthcare professionals can identify any abnormalities or irregularities that may indicate distress or compromised oxygen supply to the fetus. This information allows for timely intervention and necessary medical measures to be taken to safeguard the health of the unborn baby. Additionally, CTGs also monitor fetal movement, providing valuable insights into the overall activity and vitality of the fetus. Changes in movement patterns can be indicative of fetal well-being, and any deviations from the normal range can prompt immediate medical attention.

Furthermore, CTGs enable the assessment of uterine contractions during labor. By measuring the frequency, duration, and strength of contractions, healthcare providers can evaluate the progress of labor and identify any potential complications. This information helps in determining the need for interventions such as labor augmentation or the administration of pain relief. Timely interventions based on CTG readings can prevent adverse outcomes such as fetal distress, birth asphyxia, or maternal complications, ultimately reducing the risk of child and maternal mortality.

The simplicity and cost-accessibility of CTGs have made them an indispensable tool in the field of obstetrics. Compared to more invasive procedures or sophisticated imaging techniques, CTGs offer a non-invasive and user-friendly approach to monitoring fetal health. They can be easily employed in various healthcare settings, including clinics, hospitals, and even remote or low-resource areas where advanced medical facilities may be limited.

The utilization of CTG data for automated assessment of fetal health represents a major advancement in prenatal care. Traditionally, the interpretation of CTG exams has relied heavily on the expertise of obstetricians, which can be subjective and prone to human error. However, this study indicates that objective and reliable assessment of fetal health can be achieved through automated algorithms. By analyzing patterns in the CTG data and comparing them to the outcomes determined by expert obstetricians, the studies were able to develop algorithms that accurately classified fetal health status.

Machine learning (ML) has emerged as a transformative technology in the realm of healthcare, offering immense potential to revolutionize various aspects of the field. ML algorithms, powered by advanced computational models, have the ability to analyze vast amounts of medical data, uncover patterns, and generate valuable insights that can enhance diagnosis, treatment, and patient care. The integration of ML in healthcare facilitates the development of predictive models that aid in early detection and risk assessment of diseases. By leveraging techniques such as deep learning, neural networks, and natural language processing [9], ML algorithms can effectively process complex medical images, electronic health records (EHRs), and clinical notes, enabling clinicians to make more accurate and informed decisions [10], [11].

Moreover, ML has demonstrated remarkable proficiency in medical imaging analysis, a critical component of disease diagnosis and treatment planning [12]–[14]. Through the utilization of convolutional neural networks (CNNs) and image recognition algorithms, ML systems can detect abnormalities in medical images with a high level of accuracy, often outperforming human experts [15]. This technology holds significant potential for improving radiology and pathology practices, allowing for faster and more precise diagnoses of conditions such as cancer, cardiovascular diseases, and neurological disorders [16]–[18]. ML-based predictive models can facilitate the identification of patients at higher risk of adverse events, enabling proactive interventions and personalized treatment plans.

Methods

The dataset contains information related to cardiotocograms (CTGs) and various fetal health indicators. It consists of 2,130 instances with 21 attributes, providing a comprehensive set of features for analyzing and predicting fetal health [19].

Each instance in the dataset represents a fetal cardiotocogram examination, which is a non-invasive method used to monitor fetal well-being during pregnancy [19]. The attributes include important measurements such as baseline fetal heart rate, accelerations, decelerations, and various indices derived from the CTG signals. Additionally, the dataset includes the corresponding fetal health classification, with three classes indicating the presence of normal, suspicious, or pathological fetal conditions.

Univariate feature selection is a technique used to identify and select the most relevant features from a dataset based on their individual statistical significance. This approach operates by subjecting each feature to independent statistical tests and then ranking them according to their scores. It is commonly employed as a preliminary step before applying an estimator or predictive model to the data. One of the specific routines available in scikit-learn is *SelectKBest*. As the name suggests, this method aims to retain only the K highest-scoring features from the dataset, discarding the rest. In other words, it eliminates all but the K most relevant features according to their respective scores obtained from the univariate statistical tests. By specifying the desired value of K, the user can control the number of features to be selected. Following *SelectKBest*, the following features were dropped: uterine_contractions, light_decelerations, prolongued_decelerations, accelerations, fetal_movement, histogram_number_of_zeroes, histogram_max, histogram_number_of_peaks, and mean_value_of_short_term_variability.

Table 1. Variables

Variables	Description
baseline_value	Baseline Fetal Heart Rate (FHR) (beats per minute)
accelerations	Number of accelerations per second

fetal_movement	Number of fetal movements per second
uterine_contractions	Number of uterine contractions per second
light_decelerations	Number of light decelerations (LDs) per second
severe_decelerations	Number of severe decelerations (SDs) per second
prolongued_decelerations	Number of prolonged decelerations (PDs) per second
abnormal_short_term_variability	Percentage of time with abnormal short-term variability
mean_value_of_short_term_variability	Mean value of short-term variability
percentage_of_time_with_abnormal_long_term_variability	Percentage of time with abnormal long-term variability
mean_value_of_long_term_variability	Mean value of long-term variability
histogram_width	Width of histogram made using all values from a record
histogram_min	Histogram minimum value
histogram_max	Histogram maximum value
histogram_number_of_peaks	Number of peaks in the exam histogram
histogram_number_of_zeroes	Number of zeros in the exam histogram
histogram_mode	Histogram mode
histogram_mean	Histogram mean
histogram_median	Histogram median
histogram_variance	Histogram variance
histogram_tendency	Histogram tendency
fetal_health	Encoded as 1-Normal; 2-Suspect; 3-Pathological. (Target column)

In the given dataset, the target class of fetal health exhibits a significant imbalance. The distribution of fetal health outcomes is highly skewed, with the majority of instances belonging to the class labeled as 1.00, indicating Normal fetal health. This class accounts for the highest frequency of observations within the dataset.

The second most frequent class observed in the dataset is labeled as 2.00, representing Suspect fetal health. While not as prevalent as the Normal class, the Suspect class still demonstrates a considerable presence within the dataset.

On the other hand, the class with the least frequency in this dataset is labeled as 3.00, which corresponds to Pathological fetal health. Instances belonging to this class are significantly fewer compared to the other two classes.

To address the issue of class imbalance, the RandomOverSampler technique was employed. RandomOverSampler is a resampling method commonly used in machine learning to address class imbalance [20], [21]. It works by randomly replicating instances from the minority classes, in this case, the Suspect and Pathological classes, until a balanced distribution is achieved. By oversampling the minority classes, RandomOverSampler helps to alleviate the bias towards the majority class and create a more balanced representation of fetal health outcomes in the dataset.

Among the various features in the dataset, the one that exhibits the strongest correlation with fetal health is prolonged decelerations. This feature demonstrates a correlation coefficient of 0.485, indicating a moderately positive relationship with fetal health. A higher value of prolonged decelerations tends to be associated with a higher likelihood of abnormal fetal health.

Additionally, moderate correlations are observed between fetal health and two other features. The first is abnormal short-term variability, which shows a moderate correlation with fetal health. An increase in abnormal short-term variability is moderately associated with a higher probability of fetal health issues.

The second feature demonstrating a moderate correlation with fetal health is the percentage of time with abnormal long-term variability. This feature indicates the proportion of time during which the fetal heart rate exhibits abnormal long-term variability. A higher percentage of time with abnormal long-term variability shows a moderate positive correlation with the presence of fetal health problems.

Figure 1. correlation and linear relationships between variables

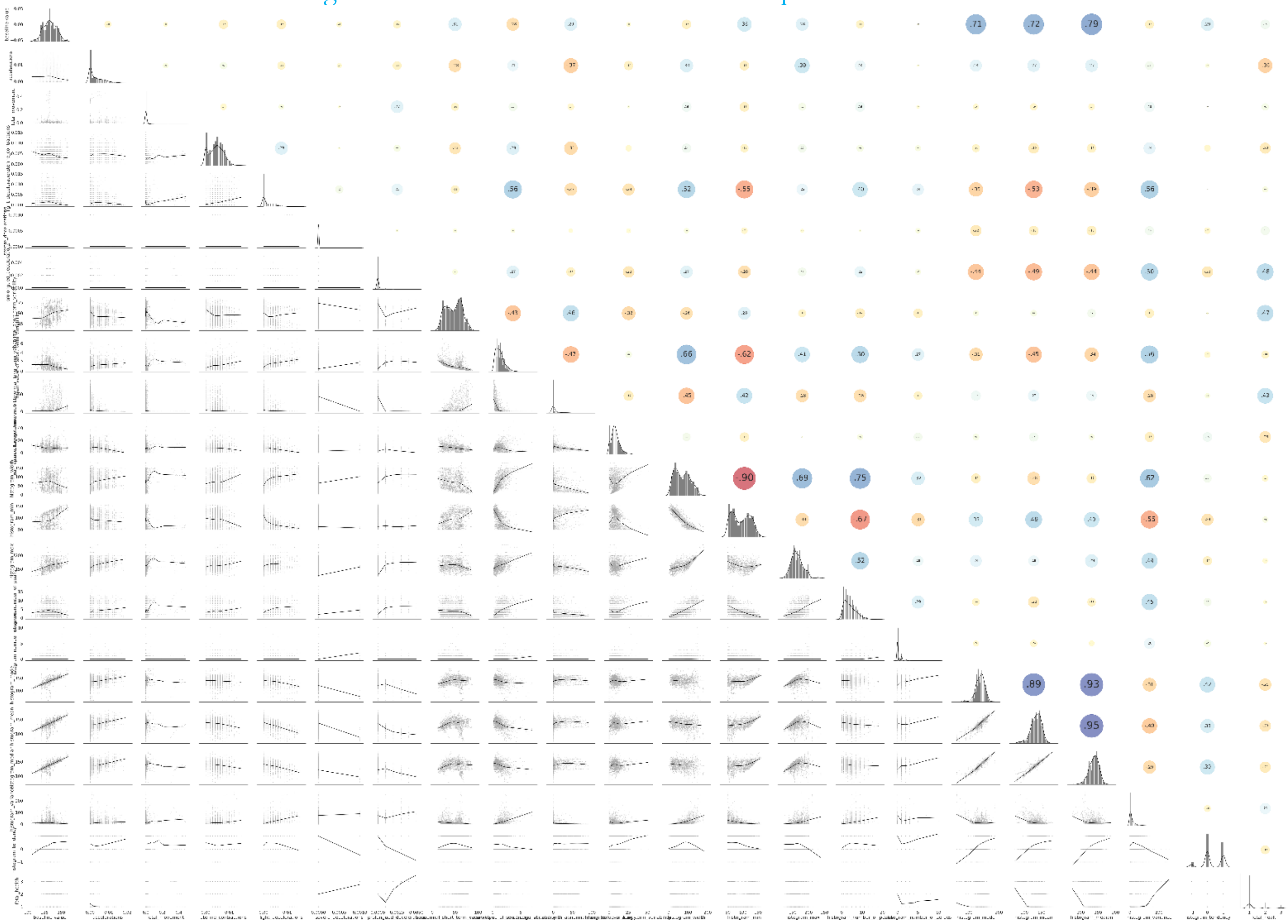
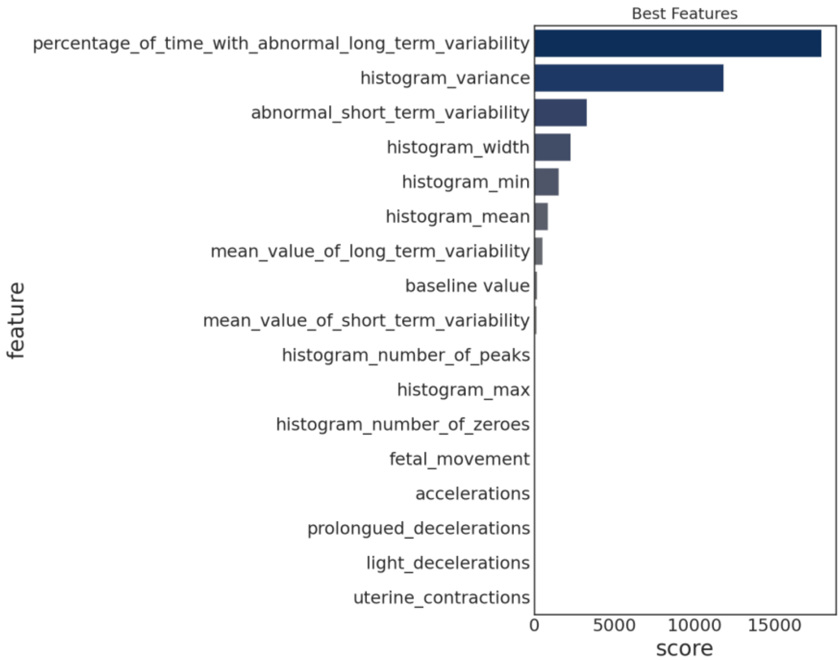


Figure 2. Univariate feature selection



Results

The results of the machine learning models' performance metrics are presented in the table. Each model's accuracy, recall, precision, F1-score, Matthews Correlation Coefficient (MCC) score, training time, prediction time, and total time are shown.

Starting with logistic regression, the model achieved an accuracy of 80.52%. This means that 80.52% of the predictions made by the model were correct. The recall and precision scores were also relatively high at 80.52% and 87.38%, respectively. Recall measures the proportion of true positive instances correctly predicted, while precision measures the proportion of instances predicted as positive that are actually true positive instances. The F1-score, which is the harmonic mean of recall and precision, was 82.47%. The MCC score, which assesses the quality of binary classifications, was 60.30%. The model's training time was relatively low at 0.103 seconds, and the prediction time was even lower at 0.010 seconds, resulting in a total time of 0.114 seconds.

Moving on to k-nearest neighbors (kNN), this model achieved a higher accuracy of 88.97%. The recall and precision scores were also high at 88.97% and 91.64%, respectively. This indicates that the kNN model performed well in correctly identifying true positive instances and minimizing false positives. The F1-score was 89.72%, indicating a good balance between precision and recall. The MCC score was 74.61%, suggesting a relatively strong correlation between the predicted and true classes. The training time was very low at 0.009 seconds, and the prediction time was even lower at 0.004 seconds, resulting in a total time of 0.013 seconds.

Figure 3. Model performances

	Accuracy	Recall	Precision	F1-Score	MCC score	time to train	time to predict	total time
Logistic	80.52%	80.52%	87.38%	82.47%	60.30%	0.103	0.010	0.114
kNN	88.97%	88.97%	91.64%	89.72%	74.61%	0.009	0.004	0.013
SVM	85.45%	85.45%	89.67%	86.66%	68.31%	10.130	0.004	10.134
Decision Tree	83.80%	83.80%	89.51%	85.33%	66.45%	0.020	0.000	0.021
Extra Trees	93.66%	93.66%	93.82%	93.71%	82.78%	1.799	0.202	2.001
Random Forest	95.77%	95.77%	95.87%	95.81%	88.58%	1.554	0.014	1.568
Gradient Boosting	93.19%	93.19%	93.79%	93.39%	82.45%	2.582	0.003	2.585
NN MLP	90.61%	90.61%	92.13%	91.07%	77.37%	12.052	0.002	12.054

The support vector machine (SVM) model achieved an accuracy of 85.45%. The recall and precision scores were also high at 85.45% and 89.67%, respectively. This indicates that the SVM model performed well in correctly identifying true positive instances and had a relatively low rate of false positives. The F1-score was 86.66%, indicating a good balance between precision and recall. The MCC score was 68.31%, suggesting a moderate correlation between the predicted and true classes. However, the training time was relatively high at 10.130 seconds, while the prediction time was very low at 0.004 seconds, resulting in a total time of 10.134 seconds.

The Extra Trees model achieved the highest accuracy among the models at 93.66%. It also had high recall and precision scores at 93.66% and 93.82%, respectively. This indicates that the Extra Trees model performed exceptionally well in correctly identifying true positive instances and had a very low rate of false positives. The F1-score was 93.71%, indicating a strong balance between precision and recall. The MCC score was 82.78%, suggesting a high correlation between the predicted and true classes.

Figure 4. Feature importance decision tree

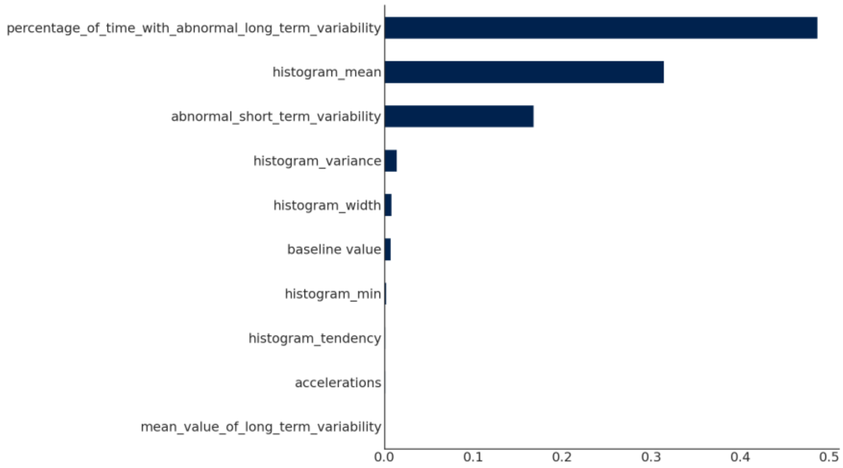
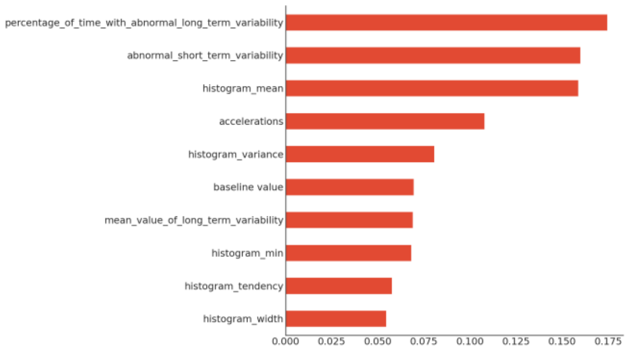
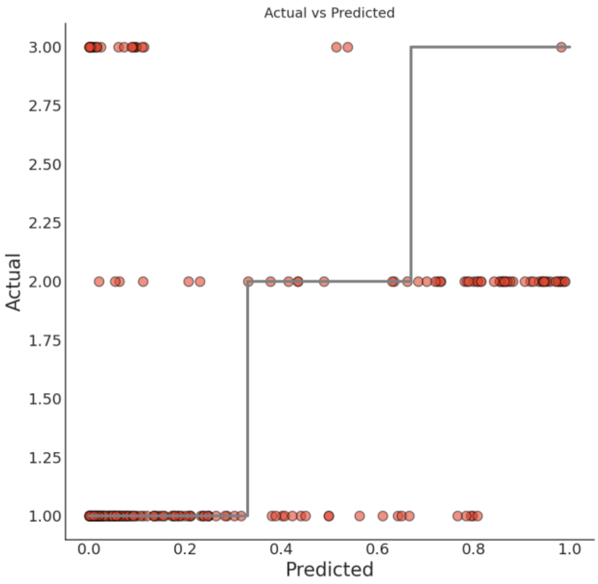


Figure 5. Feature importance and prediction in Extra Tree





The training time for the Extra Trees model was 1.799 seconds, and the prediction time was 0.202 seconds, resulting in a total time of 2.001

seconds. This model demonstrates superior performance compared to the other models in terms of accuracy and precision.

In our study, we analyzed the results of machine learning algorithms to determine the most influential features in predicting fetal health outcomes. Among all the features considered, we found that the following features exhibited significant importance in determining fetal health.

The feature "accelerations" emerged as a crucial predictor. It refers to the number of accelerations per second, indicating how frequently the fetal heart rate accelerates within a given timeframe. A higher value of accelerations suggests a healthier fetal state.

The feature "abnormal_short_term_variability" also demonstrated substantial importance. It represents the percentage of time during which abnormal short-term variability in the fetal heart rate is observed. A higher value of abnormal short-term variability indicates a higher likelihood of an adverse fetal health outcome.

Additionally, the feature "percentage_of_time_with_abnormal_long_term_variability" played a significant role. It signifies the percentage of time with abnormal long-term variability in the fetal heart rate. Higher values of abnormal long-term variability indicate an increased risk of fetal health complications.

The "histogram_width" feature, which measures the width of the histogram created using all values from a record, was found to have considerable feature importance. A wider histogram width suggests a broader range of values in the fetal heart rate, potentially indicating a healthier fetal state.

Furthermore, both the "histogram_mean" and "histogram_variance" features displayed noteworthy importance. The histogram mean represents the average value of the histogram, while the histogram variance reflects the spread or variability of the values. These features provide valuable insights into the overall distribution and central tendencies of the fetal heart rate, enabling predictions about fetal health outcomes.

Lastly, the "histogram_tendency" feature exhibited substantial significance. It describes the trend or direction of the histogram values over time. By capturing the changes and patterns in the fetal heart rate, this feature aids in predicting potential fetal health complications.

Overall, our study identified these features as having large values of feature importance in predicting fetal health outcomes. By leveraging machine learning algorithms, we gained valuable insights into the significance of these features, providing a basis for developing effective models for fetal health assessment and monitoring.

Conclusion

Fetal mortality is a significant public health problem that demands attention and action. Its effects are profound, affecting families emotionally, compromising maternal health, and impacting communities at large [22], [23]. By providing real-time and continuous data on fetal heart rate, movement, and uterine contractions, CTGs empower healthcare professionals to take proactive measures to prevent child and maternal mortality, ensuring safer pregnancies and deliveries for women worldwide.

This study demonstrates the potential of machine learning techniques in predicting fetal health using Cardiotocograms (CTGs). The developed models show promising performance in accurately classifying fetal health conditions. The findings highlight the importance of feature selection in identifying key factors for predicting fetal well-being, such as the rate of accelerations, abnormal variability, and characteristics of the fetal heart rate histogram. These insights contribute to a deeper understanding of the complex relationship between CTG features and fetal health outcomes.

While this study presents promising results in the prediction of fetal health using machine learning and Cardiotocograms (CTGs), there are several limitations that need to be acknowledged. Firstly, the dataset used in this study may not fully capture the diversity and complexity of fetal health conditions. The dataset consists of 2126 records, which might not be representative of the entire population. It is crucial to validate the performance of the developed models on larger and more diverse datasets to ensure their generalizability and robustness across different patient populations.

Secondly, the feature selection process used in this study, specifically the univariate feature selection technique (SelectKBest), may not have captured all the relevant features associated with fetal health. While this technique helps identify the most relevant features, it relies on univariate statistical analysis and does not consider potential interactions or nonlinear relationships among features. Incorporating more advanced feature selection methods, such as recursive feature elimination or dimensionality reduction techniques like principal component analysis, could potentially enhance the performance and interpretability of the predictive models.

Additionally, although the study addresses the issue of class imbalance using the SMOTE RandomOverSampler technique, it is important to acknowledge that oversampling techniques may introduce certain biases or lead to overfitting. The effectiveness of SMOTE in mitigating class imbalance relies on the assumption that synthetic samples generated through interpolation accurately represent the minority class. Therefore, it is crucial to carefully evaluate the impact of oversampling on model performance and explore alternative approaches, such as undersampling or hybrid sampling techniques, to further address class imbalance.

While the developed machine learning models demonstrate promising accuracy, recall, precision, and other performance metrics, the study primarily focuses on the predictive aspect. The clinical interpretability of the models and their integration into real-world healthcare settings require careful consideration. It is essential to conduct further research to assess the clinical utility of the developed models, including their integration into existing clinical workflows, validation through prospective studies, and evaluation of their impact on clinical decision-making and patient outcomes.

References

[1] J. Chagas e Silva, L. Lopes da Costa, and J. Robalo Silva, “Plasma progesterone profiles and factors affecting embryo-fetal mortality following embryo transfer in dairy cattle,” *Theriogenology*, vol. 58, no. 1, pp. 51–59, Jul. 2002.

[2] J. Unterscheider *et al.*, “Fetal growth restriction and the risk of perinatal mortality–case studies from the multicentre PORTO study,” *BMC Pregnancy Childbirth*, vol. 14, no. 1, p. 63, Feb. 2014.

[3] M. Enders, A. Weidner, I. Zoellner, K. Searle, and G. Enders, “Fetal morbidity and mortality after acute human parvovirus B19 infection in pregnancy: prospective evaluation of 1018 cases,” *Prenat. Diagn.*, vol. 24, no. 7, pp. 513–518, Jul. 2004.

[4] Y. Berhan and A. Berhan, “A meta-analysis of selected maternal and fetal factors for perinatal mortality,” *Ethiop. J. Health Sci.*, vol. 24 Suppl, no. 0 Suppl, pp. 55–68, Sep. 2014.

[5] S. Karakolias and C. Kastanioti, “Application of an organizational assessment tool of primary health care,” *Arch Hell Med*, vol. 35, pp. 497–505, 2018.

[6] N. Polyzos, S. Karakolias, G. Mavridoglou, P. Gkorezis, and C. Zilidis, “Current and future insight into human resources for health in Greece,” *Open J. Soc. Sci.*, vol. 03, no. 05, pp. 5–14, 2015.

[7] S. Karakolias, C. Kastanioti, M. Theodorou, and N. Polyzos, "Primary care doctors' assessment of and preferences on their remuneration," *Inquiry*, vol. 54, p. 46958017692274, Jan. 2017.
[8] E. C. W. Gregory, M. F. MacDorman, and J. A. Martin, "Trends in fetal and perinatal mortality in the United States, 2006-2012," *NCHS Data Brief*, no. 169, pp. 1–8, Nov. 2014.
[9] V. Kommaraju, K. Gunasekaran, K. Li, and T. Bansal, "Unsupervised pre-training for biomedical question answering," *arXiv preprint arXiv*, 2020.
[10] K.-H. Yu, A. L. Beam, and I. S. Kohane, "Artificial intelligence in healthcare," *Nat Biomed Eng*, vol. 2, no. 10, pp. 719–731, Oct. 2018.
[11] S. Durga, R. Nag, and E. Daniel, "Survey on Machine Learning and Deep Learning Algorithms used in Internet of Things (IoT) Healthcare," in *2019 3rd International Conference on Computing Methodologies and Communication (ICCMC)*, 2019, pp. 1018–1022.
[12] A. Gillespie *et al.*, "Gene characterization and expression of the γδ T cell co-receptor WC1 in sheep," *Dev. Comp. Immunol.*, vol. 116, p. 103911, Mar. 2021.
[13] Z. Yang *et al.*, "Modified SEIR and AI prediction of the epidemics trend of COVID-19 in China under public health interventions," *J. Thorac. Dis.*, vol. 12, no. 3, pp. 165–174, Mar. 2020.
[14] D. Ravì *et al.*, "Deep Learning for Health Informatics," *IEEE Journal of Biomedical and Health Informatics*, vol. 21, no. 1, pp. 4–21, Jan. 2017.
[15] A. Gillespie *et al.*, "Characterization of the domestic goat γδ T cell receptor gene loci and gene usage," *Immunogenetics*, vol. 73, no. 2, pp. 187–201, Apr. 2021.
[16] O. Stephen, M. Sain, U. J. Maduh, and D.-U. Jeong, "An Efficient Deep Learning Approach to Pneumonia Classification in Healthcare," *J. Healthc. Eng.*, vol. 2019, p. 4180949, Mar. 2019.
[17] H. Su and X. Zhang, "Battery-dynamics driven tdma mac protocols for wireless body-area monitoring networks in healthcare applications," *IEEE J. Sel. Areas Commun.*, vol. 27, no. 4, pp. 424–434, May 2009.
[18] Y. W. Chen and L. C. Jain, *Deep learning in healthcare: Paradigms and applications*, 1st ed. Cham, Switzerland: Springer Nature, 2019.
[19] D. Ayres-de-campos, J. Bernardes, A. Garrido, J. Marques-de-sá, and L. Pereira-leite, "SisPorto 2.0: A program for automated analysis of cardiotocograms," *J. Matern. Fetal. Neonatal Med.*, vol. 9, no. 5, pp. 311–318, Jan. 2000.
[20] J. Sun, J. Lang, H. Fujita, and H. Li, "Imbalanced enterprise credit evaluation with DTE-SBD: Decision tree ensemble based on

SMOTE and bagging with differentiated sampling rates," *Inf. Sci.* , vol. 425, pp. 76–91, Jan. 2018.

[21] J. Telo, "Intrusion Detection with Supervised Machine Learning using SMOTE for Imbalanced Datasets," *Journal of Artificial Intelligence and Machine Learning in Management*, vol. 5, no. 1, pp. 12–24, 2021.

[22] L. M. M. Nardozza *et al.*, "Fetal growth restriction: current knowledge," *Arch. Gynecol. Obstet.*, vol. 295, no. 5, pp. 1061–1077, May 2017.

[23] M. D. Givens and M. S. D. Marley, "Infectious causes of embryonic and fetal mortality," *Theriogenology*, vol. 70, no. 3, pp. 270–285, Aug. 2008.