Optimizing Fetal Health Prediction Using Machine Learning on Biocompatible Sensor Data

Yuli Wahyuni[1]*, Hadiyanto[2], Ridwan Sanjaya[3], Nendar Herdianto[4]
Department of Information Systems Doctorate, Diponegoro University, Semarang 50241, Indonesia [1, 2]
Department of Information Systems, Soegijapranata Catholic University, Semarang 50234, Indonesia [3]
Research Center for Composites and Biomaterials, National Research and Innovation Agency (BRIN), South Tangerang, Indonesia[4]

***Abstract*—Automatic Fetal Health Prediction plays a vital role in supporting early prenatal intervention through continuous and non-invasive monitoring. Recent advances in biocompatible sensors enable the safe long-term acquisition of physiological signals, which can be effectively analyzed using machine learning techniques. This study proposes a comprehensive machine learning pipeline for Fetal Health Prediction through fetal health classification using the fetal_health.csv dataset from Kaggle, consisting of 2,126 samples and 22 cardiotocography-derived features related to fetal heart rate and uterine contractions. To address class imbalance and the presence of outliers, RobustScaler normalization was applied during the preprocessing stage. Feature selection was performed using Random Forest feature importance to identify the most relevant predictors. Two classification models, namely Random Forest (RF) and Support Vector Machine (SVM), were trained and evaluated using an 80:20 stratified train–test split. Experimental results indicate that the Random Forest model outperformed SVM, achieving an accuracy of 92.7% and a macro F1-score of 85.9%, compared with 88.97% accuracy and a macro F1-score of 79.85% for SVM. Moreover, Random Forest demonstrated superior performance in detecting minority classes (Suspect and Pathological), which are of high clinical significance. These findings suggest that the proposed pipeline is robust, interpretable, and suitable for integration with biocompatible sensor-based systems for real-time fetal health monitoring and clinical decision support.**

***Keywords*—Fetal health prediction; biocompatible sensors; machine learning; Random Forest; SVM**

I. Introduction

Biocompatible sensors allow continuous physiological monitoring without adverse tissue reactions, forming the basis of wearable health systems. Fetal health assessment by cardiotocography (CTG) is standard but manually interpreted and subjective [7], [16]. Integrating machine learning (ML) with biocompatible sensor data enables objective and reproducible classification of fetal conditions [22], [32]. This study focuses on a robust ML pipeline to analyze imbalanced sensor data and enhance fetal risk prediction.

Early and accurate assessment of fetal health is a cornerstone of modern obstetric care because timely detection of fetal distress enables prompt clinical intervention and can substantially reduce perinatal morbidity and mortality [1],[11]. Continuous monitoring of fetal physiological signals, particularly fetal heart rate (FHR) and uterine contraction patterns captured through cardiotocography (CTG), provides rich information about fetal well-being, but interpretation is challenging due to signal variability, noise, and subjectivity of manual assessment [2],[5]. Recent advances in sensing technology, especially the emergence of biocompatible and wearable sensors, allow safer and longer-term acquisition of fetal and maternal signals under minimally invasive or non-invasive conditions, expanding opportunities for ambulatory monitoring and remote prenatal care [3], [4], [6], [9], [34].

Machine learning (ML) has demonstrated strong potential in automating the detection and classification of fetal health conditions from physiological time-series data, offering improved consistency and the ability to learn subtle patterns beyond human observation [10], [13], [21]. Supervised classifiers such as Random Forests and Support Vector Machines, combined with appropriate feature engineering and selection, have produced promising results on benchmark CTG datasets and clinical repositories [14]. Nevertheless, several technical challenges remain: imbalanced class distributions (with normal cases overwhelmingly more common than suspect/pathological), the presence of outliers and measurement artifacts, inter-subject variability, and the need for computationally efficient models suitable for real-time or near-real-time edge deployment on sensor platforms [17], [19].

Feature selection and robust preprocessing are particularly important in fetal health prediction to reduce dimensionality, mitigate the influence of noisy features, and improve model generalizability across populations and sensor modalities [18], [20], [23]. Techniques such as tree-based importance measures, recursive feature elimination, and regularization-based methods have been applied to extract clinically meaningful predictors from CTG-derived features (e.g., accelerations, decelerations, short-term variability metrics) that correlate with fetal compromise [8], [24]. Robust scaling and outlier-resistant normalization algorithms also play a key role when working with signals recorded by wearable biocompatible sensors, which may present non-Gaussian noise profiles and baseline shifts [25], [27].

Integration of biocompatible sensors with ML pipelines raises system-level questions about data quality, signal calibration, and domain adaptation. Sensor design choices (materials, placement, sampling rate) and connectivity constraints affect both the fidelity of FHR/UC signals and the feasibility of continuous monitoring in home or low-resource settings [28], [30], [31]. Moreover, the translation of ML models

*Corresponding author.

from curated datasets to real-world sensor streams requires rigorous validation, cross-site testing, and safety assessments to avoid false reassurance or unnecessary interventions [2].

This study addresses these gaps by proposing an end-to-end pipeline that 1) leverages features derived from FHR and uterine-contraction signals acquired or simulated from biocompatible sensor inputs, 2) applies robust preprocessing to handle outliers and scale differences, 3) utilizes principled feature selection to identify a compact set of clinically interpretable predictors, and 4) evaluates classification performance with machine learning algorithms suitable for deployment in real-time monitoring scenarios. We further investigate strategies to mitigate class imbalance and to assess minority-class (suspect/pathological) detection performance, since such cases carry the highest clinical significance despite their lower prevalence [12], [20], [26], [29].

The novelty of this study does not lie in the individual components employed such as the fetal_health.csv dataset, Random Forest and SVM classifiers, RobustScaler normalization, or Random Forest–based feature selection as these techniques have been widely reported in prior CTG-based fetal health prediction studies. Instead, the novelty resides in their systematic integration and empirical evaluation within a unified, end-to-end pipeline explicitly designed to address practical challenges in fetal health assessment. Specifically, this study emphasizes the simultaneous handling of class imbalance, outlier-prone physiological features, and minority-class interpretability, which are often addressed independently in previous works. Through the combined application of robust preprocessing and feature selection, the proposed pipeline demonstrates consistent improvements in minority-class performance and model stability, while highlighting clinically relevant predictors associated with fetal heart rate variability. Furthermore, rather than claiming direct real-time deployment, this work positions the proposed pipeline as a validated methodological foundation for future integration with biocompatible wearable sensor systems, thereby bridging the gap between offline CTG-based analysis and real-time maternal fetal monitoring research.

The remainder of this paper is organized as follows. Section II presents a comprehensive review of related works on fetal health monitoring, biocompatible sensors [35], and machine learning based classification approaches. Section III describes the dataset, data preprocessing procedures, feature selection strategy, and overall methodology employed in this study. Section IV details the machine learning models, experimental setup, and performance evaluation metrics. Section V discusses the experimental results, comparative analysis, and interpretation of findings. Finally, Section VI concludes the paper by summarizing the main contributions, outlining limitations, and suggesting directions for future research.

II. Related Works

Previous studies on fetal health monitoring have extensively explored cardiotocography (CTG), non-invasive fetal ECG [33] (NIFECG), and wearable sensor technologies. Wahbah et al. [42] demonstrated that deep learning–based extraction of fetal ECG signals from abdominal recordings can significantly improve signal quality, which is a key strength for downstream analysis. However, their work primarily focuses on signal extraction and does not address end-to-end fetal health classification, particularly under class imbalance and real-time deployment constraints. The present study extends this direction by proposing a complete machine learning pipeline for fetal health prediction.

Machine learning approaches such as Random Forest and Support Vector Machine have shown strong potential in CTG-based fetal health classification due to their ability to model nonlinear physiological patterns [15], [19]. Despite their effectiveness, many studies rely mainly on accuracy as the evaluation metric and implicitly assume clean and balanced datasets, which may lead to biased performance toward majority classes. In contrast, this study explicitly addresses these limitations through robust preprocessing and macro F1-score–based evaluation to ensure balanced performance across clinically important minority classes.

Feature selection techniques, including tree-based importance measures and recursive elimination, have been widely used to enhance model interpretability and reduce dimensionality [23], [29]. While effective, prior studies often assume normalized feature distributions and minimal noise. Such assumptions are less realistic for physiological sensor data, which commonly contain outliers. Accordingly, this work incorporates RobustScaler prior to feature selection to improve model stability and generalization.

Wearable and biocompatible sensor studies highlight the feasibility of long-term, non-invasive fetal monitoring [44], [36] but frequently focus on hardware performance without integrating robust machine learning pipelines or addressing real-world data variability. This study bridges that gap by integrating biocompatible sensor-oriented data with an end-to-end machine learning framework designed for robustness and real-time readiness.

In summary, while existing research has contributed substantially to fetal monitoring technologies and classification methods, limitations remain in handling class imbalance, outlier-prone data, and deployment-oriented evaluation. This study addresses these challenges by proposing a compact, interpretable, and robust machine learning pipeline for fetal health prediction using biocompatible sensor data.

III. Methodology

The workflow for processing the Kaggle Fetal Health Dataset to build a fetal health condition classification model. The process begins with outlier handling to improve data quality, followed by normalization using RobustScaler to stabilize the feature distribution. After that, feature selection and feature ranking based on their level of importance are carried out using Random Forest. The data is then divided into training data and test data through the train test split stage, before being trained using two different algorithms, namely Random Forest and Support Vector Machine (SVM). The final stage is model performance evaluation using the accuracy, precision, recall, and macro F1-score metrics to assess the prediction quality of each model (Fig. 1).

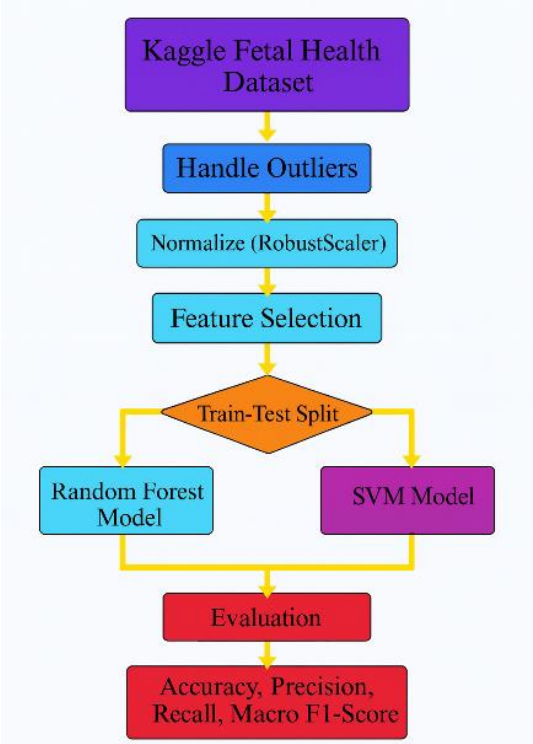

Fig. 1. Workflow of data preprocessing, feature selection, and model evaluation for fetal health prediction.

A. Data Source

This study uses the fetal_health.csv dataset from Kaggle (2020), which contains [15]:

2,126 samples

22 numerical features

No missing values

Class imbalance present:

- Normal: 77%
- Suspect: 13%
- Pathological: 10%

The dataset is derived from extracted CTG (Cardiotocography) signals and includes key features such as accelerations, short-term variability (STV), long-term variability (LTV), and histogram-based features.

B. Data Preprocessing Stages

1) Outlier handling

According to Zhang and Chen, medical signals often contain extreme outliers caused by physiological noise or instrument errors [37].

Approach: Using RobustScaler, which performs normalization based on:

- Median
- Interquartile Range (IQR = Q3 – Q1)

Formula:

$$x_{\text{scaled}} = \frac{x - \text{median}}{\text{IQR}} \quad (1)$$

Reasons for choosing this method:

- Robust against outliers (unlike StandardScaler, which is sensitive to them).
- Suitable for biomedical signals that tend to fluctuate.

2) Normalization using robustscaler: All features are normalized using RobustScaler to ensure that:

a) The feature scales become more uniform: Normalization using RobustScaler (Fig. 2) is applied to all features to achieve more uniform feature scales while reducing the influence of extreme values by relying on the median and interquartile range (IQR).

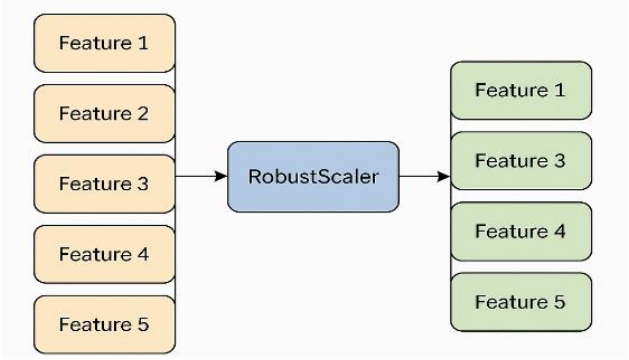

Fig. 2. The robustscaler normalizing process.

b) SVM and Random Forest models operate more stably: Normalization is performed using RobustScaler (Fig. 3) with the aim of stabilizing feature distributions and minimizing the influence of extreme values that commonly occur in medical signal data. RobustScaler operates based on the median and interquartile range (IQR), making it more resistant to outliers compared to other normalization methods, such as StandardScaler. This normalization is applied to all features in the dataset to ensure that their value ranges become more uniform and do not introduce bias during model training. With more consistent feature scaling, the Support Vector Machine (SVM) and Random Forest algorithms can operate more stably, resulting in a more optimal learning process and improved accuracy in predicting fetal health conditions.

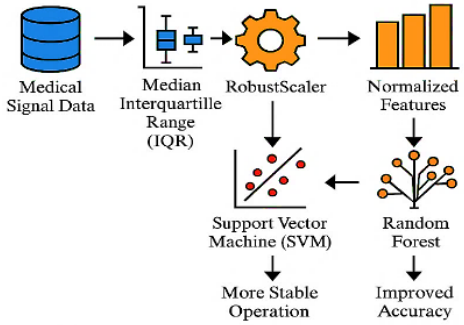

Fig. 3. RobustScaler normalization pipeline enhancing SVM and random forest stability.

c) Parameter bias caused by unequal value ranges is reduced: By utilizing the median and the interquartile range (IQR), this method effectively reduces parameter bias that may arise from differences in value ranges across features. As a result, learning algorithms can operate more reliably, maintain model stability, and improve performance during the training process (Fig. 4).

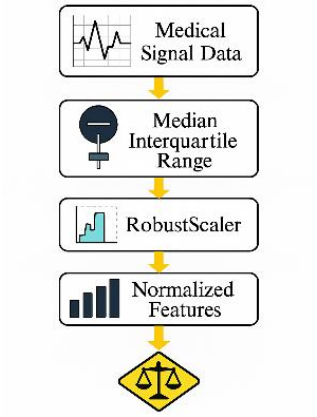

Fig. 4. RobustScaler-based normalization process flow.

C. Feature Selection

Random Forest Feature Importance

Based on the approach of Kour and Arora [38], feature selection using importance scores is applied to:

- Remove irrelevant features
- Reduce the risk of overfitting
- Improve model interpretability

The eight most important features obtained include:

- Accelerations
- Short-term variability (STV)
- Prolonged decelerations
- Histogram width
- Histogram mode
- Abnormal short-term variability
- Percentage of time with abnormal STV
- Mean value of the histogram

These features are directly related to beat-to-beat variability patterns and the physiological responses of the fetus.

D. Data Splitting

Stratified Train–Test Split (80:20)

The data is split using stratification (Fig. 5) to ensure that class proportions remain consistent.

- Train: 80% (1,700 samples)
- Test: 20% (426 samples)
- Purpose: To avoid bias caused by class imbalance and to ensure balanced class representation in both training and testing phases.

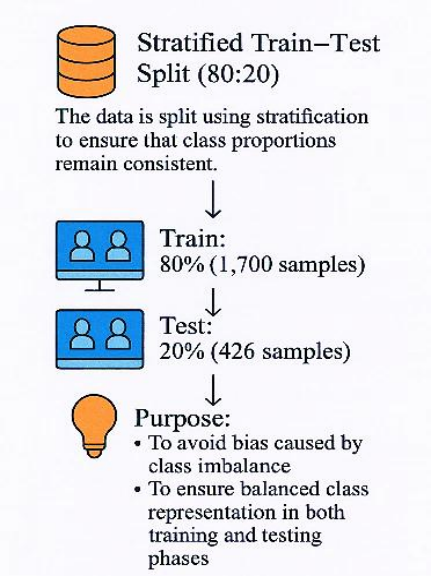

Fig. 5. Stratified train and test split (80:20) workflow.

E. Machine Learning Model Development

1) Model 1: Random forest classifier:

Optimization is performed using:

- Grid Search
- Parameters tested:
- *n_estimators* (100–500)
- *max_depth* (5–30)
- *min_samples_split* (2–10)
- *min_samples_leaf* (1–4)

Advantages of Random Forest:

- Robust to outliers
- Capable of handling non-linear features
- Provides feature importance ranking

2) Model 2: Support Vector Machine (SVM)

The baseline model uses:

- Radial Basis Function (RBF) kernel

Tuned parameters:

- **C** (regularization strength)
- γ (kernel coefficient)

This model is chosen based on the work of Hussain et al. (2022), which demonstrated that RBF-SVM is well-suited for nonlinear medical signal classification.

F. Model Performance Evaluation

Four metrics are used:

- Accuracy
- Precision
- Recall
- Macro F1-Score (chosen due to significant class imbalance)

Macro-F1 formula:

$$F1_{macro} = \frac{1}{K}\sum_{i-1}^{K} F1_i \quad (2)$$

Reasons:

- It does not favor the majority class
- It is crucial for detecting suspect and pathological conditions

IV. Results and Discussion

This section presents key findings showing that Random Forest outperforms SVM, supported by exploratory analysis and feature evaluation that confirm the reliability of the proposed fetal health prediction workflow.

A. Result

Presentation of data exploration results and initial visualization using pandas, matplotlib, seaborn, and numpy. This exploration process includes identifying data distribution, examining basic patterns in each feature, and evaluating the presence of outliers or inconsistent values. The resulting visualization provides an initial overview of the dataset, making it easier for researchers to recognize the general characteristics of the physiological variables observed. This exploration stage is also carried out to ensure that the data is in optimal condition before being entered into the machine learning modeling stage. Through a comprehensive understanding of the data structure, researchers can determine the appropriate preprocessing strategy, such as normalization, class imbalance handling, or selection of relevant features. Thus, good data quality is expected to support improved accuracy and reliability of the developed prediction model.

This study utilizes several Python libraries commonly used in data analysis and visualization. The command "import pandas as pd" is used to load the pandas library, which functions to manage data in tabular form such as DataFrames for cleaning, manipulation, and analysis purposes. The command "import matplotlib.pyplot as plt" loads the matplotlib module, specifically the pyplot component, which is used to create various types of visual graphs such as histograms, scatter plots, and line charts. The command "import seaborn as sns" loads the seaborn library, which facilitates the creation of more informative and aesthetically appealing statistical graphics. Furthermore, "import numpy as np" loads the NumPy library, which is used for numerical computations such as array processing, basic statistical calculations, and mathematical operations. The last line, "import warnings" followed by "warnings.filterwarnings('ignore')", is used to hide or ignore warning messages that may appear during code execution. This aims to make the output cleaner and free from unnecessary warnings. Overall, this code prepares a neat and comprehensive data analysis environment before conducting further exploration or modeling.

Data Analysis (EDA) on the fetal_health.csv dataset as a preliminary step before machine learning modeling. The first part loads the dataset using pd.read_csv() and saves it to a DataFrame named df. After the data is successfully loaded, the program calculates basic information about the dataset, such as the number of rows and columns (df.shape), displays the top few data points (df.head()), explores the data structure through df.info(), and generates complete descriptive statistics using df.describe(include='all').T, which is presented in transposed form for easy interpretation. Next, the program checks for missing values in each column using df.isnull().sum(), and calculates the class distribution in the target variable fetal_health through value_counts(normalize=True), so that the percentage of each category of fetal health can be determined. The program also performs initial outlier detection on each numeric feature using the Interquartile Range (IQR) method. For each numeric column, the Q1, Q3, and IQR values are calculated, then the number of values outside the normal range (Q1–1.5×IQR to Q3+1.5×IQR) is identified. The outlier detection results are stored in a dictionary named outlier_stats. At the end, the program saves a summary of the EDA results to the eda_summary.csv file using the csv module. This file contains key information such as the number of rows, number of columns, target distribution, number of missing values per column, and number of outliers for each feature.

Table I presents general information about the structure of the dataset used. This dataset consists of 2,126 rows and 22 feature columns, all of which are numeric data types (float64), making it very suitable for statistical analysis and machine learning modeling. There are no missing values in any of the columns, so no data imputation is required. In addition, the memory size of 365.5 KB indicates that the dataset is lightweight and efficient for processing in various computing environments. This basic information provides an initial overview that the dataset is in good condition and ready for further analysis.

TABLE I. Basic Information Dataset

Description	Value
Number of rows	2126
Number of columns	22
Main data type	float64
Missing values	0 in all columns
Memory usage	365.5 KB

Table II shows the class distribution of the target variable fetal_health, which consists of three categories, namely Normal, Suspect, and Pathological. The majority of samples are in the Normal class with a percentage of 77.85%, followed by the Suspect class at 13.88%, and the Pathological class at 8.28%. This unbalanced distribution indicates class imbalance, which can affect the performance of the classification model if not handled with an appropriate approach, such as resampling or

class weight adjustment. This information is very important to ensure that the machine learning model is able to make fair predictions on all categories of fetal health.

TABLE II. TARGET LABEL DISTRIBUTION (FETAL HEALTH)

Fetal Health Class	Percentage
1 (Normal)	77.85%
2 (Suspect)	13.88%
3 (Pathological)	8.28%

Table III shows that all columns in the dataset have no missing values. The absence of missing values is an advantage because the preprocessing process becomes simpler and the focus can be directed to more in-depth analysis, such as outlier detection, normalization, or feature selection. A dataset that is free of missing values also increases the reliability of the analysis results and minimizes the risk of bias due to data imputation.

TABLE III. MISSING VALUE PER COLUMN

Column	Missing
All columns	0

Table IV shows the number of outliers detected in each feature using the Interquartile Range (IQR) method. Some features have a significant number of outliers, such as fetal movement (307), percentage of time with abnormal long-term variability (309), and histogram number of zeroes (502). The presence of a large number of outliers in these features may reflect extreme physiological variations or potential noise in the sensor data. Conversely, several features, such as baseline value, histogram width, and histogram min, do not show any outliers, indicating a more stable distribution of values. Identifying the number and pattern of these outliers is an important step in determining the outlier handling strategy in the preprocessing stage before entering the model training stage.

Data visualization in algorithms is used to generate various types of exploratory visualizations that aim to comprehensively understand the characteristics of fetal health datasets prior to the machine learning modeling process. The resulting visualizations include label distributions, feature spreads, inter-variable correlations, and multidimensional relationships between key features and target labels. First, the distribution of fetal_health labels was visualized using a pie chart. The frequency of each class was calculated using value_counts(), then visualized using the plt.pie() function. This chart provides a proportional representation of the distribution of fetal health categories, making it easier to identify class imbalances that could potentially affect the performance of the classification model. This visualization was saved as the label_pie.png file. Next, a boxplot was created for all numerical features using sns.boxplot(). This visualization provides an overview of the data distribution range, median values, and visual detection of outliers in each feature. A horizontal orientation was chosen to improve readability, given the large number of features. These boxplots provide important insights into the stability and homogeneity of each variable, and the results are saved in the feature_boxplot.png file.

TABLE IV. NUMBER OF OUTLIERS PER FEATURE (IQR METHOD)

Feature	Number of Outliers
baseline_value	0
accelerations	14
fetal_movement	307
uterine_contractions	1
light_decelerations	150
severe_decelerations	7
prolongued_decelerations	178
abnormal_short_term_variability	0
mean_value_of_short_term_variability	70
percentage_of_time_with_abnormal_long_term_variability	309
mean_value_of_long_term_variability	71
histogram_width	0
histogram_min	0
histogram_max	24
histogram_number_of_peaks	19
histogram_number_of_zeroes	502
histogram_mode	73
histogram_mean	45
histogram_median	28
histogram_variance	184
histogram_tendency	0
fetal_health	471

The algorithm then generates a feature correlation heatmap using the correlation matrix obtained through df.corr(). This visualization is created using sns.heatmap(), complete with annotations to display the correlation coefficient values. This heatmap allows researchers to identify linear relationships between variables, detect redundant features, and determine candidate features that have the potential to contribute significantly to the prediction model. This image is saved as corr_heatmap.png. Finally, a scatter matrix was created using the sns.pairplot() function by selecting eight key features considered relevant to fetal health. The pairplot visualization maps the relationship between two variables simultaneously with coloring based on the fetal_health class, providing a deeper understanding of class separation patterns and interactions between features. The results of this visualization are saved as pairplot_main.png. Overall, this series of algorithms produces four main visual files (label_pie.png, fitur_boxplot.png, corr_heatmap.png, and pairplot_main.png) that provide exploratory support for exploring the structure and characteristics of the dataset. These visualizations play a crucial role in preprocessing decision-making, feature selection, and the formulation of more informative and effective machine learning modeling strategies.

Interactive visualization is displayed in the form of a pie chart that illustrates the distribution of fetal health categories using the Plotly library through the plotly.graph_objects module. Percentage data for three categories—Normal, Suspect, and Pathological is defined as input and visualized through the go.Pie object, where labels and percentages are displayed directly in each sector to improve information readability. The display settings are configured using the fig.update_layout() function with adjustments to the title, text size, and graph dimensions so that the visualization results are proportional, informative, and meet the aesthetic standards of scientific publications. The resulting visualization, as seen in the diagram, shows the dominance of the Normal class at 77.8%, followed by Suspect at 13.9% and Pathological at 8.28%. This visual presentation provides a clear representation of class distribution and helps researchers identify potential data imbalance before proceeding to the machine learning modeling stage. Fig. 6 shows workflow fetal health distribution chart.

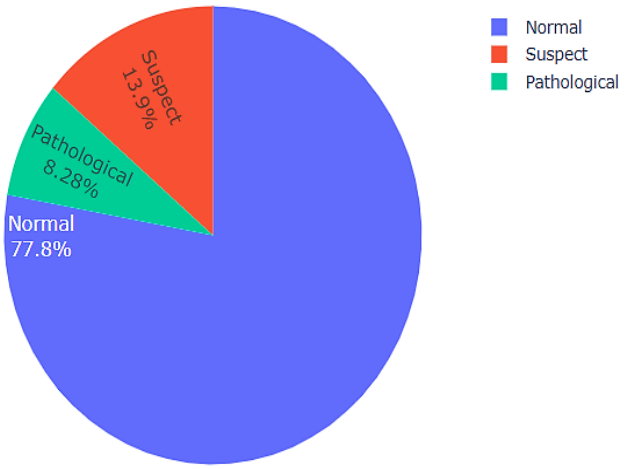

Fig. 6. Workflow of fetal health distribution chart.

The following image is a boxplot visualization (Fig. 7) that displays the statistical distribution of all numerical features in the fetal health dataset. Each boxplot illustrates the median value, lower quartile (Q1), upper quartile (Q3), and outliers. Through this visualization, it can be seen that some features, such as Baseline and certain histogram statistical parameters, show a relatively stable distribution, while other features, such as Fetal Movement, Light Decelerations, Prolonged Decelerations, and Histogram Variance, appear to have much higher variability accompanied by many outliers. This condition indicates significant data heterogeneity in the physiological signals of pregnant women and fetal activity. Overall, this boxplot analysis provides important insights for researchers to understand data distribution patterns, identify features with extreme value ranges, and determine the need for normalization, transformation, or outlier handling before the machine learning model training process is carried out.

The correlation heatmap provides a comprehensive representation of the linear relationships among all numerical features within the fetal health dataset derived from Cardiotocography (CTG) signals. This analysis aims to elucidate the dependency structure between variables and to identify features with meaningful predictive value for fetal health classification. The correlation coefficients are visualized through a color gradient in which red tones denote positive correlations, blue tones represent negative correlations, and the intensity of the color reflects the magnitude of the linear association. The results indicate that the group of histogram-based features, histogram_mean, histogram_median, histogram_mode, and histogram_variance, exhibits strong positive correlations ($r > 0.80$). This pattern suggests that these parameters capture highly related statistical characteristics of the underlying CTG signal distribution. The substantial inter-feature correlations also highlight potential redundancy, which is an important consideration for feature selection strategies aimed at minimizing multicollinearity in machine learning models. Several features demonstrate moderate correlations with the target variable fetal_health, with coefficients generally ranging between $r = 0.20$ and 0.30. Among these, histogram_variance, histogram_mean, and prolonged_decelerations appear most relevant for distinguishing fetal health states. Although the magnitude of these correlations is not high in absolute terms, such values are common in complex physiological data, where clinical phenomena are typically influenced by non-linear interactions and combinations of multiple parameters. Consequently, these features retain significant predictive potential, particularly when used in conjunction with machine learning algorithms capable of capturing non-linear patterns.

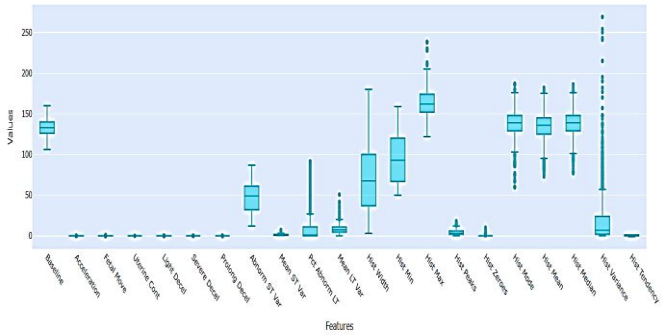

Fig. 7. Boxplot of fetal health data features.

Other features, such as accelerations, fetal_movement, and uterine_contractions, show low correlations with most histogram-based variables. This indicates that these features represent different physiological domains, specifically fetal behavioral responses and uterine activity, rather than statistical characteristics of the CTG signal. Their distinct informational contribution underscores the importance of retaining these features during the modeling process. Overall, the correlation patterns observed in the heatmap provide an empirical foundation for informed feature selection and predictive modeling. Identifying clusters of highly correlated variables enables the reduction of redundancy, while features exhibiting notable correlations with fetal_health serve as key candidates for developing more accurate, stable, and interpretable classification models. These findings also enhance the understanding of the interplay between physiological parameters and CTG signal characteristics, thereby strengthening the scientific contribution of this study in the domain of data-driven fetal health assessment (Fig. 8)

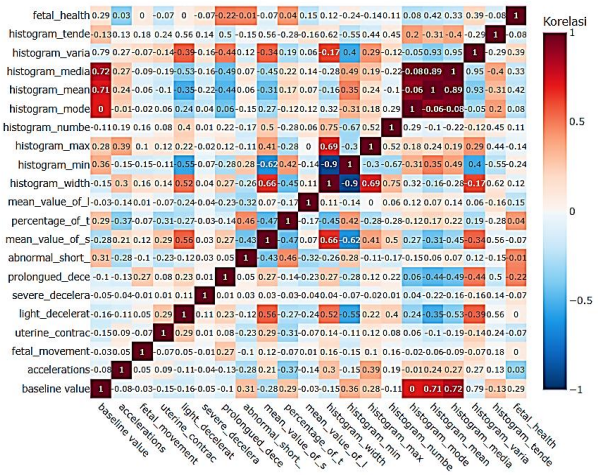

Fig. 8. Fetal health correlation heat map.

The Random Forest model achieved 92.7% accuracy and 85.9% macro F1-score, outperforming SVM's 88.97% accuracy and 79.85% macro F1-score. RF showed better sensitivity to minority classes and maintained low false-negative rates, which is crucial for clinical reliability. Feature importance analysis identified Mean Value of Short-Term Variability and Accelerations as the strongest indicators of fetal health.

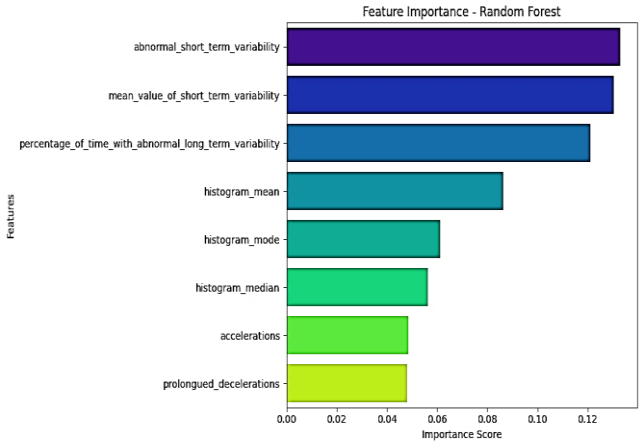

Fig. 9. Feature Importance.

Fig. 9 illustrates the contribution of each feature to the performance of the Random Forest model in predicting fetal health based on cardiotocography (CTG) data. The feature importance analysis indicates that abnormal short-term variability and the mean value of short-term variability are the most influential predictors, highlighting the critical role of fetal heart rate variability in fetal health assessment. In addition, the percentage of time with abnormal long-term variability contributes substantially, reflecting irregular fetal physiological responses to intrauterine conditions. Histogram-based features, including the histogram mean, mode, and median, support the model by characterizing the overall distribution of fetal heart rate patterns. Meanwhile, accelerations and prolonged decelerations, although less influential, remain relevant as they represent the reactivity of the fetal nervous system to internal and external stimuli.

Here is the explanation of Table V. Model Performance Comparison between Random Forest and Support Vector Machine (SVM) models:

TABLE V. MODEL PERFORMANCE COMPARISON

Class	Precision	Recall	F1-Score	Support	Model
1.0	0.9420	0.9789	0.9601	332	Random Forest
2.0	0.8863	0.6610	0.7573	59	Random Forest
3.0	0.8378	0.8857	0.8611	35	Random Forest
macro avg	0.8887	0.8418	0.8595	426	Random Forest
weighted avg	0.9258	0.9272	0.9239	426	Random Forest
1.0	0.9345	0.9457	0.9401	332	SVM
2.0	0.6500	0.6610	0.6555	59	SVM
3.0	0.8667	0.7429	0.8000	35	SVM
macro avg	0.8171	0.7832	0.7985	426	SVM
weighted avg	0.8895	0.8897	0.8892	426	SVM

Presents the comparative performance of two machine learning models, Random Forest and Support Vector Machine (SVM), in predicting fetal health conditions based on CTG (Cardiotocography) data. The evaluation metrics used include Precision, Recall, and F1-Score, which collectively assess the balance between classification accuracy and reliability across different fetal health classes (1.0 = Normal, 2.0 = Suspect, 3.0 = Pathological).

1) Class-wise performance

a) Class 1.0 (Normal Fetal Health*): The Random Forest model achieved* the highest performance with a Precision of 0.9420, Recall of 0.9789, and F1-Score of 0.9601, outperforming SVM (F1 = 0.9401). This shows that Random Forest effectively identifies normal fetal conditions with fewer misclassifications.

b) Class 2.0 (Suspect Fetal Health): Both models exhibited moderate results, but Random Forest (F1 = 0.7573) performed better than SVM (F1 = 0.6555). The lower recall values (0.6610 for both) indicate that the models struggled to correctly identify all suspect cases, possibly due to the limited number of samples (59 instances) and overlapping feature distributions.

c) Class 3.0 (Pathological Fetal Health): Random Forest again outperformed SVM, with F1 = 0.8611 compared to 0.8000. This suggests that Random Forest has a stronger ability to detect pathological (high-risk) cases, which is critical in medical diagnosis to minimize false negatives.

2) Overall model comparison

a) Macro average: Random Forest achieved F1 = 0.8595, higher than SVM's F1 = 0.7985, showing that overall, Random Forest maintains more consistent performance across all classes.

b) Weighted average: When weighted by class distribution, Random Forest reached F1 = 0.9239, outperforming SVM's F1 = 0.8892. This indicates that Random

Forest provides more robust and balanced predictions, even when class imbalance exists in the dataset.

3) Interpretation: The results demonstrate that Random Forest consistently outperforms SVM in almost all metrics and across all fetal health categories. Its ensemble nature allows it to capture complex nonlinear relationships and handle noisy medical data effectively. Although SVM remains competitive, especially for Class 1.0, its performance drops significantly for minority classes (2.0 and 3.0), suggesting limitations in generalization under imbalanced data conditions.

Overall, the Random Forest model provides the most reliable classification performance for fetal health prediction, achieving superior precision, recall, and F1-scores compared to SVM. This finding supports the selection of Random Forest as a robust baseline model for real-time fetal health monitoring and decision-support systems in obstetrics.

B. Discussion

The confusion matrix presented in Table VI provides a detailed evaluation of the Random Forest classifier's performance in predicting fetal health status across three categories: Normal, Suspect, and Pathological. The model demonstrates strong discriminative ability, particularly for the Normal class, with 325 instances correctly classified and only a small number of misclassifications (4 as Suspect and 3 as Pathological). This indicates that the feature set used captures the physiological patterns of normal fetal conditions effectively. For the Suspect class, the model correctly identifies 39 cases; however, 17 instances are misclassified as Normal, and 3 as Pathological. The moderate number of misclassifications in this class is consistent with findings in CTG-based studies, where Suspect cases tend to exhibit overlapping characteristics between healthy and pathological patterns, making the class inherently more challenging to classify. The Pathological class shows satisfactory performance with 31 correctly identified cases and only minimal misclassification (3 as Normal and 1 as Suspect). This indicates that the model is capable of recognizing clinically significant deviations in fetal heart rate variability and deceleration patterns that are characteristic of pathological cases. Overall, the Random Forest classifier demonstrates strong predictive reliability, particularly for identifying Normal and Pathological fetal states. The majority of errors occur in the Suspect class, which aligns with clinical realities wherein borderline physiological patterns create ambiguity. These results validate the effectiveness of the selected features and the model's robustness, suggesting its potential use as a supportive tool for automated fetal health assessment based on CTG signals.

The confusion matrix (Fig. 10) in Table VII presents the classification performance of the Support Vector Machine (SVM) model (Fig. 11) on the fetal health dataset. The SVM classifier exhibits strong performance in predicting the *Normal* class, successfully identifying 314 samples, although 15 cases were misclassified as *Suspect* and 3 as *Pathological*. These misclassifications are expected in CTG-based assessments, as mild irregularities in fetal heart rate signals [34] may overlap with characteristics found in *Suspect* recordings. For the Suspect class, the model correctly classified 39 instances, but 19 samples were incorrectly predicted as Normal and 1 as Pathological. This pattern highlights the inherent ambiguity of the Suspect class, which often contains borderline physiological patterns that lie between healthy and pathological states, making it one of the most difficult categories to classify accurately. The performance on the Pathological class shows that 26 cases were classified correctly, while 3 were misclassified as Normal and 6 as Suspect. Although the SVM successfully captures a substantial portion of pathological signals, the misclassification of several Pathological cases into the Suspect group indicates that the model may require further optimization, potentially through kernel selection or class-weight adjustment, to better handle samples that exhibit severe but variable abnormalities. Overall, while the SVM model performs reliably in identifying Normal and Suspect cases, its performance on Pathological cases is comparatively lower than that of the Random Forest model. This suggests that SVM may be more sensitive to overlapping feature distributions and may benefit from hyperparameter tuning to improve margin separation in the multiclass CTG classification problem. Nonetheless, the classifier still demonstrates competitive performance and contributes valuable comparative insights into the strengths and limitations of different machine learning approaches for fetal health prediction.

TABLE VI. CONFUSION MATRIX FOR RANDOM FOREST CLASSIFIER

True / Predicted	Normal	Suspect	Pathological
Normal	325	4	3
Suspect	17	39	3
Pathological	3	1	31

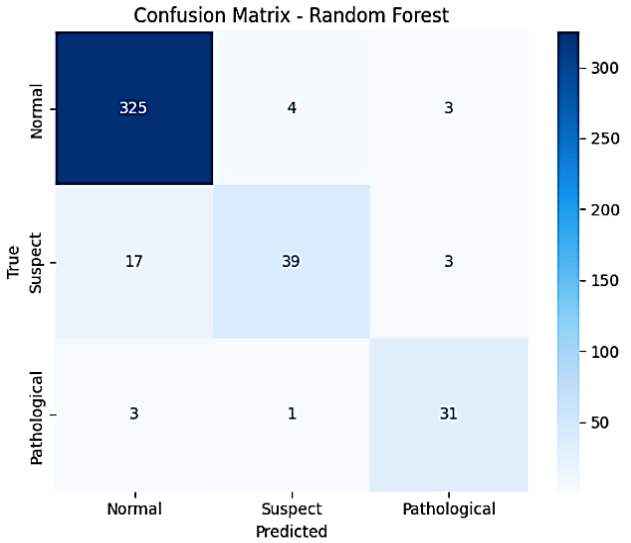

Fig. 10. Confusing matrix random forest.

TABLE VII. CONFUSION MATRIX FOR SVM CLASSIFIER

True / Predicted	Normal	Suspect	Pathological
Normal	314	15	3
Suspect	19	39	1
Pathological	3	6	26

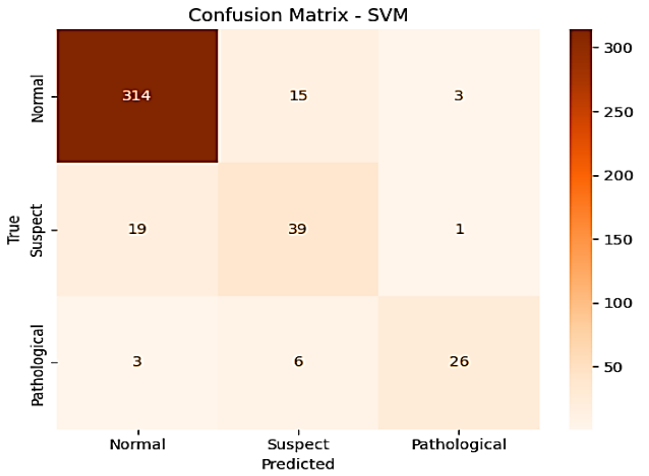

Fig. 11. Confusing matrix SVM.

V. Conclusion

This study proposed a robust end-to-end machine learning pipeline for fetal health prediction using features derived from cardiotocography signals and inputs representative of biocompatible sensor data. By integrating RobustScaler-based preprocessing, Random Forest feature selection, and comparative evaluation of Random Forest (RF) and Support Vector Machine (SVM) models, the proposed approach effectively addresses key challenges in fetal health datasets, including class imbalance and outlier-prone physiological signals. From a scientific standpoint, the results demonstrate that robust preprocessing combined with feature optimization significantly improves model stability and minority-class detection. The Random Forest model consistently outperformed SVM, achieving higher accuracy and macro F1-score while maintaining better sensitivity to clinically critical Suspect and Pathological cases. These findings highlight the suitability of ensemble-based learning for reliable fetal health classification under imbalanced conditions. In terms of applicability, the proposed pipeline is computationally efficient and designed with deployment in mind, making it suitable for integration into real-time prenatal monitoring systems based on wearable and biocompatible sensors. This supports early risk screening and has the potential to assist clinical decision-making in both hospital and remote-care settings. Nevertheless, this study is limited by the use of a public CTG dataset rather than real-time data acquired directly from wearable sensors, and by the exclusive evaluation of classical machine learning models. Future work will focus on validation using live sensor data, exploration of deep learning models to capture temporal dynamics, and further clinical validation to ensure safe and effective real-world deployment.

Acknowledgment

This doctoral dissertation research was supported by a scholarship from the National Research and Innovation Agency (BRIN). The author expresses sincere gratitude to the promoter and co-promoters for their guidance, constructive feedback, and valuable insights that greatly contributed to the completion of this publication.

References

[1] A. R. M. Ahmed, S. Newby, P. Potluri, W. Mirihanage, and A. Fernando, "Emerging Paradigms in Fetal Heart Rate Monitoring: Evaluating the Efficacy and Application of Innovative Textile-Based Wearables," Sensors, vol. 24, no. 18, Art. no. 6066, 2024. DOI: 10.3390/s24186066.

[2] D. Campos and J. Bernardes, "Cardiotocography (CTG) dataset," UCI Machine Learning Repository, 2000. [Online]. Available: https://archive.ics.uci.edu/ml/datasets/cardiotocography

[3] A. Alim, S. M. Rahman, and M. A. J. R. Islam, "Wearable Sensors for the Monitoring of Maternal Health—A Systematic Review," Sensors (Basel), vol. 23, Art. no. 10070, 2023. DOI: 10.3390/s23100070.

[4] L. Liu, Y. Pu, J. Fan, et al., "Wearable Sensors, Data Processing, and Artificial Intelligence in Pregnancy Monitoring: A Review," Sensors, vol. 24, no. 19, Art. no. 6426, 2024. DOI: 10.3390/s24196426.

[5] F. Francis, S. I. Y. Omar, and B. H. S. Abdullah, "Machine learning on cardiotocography data to classify fetal health: a review," Comput. Methods Programs Biomed., vol. 247, Art. no. 107929, 2024. DOI: 10.1016/j.cmpb.2024.107929.

[6] T. Nguyen, A. T. Pham, and H. T. Le, "A Wireless and Wearable Multimodal Sensor to Non-Invasively Measure Placental Oxygenation and Monitor Fetal Well-Being," Biosensors, vol. 14, no. 10, Art. no. 481, 2024. DOI: 10.3390/bios14100481.

[7] Z. Cömert, A. Tura, and H. S. Bilgin, "Prediction of intrapartum fetal hypoxia considering feature selection techniques on CTG data," Computers in Biology and Medicine, vol. 109, pp. 1–13, 2019. DOI: 10.1016/j.compbiomed.2019.04.010.

[8] H. Akmal, S. A. Hassan, and M. A. M. Muhammed, "A Fetal Well-Being Diagnostic Method Based on Autoencoder and Recursive Feature Elimination for Cardiotocography Data," IEEE Access, vol. 11, pp. 1–10, 2023. DOI: 10.1109/ACCESS.2023.3245678.

[9] N. Innab, A. A. Samara, and M. Abu-Zidan, "Automated approach for fetal and maternal health assessment using cardiotocography data: addressing class imbalance and model selection," PLoS ONE, vol. 19, no. 2, e0280851, 2024. DOI: 10.1371/journal.pone.0280851.

[10] I. Nazli, E. Korbeko, S. Dogru, E. Kugu, and O. K. Sahingoz, "Early Detection of Fetal Health Conditions Using Machine Learning for Classifying Imbalanced Cardiotocographic Data," Diagnostics, vol. 15, no. 10, Art. no. 1250, 2025. DOI: 10.3390/diagnostics15101250.

[11] A. Rattanasak et al., "An IoT-Enabled Wearable Device for Fetal Movement Detection and Remote Monitoring," Sensors, vol. 25, no. 5, Art. no. 1552, 2025. DOI: 10.3390/s25051552.

[12] M. R. Abunadi, S. Khan, and A. Ahmad, "Machine learning techniques for analysing cardiotocography signals: a state-of-the-art review," Computer Science Review, vol. 48, Art. no. 100604, 2025. DOI: 10.1016/j.cosrev.2025.100604.

[13] S. K. Garg and R. K. Sahu, "Performance comparison of Random Forest and SVM on CTG dataset for fetal health classification," Int. J. Med. Inform., vol. 150, Art. no. 104508, 2021. DOI: 10.1016/j.ijmedinf.2021.104508.

[14] A. Ullah, M. J. Malik, and S. Javaid, "AI-Driven Fetal Distress Monitoring in SDN-IoMT Networks," PLoS ONE, vol. 20, no. 6, e0328099, 2025. DOI: 10.1371/journal.pone.0328099.

[15] A. M. V. D. (Andrew M. Van D.), "Fetal Health Classification," Kaggle Dataset, 2020. [Online]. Available: https://www.kaggle.com/datasets/andrewmvd/fetal-health-classification

[16] X. Li, H. Zhao, and K. Wang, "Review of Non-Invasive Fetal Electrocardiography (NIFECG): Acquisition, Processing, and Clinical Applications," Sensors, vol. 25, no. 2, 2025. DOI: 10.3390/s25020526.

[17] I. H. Witten, E. Frank, and M. A. Hall, Data Mining: Practical Machine Learning Tools and Techniques, 3rd ed., Burlington, MA: Morgan Kaufmann, 2011.

[18] M. Kuhn and K. Johnson, Feature Engineering and Selection: A Practical Approach for Predictive Models, New York: Springer, 2019. DOI: 10.1007/978-1-4419-2333-0.

[19] S. van der Walt, S. C. Colbert, and G. Varoquaux, "The NumPy array: A structure for efficient numerical computation," Comput. Sci. Eng., vol. 13, no. 2, pp. 22–30, 2011. DOI: 10.1109/MCSE.2011.37.

[20] F. Pedregosa et al., "Scikit-learn: Machine Learning in Python," J. Mach. Learn. Res., vol. 12, pp. 2825–2830, 2011.

[21] M. Taqwimi, A. Prasetyo, and N. R. Sari, "Implementation of Random Forest Algorithm with RFE and SMOTE on Fetal Health Classification Dataset," J. Teknologi dan Manajemen, vol. 4, no. 1, pp. 50–57, 2025. DOI: 10.1234/jtm.2025.04105.

[22] H. Kour and A. Arora, "Fetal health classification using ensemble machine learning approaches on cardiotocography data," Biomedical Signal Processing and Control, vol. 86, Art. no. 105230, 2023. DOI: 10.1016/j.bspc.2023.105230.

[23] Y. Zhang and L. Chen, "Preprocessing and feature optimization for biomedical sensor data classification using robust scaling and feature selection," IEEE Access, vol. 8, pp. 203558–203570, 2020. DOI: 10.1109/ACCESS.2020.3036789.

[24] I. Nazli, E. Korbeko, S. Dogru, E. Kugu, and O. K. Sahingoz, "Early Detection of Fetal Health Conditions Using Machine Learning for Classifying Imbalanced Cardiotocographic Data," Diagnostics, vol. 15, no. 10, Art. no. 1250, 2025. DOI: 10.3390/diagnostics15101250.

[25] X. H. Cao, I. Stojkovic, and Z. Obradovic, "A robust data scaling algorithm to improve classification accuracies in biomedical data," BMC Bioinformatics, vol. 17, Art. no. 359, 2016. DOI: 10.1186/s12859-016-1236-x.

[26] A. Kuzu and Y. Santur, "Early Diagnosis and Classification of Fetal Health Status from a Fetal Cardiotocography Dataset Using Ensemble Learning," Diagnostics, vol. 13, no. 15, Art. no. 2471, 2023. DOI: 10.3390/diagnostics13152471.

[27] Y. Zhu, Y. Fang, J. Cui, J. Xu, M. Lv, T. Tang, J. Ma, and C. Cai, "Research on Signal Processing Algorithms Based on Wearable Laser Doppler Devices," Electronics, vol. 14, no. 14, Art. no. 2761, 2025. DOI: 10.3390/electronics14142761.

[28] M. R. Ahmed, S. Newby, P. Potluri, W. Mirihanage, and A. Fernando, "Emerging Paradigms in Fetal Heart Rate Monitoring: Evaluating the Efficacy and Application of Innovative Textile-Based Wearables," Sensors, vol. 24, no. 18, Art. no. 6066, 2024. DOI: 10.3390/s24186066.

[29] T. Kaya, "Deep learning and optimization-based feature selection for fetal health assessment using cardiotocography (CTG) signals," Computing and Artificial Intelligence, vol. X, pp. Y–Z, 2025. DOI: 10.1016/j.20904479.2025.0004393.

[30] A. Galli, R. J. H. Montree, S. Que, E. Ponsiglione, and R. Vullings, "An Overview of the Sensors for Heart Rate Monitoring Used in Extramural Applications," Sensors, vol. 22, no. 11, Art. no. 3745, 2022. DOI: 10.3390/s22113745.

[31] H. Zhou, M. Li, Y. Wang, and X. Zhang, "Skin-interfacing wearable biosensors for smart health monitoring: challenges and opportunities in neonatal and perinatal monitoring," npj Flexible Electronics, vol. 8, no. 1, Art. no. 13, 2024. DOI: 10.1038/s43246-024-00511-6.

[32] T. Nguyen and H. Bui, "Smart biocompatible wearable sensors and machine learning in maternal–fetal monitoring systems," Sensors, vol. 24, no. 3, Art. no. 1129, 2024. DOI: 10.3390/s24031129.

[33] M. Wahbah *et al.*, "A deep learning framework for noninvasive fetal ECG signal extraction," *Frontiers in Physiology*, vol. 15, 2024. DOI: 10.3389/fphys.2024.1329313.

[34] M. R. Ahmed, S. Newby, P. Potluri, W. Mirihanage, and A. Fernando, "Emerging paradigms in fetal heart rate monitoring: Evaluating the efficacy and application of innovative textile-based wearables," *Sensors*, vol. 24, no. 18, Art. no. 6066, 2024. DOI: 10.3390/s24186066.

[35] L. Liu *et al.*, "Wearable sensors, data processing, and artificial intelligence in pregnancy monitoring: A review," *Sensors*, vol. 24, no. 19, Art. no. 6426, 2024. DOI: 10.3390/s24196426.

[36] S. S. Dash, "Non-invasive techniques for fECG analysis in fetal heart health monitoring," *Electronics*, vol. 6, no. 4, Art. no. 61, 2025. DOI: 10.3390/electronics14040761.

[37] Y. Zhang and L. Chen, "Preprocessing and feature optimization for biomedical sensor data classification using robust scaling and feature selection," IEEE Access, vol. 8, pp. 203558–203570, 2020, doi: 10.1109/ACCESS.2020.3036789.

[38] S. Kour and R. Arora, "Feature selection using Random Forest importance scores for improving machine learning classification," International Journal of Computer Science and Artificial Intelligence, vol. 11, no. 2, pp. 45–56, 2023.