Indonesian Journal of Computer Science
ISSN 2549-7286 (*online*)
Jln. Khatib Sulaiman Dalam No. 1, Padang, Indonesia
Website: ijcs.stmikindonesia.ac.id | E-mail: ijcs@stmikindonesia.ac.id

Comprehensive Classification of Fetal Health Using Cardiotocogram Data Based on Machine Learning

Ahmed A. H. Alkurdi[1,2], Adnan Mohsin Abdulazeez[3]
ahmed.alaa@auas.edu.krd, adnan.mohsin@dpu.edu.krd
[1]Information Technology Department, Technical College of Informatics, Akre University for Applied Sciences, Duhok, Iraq
[2]Department of Information Technology, Technical College of Duhok, Duhok Polytechnic University, Duhok, KRG-Iraq
[3]Technical College of Engineering-Duhok, Duhok Polytechnic University, Duhok, KRG-Iraq

Article Information

Submitted : 23 Jan 2024
Reviewed: 26 Jan 2024
Accepted : 10 Feb 2024

Keywords

Machine Learning, Classification, Fetal Health, Cardiotocogram, Medical

Abstract

In the realm of obstetrics, the evaluation of fetal health remains a paramount yet challenging endeavor. Traditional approaches, such as electronic fetal monitoring (EFM), despite their widespread adoption, continue to grapple with uncertainties regarding their impact on neonatal outcomes and the reduction of emergency cesarean deliveries. This ambiguity is compounded by a prevailing confusion within the obstetric community about interpreting fetal heart rate patterns, often leading to inconsistent and subjective assessments. Addressing these complexities, our study presents an innovative machine learning-based techniques for the comprehensive classification of fetal health using cardiotocogram (CTG) data, offering a more objective and nuanced alternative to conventional methods. The core of our proposed solution is a novel model employing a sophisticated ensemble of machine learning classifiers, including Multi-Support Vector Machine (Multi-SVM), Decision Tree, Random Forest with Hyperparameter Tuning, XGBoost, and Neural Networks. This model is unique in its application, processing datasets in four different forms: raw datasets, datasets processed with MinMaxScaler, datasets subjected to feature selection using SelectKBest, and a combination of MinMaxScaler processing and SelectKBest feature selection. Such meticulous preprocessing, encompassing normalization and feature selection, is pivotal in ensuring equitable contribution from each feature, thereby optimizing the model's learning process and predictive accuracy. The effectiveness of our model is rigorously evaluated using a dataset comprising 2126 individual records from CTG exams, classified by specialist obstetricians into three types: Normal, Suspect, and Pathological. These records are exhaustively analyzed using various metrics, including Accuracy, Precision, Recall, F1-Score, ROC AUC, and Confusion Matrix. Among the classifiers, XGBoost emerged as the most proficient, consistently outperforming others across multiple metrics. This indicates its superior ability to accurately identify and categorize the different states of fetal health. Our findings thus underscore the significant promise of machine learning in revolutionizing fetal health monitoring, offering a more reliable, objective, and comprehensive method for assessing fetal well-being, with profound implications for prenatal care and clinical decision-making.

Attribution-ShareAlike 4.0 International License

A. Introduction

The exploration and debate surrounding the assessment of fetal health, particularly through electronic fetal monitoring (EFM), constitute a prominent area of inquiry in obstetrics. Despite advancements in technology, the efficacy of EFM in enhancing neonatal outcomes and diminishing the necessity for emergency cesarean deliveries continues to be a subject of uncertainty [1]. A comprehensive systematic review and network meta-analysis, encompassing 33 trials with a total of 118,863 patients, compared various fetal surveillance methodologies, including intermittent auscultation and cardiotocography (CTG). This analysis revealed that intermittent auscultation may reduce the rate of emergency cesarean deliveries without adversely affecting neonatal and maternal outcomes [2]. Nevertheless, it was observed that newer surveillance methods, such as fetal heart electrocardiography, did not significantly lower the rates of emergency cesarean deliveries during labor or decrease newborn morbidity when used with simple intermittent auscultation [3].

Further complicating the landscape, a critical commentary on EFM highlighted the prevailing confusion within the obstetric community regarding the interpretation of fetal heart rate (FHR) patterns and the associated challenges in responding effectively [3]. The conventional three-category classification system for FHR patterns (Categories I-III) has been subjected to criticism for its lack of comprehensive vetting and insufficient consideration of essential physiological principles, including fetal behavior and the potential for detecting neurological injury. The prevailing focus on fetal acidemia as the primary indicator of fetal distress has been challenged, suggesting the need for a more expansive understanding of fetal-maternal physiology and advocating for a less defensive posture towards monitoring. Additionally, the persistent concern of obstetrical malpractice, often centered around the interpretation of EFM tracings, introduces a further layer of complexity to this issue [4].

The discipline of obstetrics is presently confronted with persistent obstacles in the identification and implementation of effective methodologies for the surveillance and interpretation of fetal well-being. The application of machine learning techniques in the classification of fetal well-being by analyzing cardiotocogram data offers a novel approach to gain deeper insights into fetal health. This possesses the capacity to surmount specific limitations and challenges linked to conventional methodologies.

This paper aims to enhance the assessment of fetal health by employing a range of sophisticated machine learning classifiers to analyze Cardiotocogram (CTG) data. The objective is to develop a model that surpasses traditional electronic fetal monitoring by eliminating the prevalent ambiguities and subjective interpretations, thereby offering a more reliable, objective, and comprehensive evaluation of fetal well-being.

This paper is systematically organized for clarity and depth. Section 1, the Introduction, delves into the difficulties and challenges prevalent in the realm of fetal health assessment. In Section 2, the Literature Review, a thorough examination of existing studies pertinent to the application of machine learning in fetal health monitoring and classification is presented. The methodological framework is articulated in Section 3, Proposed Model, which is subdivided to expound on preprocessing techniques and classifier methodologies. Section 4, Dataset, meticulously details the composition and classification of the data utilized. The Evaluation Metrics, encompassed in Section 5, elaborate on the criteria and methods employed to assess model performance. Section 6, Results and Discussion, offers an insightful analysis of the model's efficacy. Comparative insights with extant research are drawn in Section 7, while Section 8, the Conclusion, encapsulates the study's contributions and implications in the domain of fetal health monitoring through machine learning paradigms.

B. Literature Review

The literature review integrates a diverse array of studies, each contributing significantly to the landscape of fetal health monitoring using deep learning, machine learning and artificial intelligence.

Several studies focus on employing neural networks and other computational techniques for FHR classification. The study achieved by [5] underscores the effectiveness of neural networks in FHR pattern classification, particularly in intrauterine growth restriction cases. Similarly, another study fulfilled by [6] demonstrates the superiority of CNNs over traditional methods like SVM and MLP in classifying high one-dimensional FHR records. The advancement of neural networks is further emphasized in the study [7], which employs Continuous Wavelet Transform and 2D CNN for predicting fetal acidemia.

Another significant theme is the application of machine learning algorithms in CTG data classification. [8] leverages Bagging with Random Forest, achieving an accuracy of 99.02%. while a study conducted by [9] examines algorithms like XGB, SVM, KNN, and LSTM, providing a comprehensive performance comparison. Authors [10] developed a model with 99.59% accuracy using SVM and oversampling techniques, accompanied by an analysis of feature importance.

The importance of ensemble learning techniques is further highlighted in a research paper concluded by [11], which achieved remarkable accuracy levels above 99.5%. This approach significantly improves prediction performance, demonstrating the potential of combining multiple classifiers in fetal health diagnostics.

The researchers conducted a comparative analysis of three supervised learning algorithms: Decision Tree (DT), Support Vector Machine (SVM), and Naïve Bayes (NB), in the context of heart disease prediction in [12]. The study concluded that the Decision Tree algorithm demonstrated superior performance in terms of predictive accuracy and exhibited a more efficient training time relative to SVM and Naïve Bayes. This finding underscores the potential applicability of the Decision Tree method in the development of real-time clinical decision support systems, thereby advancing the utility of data mining techniques in the healthcare domain for early disease detection and enhancing patient care outcomes.

Several studies concentrate on the importance of feature selection and classification during different stages of labor. Authors [13] highlight the need for robust classification models in their paper, considering the unique dynamics of FHR during different labor stages. While [14] focuses on the use of ML algorithms like SVM, RF, MLP, and KNN for accurate fetal health classification. Furthermore,

[15] introduces a novel feature selection method using a crowding distance-based multi-objective genetic algorithm.

In a research paper by [16], Long Short-Term Memory networks are utilized for segmental classification of FHR. Also [17] presents a new method for automated FHR analysis using a weighted median filter baseline. On the other hand, [18] introduces methods for classifying FHR signals using generative models and Bayesian theory.

The study undergone by [19] applies deep learning models like MCNN and Stacked MCNN to analyze CTG data, and [20] introduces methods to detect periodic changes in FHR. particularly accelerations, decelerations, and the Sinusoidal Heart Rate (SHR) pattern, with the goal of enhancing automated clinical decision support systems. [21] employs Sparse SVM classification for fetal acidosis detection during delivery. The sparse SVM is utilized to select a limited set of relevant features from a comprehensive set including clinical, frequency domain, scaling, and multifractal features, computed from a large database of 1,288 subjects. This approach allows for efficient fetal acidosis detection with improved sensitivity and specificity compared to traditional clinical practice.

Another study performed by [22], investigates the effectiveness of ML techniques in identifying high-risk fetuses using CTG data. The study used data from 2126 pregnant women in their third trimester, obtained from the University of California Irvine Machine Learning Repository. This data included various attributes of fetal heart rate (FHR) and uterine contractions (UCs). Ten different ML classification models were trained to predict normal, suspect, and pathological fetal states based on CTG data. The study concluded that the XGBoost-based classification model had the highest prediction accuracy for adverse fetal outcomes. Finally, [23] in their study explore the effectiveness of combining conventional and nonlinear features for improving FHR classification accuracy. Using a database of 217 FHR records, the study demonstrated that the inclusion of nonlinear features such as Lempel Ziv complexity, Sample entropy, and fractal dimension (estimated by Higuchi method) significantly improves the accuracy of FHR classification.

In summary, these studies collectively illustrate the evolving landscape of fetal health monitoring, where various machine learning and artificial intelligence technologies are being increasingly integrated to enhance the accuracy and reliability of fetal health classification, promising significant improvements in prenatal care and perinatal mortality reduction.

C. Proposed Model

The projected model for fetal heart classification is developed to assess the performance of Multi-Support Vector Machine (Multi-SVM), Decision tree, Random Forest with Hyperparameter Tuning, XGBoost and Neural Networks. The Dataset is fed into the aforementioned classifiers in four forms as mentioned below.

1- Raw Dataset without Preprocessing and Feature Selection.
2- Dataset after preprocessing with MinMaxScaler.
3- Dataset after feature selection using SelectKBest.
4- Dataset after preprocessing using MinMaxScaler and feature selection using SelectKBest.

Afterwards the data is split and run through the classifiers.

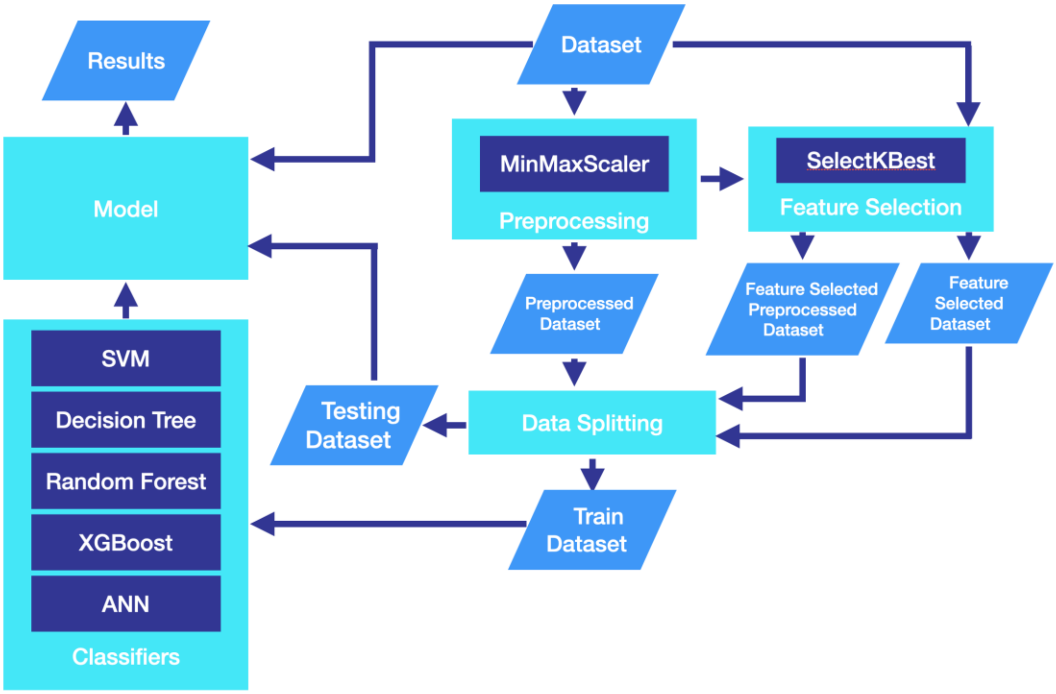

Figure 1. Proposed Model

1. Preprocessing: Normalization

Normalization and scaling are essential preprocessing steps in machine learning, especially for models dealing with complex data such as fetal health classification using cardiotocograms. Normalization adjusts the scale of data attributes, ensuring that each feature contributes equally to the model's learning process [24], [25]. This is crucial because features with larger numerical ranges could disproportionately influence the model's performance. Techniques like MinMaxScaler, StandardScaler, and RobustScaler are employed to rescale data to a specific range or to adjust it based on statistical properties such as mean and standard deviation [26], [27].

The choice of normalization method can considerably impact the model's ability to learn from the data. For instance, the MinMaxScaler linearly transforms features to fall within a predefined range, making it less susceptible to outliers. These methods not only aid in reducing the effects of outliers but also ensure that the significant relationships inherent in the original data are preserved [28], [29].

2. Feature Selection

Feature selection is a critical aspect of preparing datasets for machine learning models, especially in complex fields such as medical analysis or time series classification. It involves identifying and extracting the most significant features from a large dataset, which can greatly enhance the performance of predictive models [30] [31]. Feature selection methods are broadly categorized into lexicon-based methods, which require human input, and statistical methods, which automatically provide markers. Among the statistical methods, Mutual Information (MI) stands out as a particularly effective filter-based approach. It quantifies the importance of feature information by evaluating the correlation between selected features and class labels, assuming that features with a strong correlation will improve classification performance [32], [33].

In supervised classification of multivariate time series, such as in fetal health classification using cardiotocogram data, mutual information is used to find the relevance of each feature subset. This is especially useful when the features are time series, as it involves adapting nonparametric mutual information estimators for time series scenarios [34]. The goal is to select time series subsets that maximize a score function, focusing on those that share high information with the classification variable and are less redundant with each other [35], [36]. This method of feature selection using mutual information can significantly reduce the number of features while maintaining or increasing classification accuracy.

3. Classifiers

The suggested model proposed in this study relies on a diverse range of classifiers. The machine learning methods examined in this paper encompass the Multi-Support Vector Machine, Decision tree, Random Forest with hyperparameter tweaking, Gradient Boosting Machines, and Neural Network.

MultiSVM

The Multi-Support Vector Machine (Multi-SVM) is a machine learning methodology that improves the capabilities of the traditional Support Vector Machine (SVM) by enabling it to efficiently handle multi-class classification problems [37], [38]. This is particularly relevant in situations such as the classification of fetal health using cardiotocogram data, where the main objective is to classify different health states. The Support Vector Machine (SVM) is a commonly used binary classifier that separates data into two discrete classes. The operational method entails the identification of a hyperplane within a space of n dimensions that effectively discriminates between the different classes. In contrast, a Multi-SVM exhibits the capacity to efficiently handle numerous classes, making it highly suitable for classification problems of a more complex type [39], [40].

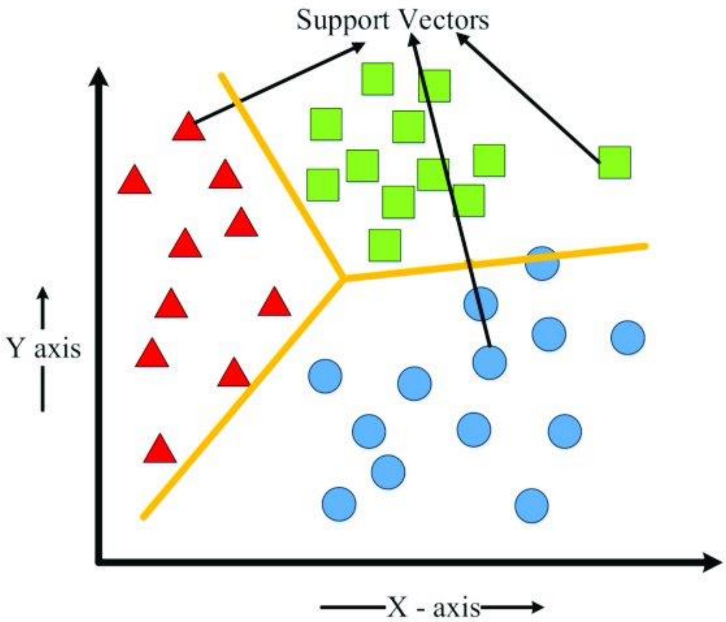

Figure 2. Multi-SVM [41]

Numerous studies conducted across diverse fields, including medicine and agriculture, have provided evidence supporting the efficacy of Multi-SVM in effectively addressing multi-class scenarios. The data classification or identification process involves the utilization of a support-vector engine to assign labels to instances, drawing from a variety of factors. In agricultural contexts, Multi-SVM has been employed to classify various crop diseases and flower varieties, hence showcasing its versatility across a range of domains [42], [43]. The utilization of Multi-SVM in the classification of fetal health based on cardiotocogram data entails the categorization of different fetal health states. This approach capitalizes on the algorithm's proficiency in effectively handling intricate multi-class situations.

Decision Tree

The Decision Tree classifier, specifically the C4.5 variant, holds considerable importance in the domain of machine learning for the categorization of medical data [44]. It finds use in several areas, such as the classification of fetal health using cardiotocogram data. The hierarchical classifier being discussed is an expansion of the ID3 (Iterative Dichotomiser 3) technique, which was first devised by Quinlan in 1993. The categorization of datasets performed by this tool is widely recognized for its effectiveness and has become a common practice in the field of supervised classification. The C4.5 algorithm functions by initiating the construction of the decision tree from the root node and thereafter expanding it in a top-down manner. It assesses each attribute to decide how correctly it can classify the training samples [45], [46].

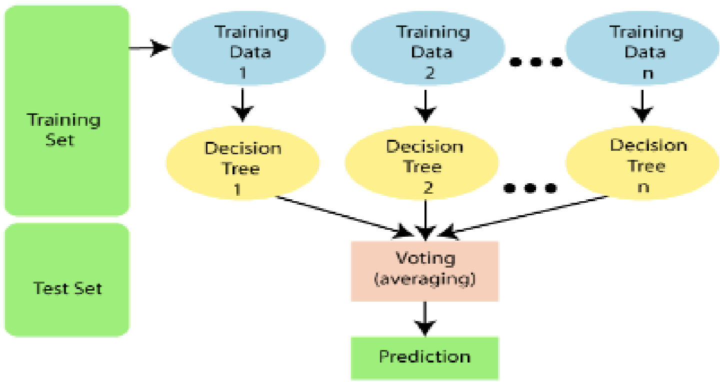

Figure 3. Decision Tree [47]

Decision tree classifiers, specifically the C4.5 algorithm, have demonstrated effectiveness in practical applications, particularly in the classification of various forms of cancer such as hepatocellular carcinoma and metastatic carcinoma . The procedures in question have undergone validation and have been subject to favorable comparisons with other contemporary techniques. The effective recognition of various diseases by the decision tree classifier holds significant importance in facilitating appropriate medical diagnosis and subsequent treatment. Consequently, it serves as a vital tool in the realm of medical data analysis and classification tasks [48].

Random Forest

The Random Forest algorithm is a widely recognized ensemble classification method that has garnered significant interest across multiple disciplines. One such application is medical data analysis, specifically in the context of fetal health classification utilizing cardiotocogram data. The technique integrates numerous decision trees in order to enhance the precision of forecasting, rendering it very proficient in handling extensive datasets and intricate classification challenges. The Random Forest algorithm combines predictions from many decision trees, resulting in improved accuracy and greater performance [49].

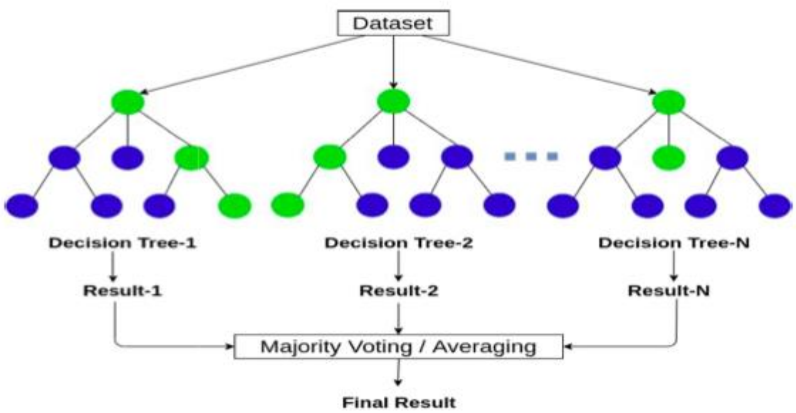

Figure 4. Random Forest [50]

One notable advantage of the Random Forest algorithm is its capacity to mitigate overfitting, a prevalent challenge encountered in machine learning models, particularly when confronted with extensive and noisy datasets. The aforementioned capabilities, when coupled with its expedited training duration in comparison to individual classifiers such as Decision Trees and Support Vector Machines, renders it a favored option among academics and practitioners. The combination of Random Forest and AdaBoost algorithms has demonstrated improved classification accuracy, rendering it a resilient and effective tool in the domain of machine learning for intricate classification tasks [51].

Hyperparameter tuning

hyperparameter tuning is an important aspect of optimizing machine learning models. Hyperparameters, distinct from model parameters, must be set before the training and considerably impact the model's performance on specific tasks. The aim is often to identify whether it's essential to tune a hyperparameter or if it can be safely set to a default value [52][53].

The study employs a methodology focusing on the non-inferiority test and tuning risk, which is the performance loss sustained when a hyperparameter is not adjusted but set to a default value. This approach involves determining reasonable default parameters and assessing whether leaving certain hyperparameters at these default values is comparable to tuning them, sometimes even outperforming the tuned models under limited iterations .

To establish default values, a heuristic procedure is used, analyzing subsets of empirical performance measurements representing good performance. The process identifies the most frequently occurring parameter setting within these subsets, considering it a good candidate for a default value. This methodology is critical for cases where computational resources limit the viability of large scale hyperparameter tuning [53]–[55].

Gradient Boosting Machines (GBMs)

Gradient Boosting Machines (GBMs) are a very effective category of machine learning algorithms that have demonstrated exceptional efficacy in addressing both classification and regression tasks. These algorithms function as an ensemble learning technique, employing a blend of additive models commonly known as weak learners [56], [57]. The primary advantage of Gradient Boosting Machines (GBMs) resides in their capacity to iteratively acquire knowledge from past misclassifications, progressively creating a prediction model that is more resilient and precise. The iterative approach described above incorporates the practice of feature selection, a technique that improves the performance of the model by selecting and prioritizing the most pertinent elements within the input data. In practical applications, these techniques are commonly created utilizing libraries like as scikit-learn, which provide a diverse selection of algorithms for classification, regression, and clustering [58].

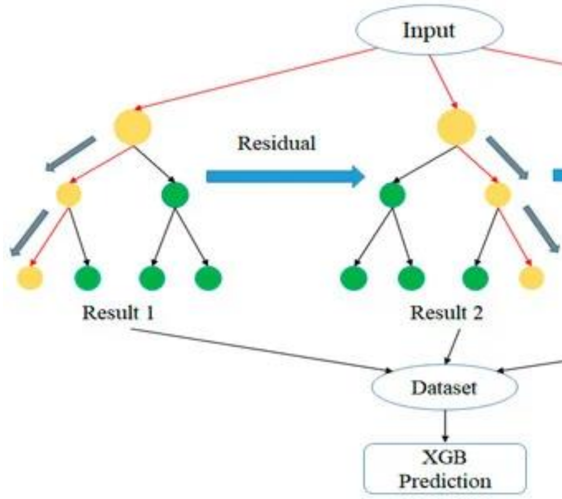

Figure 5. Gradient Boosting Machines (GBMs) [59]

One notable utilization of Gradient Boosting Machines (GBMs) can be observed within the domain of credit card transaction fraud detection. The integration of a Bayesian-based hyperparameter optimization technique in an approach called optimized light gradient boosting machine (OLightGBM) allows for the precise tuning of model parameters. The optimization of credit card fraud detection is of utmost importance due to the intricate and multifaceted nature of this issue[60]. In several real-world scenarios, the OLightGBM algorithm has exhibited notable efficacy in comparison to alternative machine learning methodologies, specifically in relation to measures of accuracy, precision, and AUC. The evidence of efficacy is derived from empirical investigations utilizing authentic datasets of credit card transactions, encompassing instances of both fraudulent and authorized transactions [61].

Neural Networks

Neural networks are of utmost importance in the domain of machine learning and artificial intelligence. These networks are specifically engineered to replicate the cognitive capabilities of the human brain's neural system, rendering them very proficient in managing intricate tasks that pose difficulties for conventional expert systems [62], [63]. Artificial neural networks (ANNs) are widely recognized for their exceptional capacity to store and handle ambiguous information. The applications of these technologies encompass a diverse array of fields, such as pattern recognition, signal processing, intelligent control, and optimization. The ability of artificial neural networks (ANNs) to adapt and learn from data renders them an essential element in the development of intelligent systems [64], [65].

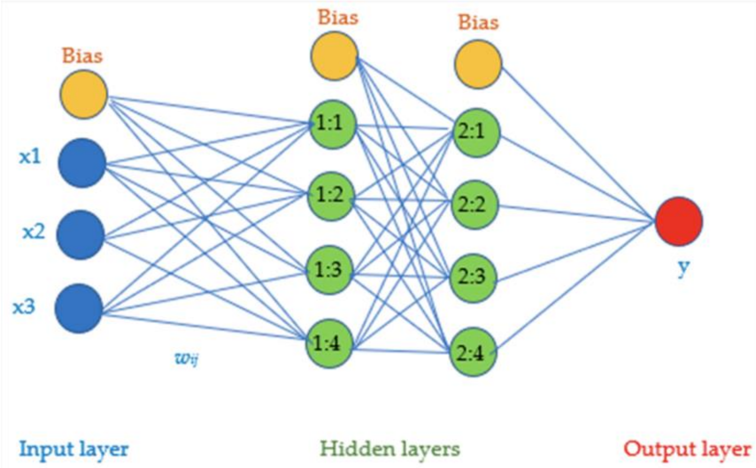

Figure 6. Neural Network [66]

Several types of Artificial Neural Networks (ANNs) have acquired significance within the field, mostly due to their distinct uses and capabilities. Deep Neural Networks (DNNs), Convolutional Neural Networks (CNNs), and Recurrent Neural Networks (RNNs) are extensively studied and widely utilized models in the field of neural networks [67]. Artificial neural networks, which are frequently trained using the error back-propagation (BP) method, are generally acknowledged for their efficacy in many tasks such as handwriting identification, speech recognition, product inspection, fault detection, and medical diagnosis. Nevertheless, the rate at which the backpropagation (BP) process, commonly utilizing the gradient descent method, undergoes training can provide a constraint. The performance of the network is substantially influenced by its structure, underscoring the importance of meticulous design and optimization in order to attain optimal outcomes [68], [69].

D. Dataset

The dataset [70] utilized for the study comprises 2126 individual records extracted from Cardiotocogram (CTG) exams, which have been meticulously classified by three expert obstetricians into three distinct categories: Normal, Suspect, and Pathological. This categorization reflects the varying degrees of fetal health as discerned through CTG, a non-invasive diagnostic tool that measures fetal heart rate (FHR), uterine contractions, and other vital parameters.

Each record in the dataset encapsulates a comprehensive array of features, including but not limited to:

1. The Baseline Fetal Heart Rate (FHR) is a characteristic that denotes the mean fetal heart rate observed within a specific time interval.
2. The phenomena of accelerations and decelerations are utilized to measure the temporary variations in fetal heart rate (FHR), with accelerations generally indicating a favorable physiological response and decelerations potentially indicating probable fetal discomfort.
3. Fetal Movements: The examination of fetal movements as a potential indicator of fetal well-being.

4. Measurement of uterine contractions in the maternal body provides valuable information regarding the frequency of contractions, which in turn offers insights into the progression of labor and the well-being of the fetus.
5. Assessment of Short-term and Long-term Variability: These characteristics evaluate the temporal intervals between consecutive heartbeats and the extended changes in fetal heart rate (FHR) observed over a period of time.

The characteristics of the dataset mostly consist of quantitative attributes, which include a combination of continuous and discrete variables. The authors present a comprehensive viewpoint on the health of the fetus, utilizing the diverse capabilities of cardiotocography (CTG) technology to observe and analyze intricate physiological data.

The extensive dataset at hand offers a distinctive prospect for the development of machine learning models that possess the ability to classify the status of fetal health. The classification into Normal, Suspect, and Pathological states not only facilitates a comprehensive comprehension of fetal well-being but also assists in the potential detection of fetal distress, so offering a substantial contribution to prenatal care and decision-making procedures. The dataset's extensive scope, which includes a diverse set of variables acquired from cardiotocography (CTG), renders it an indispensable asset for obstetric research, specifically in the field of fetal monitoring and health evaluation.

E. Evaluation Metrics

The model performance is measured for each of the dataset forms according to Accuracy, Precision, Recall, F1-Score, ROC AUC (Receiver Operating Characteristic Area Under the Curve), and Confusion Matrix.

$$\text{Accuracy} = \frac{(True\ Positives + True\ Negatives)}{Total\ Samples} \tag{1}$$

$$\text{Precision} = \frac{True\ Positives}{True\ Positives + False\ Positives} \tag{2}$$

$$\text{Recall} = \frac{True\ Positives}{True\ Positives + False\ Negatives} \tag{3}$$

$$\text{F1 Score} = \frac{2 * (Precision * Recall)}{(Precision + Recall)} \tag{4}$$

$$\text{Specificity} = \frac{True\ Negative}{True\ Negative + False\ Positive} \tag{5}$$

The Receiver Operating Characteristic (ROC) curve and its associated metric, the Area Under the Curve (AUC), serve as pivotal tools in evaluating the diagnostic ability of binary classifiers. The ROC curve, a graphical representation, plots the True Positive Rate (TPR) against the False Positive Rate (FPR) across varying threshold settings, providing insight into the trade-off between sensitivity (recall) and specificity. The AUC, a scalar value ranging from 0 to 1, quantifies the overall ability of the model to discriminate between the positive and negative classes. A receiver operating characteristic (ROC) curve with an area under the curve (AUC) value of 1 signifies optimal discrimination, indicating that the model can perfectly distinguish between positive and negative instances. Conversely, an AUC value of 0.5 indicates a performance equivalent to random chance, where the model's predictive ability is no better than a random guess. The aforementioned metric holds significant value in the evaluation of classifiers that encounter class imbalance, since it provides an assessment that is independent of the threshold and remains unbiased by the distribution of classes.

The confusion matrix holds significant importance in statistical classification as it provides a visual and quantitative representation of the performance of a classification model. The confusion matrix is a square matrix that depicts the tally of correct and incorrect guesses, juxtaposing them against the actual numbers. The matrix typically comprises four components, specifically True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN). These several measurements, including as accuracy, precision, recall, and the F1 score, collectively contribute to a full evaluation of the given elements. The organization of the matrix enables the detection of errors committed by the classifier, such as its tendency to incorrectly categorize one class as another. Consequently, this aids in directing subsequent model refinement and offers significant observations regarding the performance of the model across several categories, especially in datasets characterized by imbalanced distributions.

F. Results and Discussion

This research study provides a thorough assessment of the utilization of various machine learning models on a dataset including 2126 individual records acquired from Cardiotocogram (CTG) testing. The obstetricians with expertise in the field classified the data into three distinct groups, namely Normal, Suspect, and Pathological. The classified records provided a comprehensive and extensive basis for assessing the effectiveness of each model in distinguishing between different fetal health problems. The following is a comprehensive exposition of the attained outcomes.

Table 1. Results Without preprocessing and feature selection

Model	*Accuracy*	*Precision*	*Recall*	*F1 Score*	*Confusion Matrix*	*ROC AUC*	*Specificity*
Support Vector Machine	0.857	0.842	0.857	0.844	[321, 10, 2] [36, 26, 2] [5, 6, 18]	0.934	0.964
Decision Tree	0.932	0.936	0.932	0.933	[314, 16, 3] [8, 55, 1] [1, 0, 28]	0.936	0.943
Random Forest	0.946	0.945	0.946	0.945	[326, 6, 1] [11, 51, 2] [1, 2, 26]	0.985	0.979
XGBoost	0.960	0.960	0.960	0.960	[325, 7, 1] [8, 55, 1] [0, 0, 29]	0.989	0.976
Neural Network	0.817	0.7884	0.8192	0.7616	[332, 0, 1] [60, 1, 3] [14, 0, 15]	0.8715	0.997

Table 2. Results after preprocessing without feature selection

Model	*Accuracy*	*Precision*	*Recall*	*F1 Score*	*Confusion Matrix*	*ROC AUC*	*Specificity*
Support Vector Machine	0.911	0.911	0.911	0.909	[321, 12, 0] [18, 46, 0] [3, 5, 21]	0.965	0.964
Decision Tree	0.918	0.920	0.918	0.919	[313, 17, 3] [14, 50, 0] [1, 0, 28]	0.912	0.940
Random Forest	0.953	0.952	0.953	0.952	[327, 5, 1] [9, 53, 2] [1, 2, 26]	0.986	0.982
XGBoost	0.960	0.960	0.960	0.960	[325, 7, 1] [8, 55, 1] [0, 0, 29]	0.989	0.976
Neural Network	0.890	0.890	0.873	0.879	[308, 23, 2] [13, 49, 2] [2, 5, 22]	0.953	0.925

Table 3. Results after feature selection without preprocessing

Model	*Accuracy*	*Precision*	*Recall*	*F1 Score*	*Confusion Matrix*	*ROC AUC*	*Specificity*
Support Vector Machine	0.883	0.874	0.883	0.875	[322, 9, 2] [29, 33, 2] [3, 5, 21]	0.936	0.967
Decision Tree	0.923	0.923	0.923	0.922	[317, 14, 2] [13, 48, 3] [1, 0, 28]	0.911	0.952
Random Forest	0.944	0.943	0.944	0.943	[323, 7, 3] [10, 53, 1] [1, 2, 26]	0.983	0.970
XGBoost	0.948	0.948	0.948	0.948	[323, 8, 2] [10, 53, 1] [1, 0, 28]	0.986	0.970
Neural Network	0.833	0.686	0.807	0.742	[329, 2, 2] [51, 10, 3] [12, 1, 16]	0.893	0.970

Table 4. Results after preprocessing and feature extraction

Model	*Accuracy*	*Precision*	*Recall*	*F1 Score*	*Confusion Matrix*	*ROC AUC*	*Specificity*
Support Vector Machine	0.901	0.901	0.901	0.900	[318, 14, 1] [200, 44, 0] [2, 5, 22]	0.964	0.955
Decision Tree	0.918	0.921	0.918	0.919	[313, 16, 4] [11, 50, 3] [1, 0, 28]	0.916	0.940
Random Forest	0.939	0.938	0.939	0.938	[323, 8, 2] [12, 51, 1] [1, 2, 26]	0.982	0.970

XGBoost	0.948	0.948	0.948	0.948	[323, 8, 2] [10, 53, 1] [1, 0, 28]	0.986	0.970
Neural Network	0.899	0.897	0.896	0.896	[317, 15, 1] [18, 44, 2] [2, 5, 22]	0.951	0.952

Among the models evaluated, XGBoost shown remarkable competency, surpassing other models in crucial measures including accuracy, precision, recall, and F1 score. This indicates its superior capability in correctly identifying the three categories of fetal health. The high ROC AUC value of XGBoost further reinforces its effectiveness in distinguishing between the categories with a reduced rate of false positives and false negatives.

The confusion matrix, integral to understanding the model's performance, offers an in-depth view of its classification accuracy. In the context of this study, the confusion matrix for each model elucidates how many CTG records were correctly classified into each of the three categories. For example, XGBoost's confusion matrix with high true positive rates for each category indicates its precision in correctly classifying cases as Normal, Suspect, or Pathological. Conversely, lower false positive and false negative rates suggest fewer instances of misclassifying a Normal condition as Pathological or missing a Pathological condition by classifying it as Normal. This precise classification is crucial in clinical settings, where accurate diagnosis can significantly impact patient care and outcomes.

The dataset's composition, with records categorized as Normal, Suspect, and Pathological, presented a realistic and challenging scenario for the models. The varied nature of the data, reflective of real-world clinical conditions, underscored the need for a model that is not only accurate but also capable of handling the distinctions and complexities inherent in medical diagnostics.

In conclusion, the study effectively demonstrates the potential of machine learning models, particularly XGBoost, in the field of fetal health monitoring using CTG data. The ability of these models to accurately classify and differentiate between various fetal health conditions holds significant promise for enhancing prenatal care and ensuring better outcomes. The use of such advanced analytical tools could potentially transform the landscape of fetal health monitoring, making it

G. Comparison With Other Studies

In this study, the performance of the machine learning models shows a distinct improvement in specific areas when directly compared with similar metrics available from previous studies.

Table 5. Comparison With Other Studies

Study	*Model*	*Accuracy*	*Precision*	*Recall*	*F1-Score*	*ROC AUC*	*Specificity*
This Study	**SVM**	**0.911**	**0.911**	**0.911**	**0.909**	**0.965**	**0.964**
	DT	**0.932**	**0.936**	**0.932**	**0.933**	**0.936**	**0.940**
	RM	**0.953**	**0.952**	**0.953**	**0.952**	**0.986**	**0.982**

Study	*Model*	*Accuracy*	*Precision*	*Recall*	*F1-Score*	*ROC AUC*	*Specificity*
	XGB	0.960	0.960	0.960	0.960	0.989	0.976
	NN	0.899	0.897	0.896	0.896	0.0951	0.925
2 *[6]*	SVM	0.79	NA	88.18	NA	NA	77.36
	MLP	0.85	NA	91.83	NA	NA	89.25
	CNN	0.93	NA	98.40	NA	NA	84.77
3 *[8]*	SVM	0.98	NA	NA	0.99	0.96	NA
	k-NN	0.97	NA	NA	0.98	0.97	NA
	ANN	0.98	NA	NA	0.98	0.99	NA
	RF	0.98	NA	NA	0.99	1.00	NA
	CART	0.98	NA	NA	0.99	0.94	NA
	C4.5	0.98	NA	NA	0.99	0.93	NA
	REP Tree	0.98	NA	NA	0.98	0.94	NA
	RT	0.98	NA	NA	0.98	0.96	NA
4 *[9]*	XGB	0.98	0.99	0.94	0.98	0.98	NA
	SVM	0.98	0.98	0.95	0.98	0.98	NA
	KNN	0.98	1.00	0.94	0.98	0.99	NA
	LGBM	0.99	1.00	0.97	0.99	0.99	NA
	RF	0.98	0.99	0.95	0.99	0.99	NA
	ANN	0.17	0.13	0.14	0.13	0.38	NA
	LSTM	0.34	0.34	1.00	0.17	0.50	NA
6 *[11]*	LR	0.99	0.99	1.00	0.99	NA	NA
	RF	0.99	0.99	0.99	0.99	NA	NA
	GB	1.00	1.00	1.00	1.00	NA	NA
	XGBoost	1.00	1.00	1.00	1.00	NA	NA
7 *[71]*	GB	0.96	0.96	0.91	0.95	0.99	NA
	CatBoost	0.96	0.96	0.90	0.95	0.99	NA
	LGBM	0.95	0.95	0.90	0.95	0.99	NA
	EGB	0.95	0.95	0.90	0.95	0.99	NA
	Cascade Forest	0.95	0.93	0.87	0.90	0.98	NA
	RF	0.94	0.94	0.86	0.94	0.98	NA
	ET	0.94	0.94	0.85	0.93	0.99	NA
	DT	0.91	0.91	0.85	0.91	0.88	NA
	LR	0.89	0.88	0.75	0.88	0.92	NA
	KNN	0.89	0.88	0.75	0.88	0.91	NA
	LDA	0.88	0.88	0.74	0.88	0.96	NA
	Ada Boost	0.88	0.87	0.74	0.87	0.87	NA
	Stacker Model	0.95	0.95	0.91	0.95	0.99	NA
	Blender Model	0.96	0.96	0.92	0.96	0.99	NA
8 *[13]*	MLP	0.93	0.93	0.93	NA	NA	0.96
	RF	0.97	0.97	0.96	NA	NA	0.98
	SVM	0.96	0.97	0.96	NA	NA	0.98
	Bagging	0.94	0.94	0.94	NA	NA	0.97

Study	Model	Accuracy	Precision	Recall	F1-Score	ROC AUC	Specificity
10 [15]	**LR**	**0.86**	**NA**	**NA**	**NA**	**NA**	**NA**
	SVM	**0.79**	**NA**	**NA**	**NA**	**NA**	**NA**
	KNN	**0.91**	**NA**	**NA**	**NA**	**NA**	**NA**
	XGBoost	**0.94**	**NA**	**NA**	**NA**	**NA**	**NA**
	DT	**0.92**	**NA**	**NA**	**NA**	**NA**	**NA**
	RF	**0.92**	**NA**	**NA**	**NA**	**NA**	**NA**
	GNB	**0.83**	**NA**	**NA**	**NA**	**NA**	**NA**
11 [16]	**SVM Voting**	**0.81**	**NA**	**0.53**	**NA**	**NA**	**0.84**
	SVM Probability	**0.59**	**NA**	**0.63**	**NA**	**0.65**	**0.58**
	Deep Learning	**0.83**	**NA**	**1.00**	**NA**	**0.93**	**0.82**
17 [22]	**MLP**	**NA**	**0.96**	**0.85**	**0.90**	**NA**	**NA**
	XGBoost	**NA**	**0.98**	**0.94**	**0.96**	**NA**	**NA**
	DT	**NA**	**0.96**	**0.95**	**0.95**	**NA**	**NA**
	RM	**NA**	**0.96**	**0.95**	**0.95**	**NA**	**NA**
	LR	**NA**	**0.96**	**0.84**	**0.90**	**NA**	**NA**
	SVM	**NA**	**0.97**	**0.84**	**0.90**	**NA**	**NA**
	SVM RBF	**NA**	**0.98**	**0.91**	**0.94**	**NA**	**NA**
	KNN	**NA**	**0.96**	**0.90**	**0.93**	**NA**	**NA**
	Naive Bayes	**NA**	**0.97**	**0.76**	**0.85**	**NA**	**NA**
	AdaVoost	**NA**	**0.96**	**0.89**	**0.92**	**NA**	**NA**
18 [10]	**XGB**	**0.98**	**NA**	**NA**	**0.97**	**1.00**	**NA**
	LGBM	**0.98**	**NA**	**NA**	**0.98**	**1.00**	**NA**
	SVC	**0.92**	**NA**	**NA**	**0.92**	**0.98**	**NA**
19 [23]	**Naive Bayes**	**NA**	**0.69**	**72.30**	**0.70**	**0.75**	**75.60**
	SVM	**NA**	**0.70**	**73.40**	**0.71**	**0.78**	**76.30**
	DT	**NA**	**0.62**	**60.60**	**0.71**	**0.78**	**71.50**

- Support Vector Machine (SVM): This study achieved an SVM accuracy of 0.911. When compared to Study 2, where the SVM accuracy was 0.79, the current study shows a significant improvement. However, in Studies 3 and 4, where SVM accuracies reached 0.98, the current study's SVM performs slightly lower, although it's important to note that precision and recall metrics are not available for these studies.

- Decision Tree (DT): The DT model in this study showed an accuracy of 0.932. In Study 7, the DT model had a lower accuracy of 0.91. This comparison indicates an enhancement in the DT model's classification ability in the current study.

- XGBoost: The XGBoost model in this study, with an accuracy of 0.960, is on par with the high performance seen in Study 4 (0.98 accuracy). This consistency across studies highlights the model's effectiveness in fetal health classification.

- Random Forest (RM): The RM model's accuracy of 0.953 in this study does not have a direct comparison in previous studies where all metrics are reported. However, it is noteworthy that this accuracy is higher than the general trend observed in previous studies for Random Forest models.

The current study demonstrates a notable improvement in the classification of fetal health using cardiotocogram data, especially in terms of SVM and DT models. The XGBoost model maintains a high standard of performance consistent with previous research. These results highlight the advancements in machine learning models' accuracy and precision in fetal health monitoring.

This research presents a novel approach to classify fetal cardiac patterns using a range of machine learning classifiers, including Multi-Support Vector Machine (Multi-SVM), Decision Tree, Random Forest with Hyperparameter Tuning, XGBoost, and Neural Networks. The model exhibits a unique methodology by employing four separate variations in its approach. These variations involve inputting datasets in their original form, datasets that have been preprocessed using MinMaxScaler, datasets that have undergone feature selection through the utilization of SelectKBest, finally datasets that have undergone preprocessing and feature selection. The preprocessing stages, including normalization and feature selection, are carefully planned to guarantee that each feature contributes equally to the learning process of the model, hence improving its prediction performance. The paper's contribution to the field of fetal health monitoring utilizing CTG data is highlighted by the incorporation of novel classifiers and preprocessing approaches.

H. Conclusion

The results of the study indicate that machine learning models, specifically XGBoost, have considerable promise in accurately categorizing and distinguishing distinct fetal health problems using CTG data. The dataset consisted of 2126 individual CTG exam records that were categorized as Normal, Suspect, or Pathological. This dataset served as a complete foundation for assessing the efficacy of each model. The XGBoost model demonstrated improved performance compared to other models in important evaluation metrics like accuracy, precision, recall, and F1 score. This suggests that the XGBoost model has a higher ability to accurately classify the three categories of fetal health. The findings of this study underscore the capacity of sophisticated analytical techniques to revolutionize the field of fetal health monitoring, presenting great opportunities for improving prenatal care and achieving more favorable outcomes. The study successfully demonstrates the proficiency of machine learning models in efficiently managing the intricacies and subtleties of medical diagnostics within the domain of fetal health monitoring.

References

[1] AWHONN POSITION STATEMENT, "Fetal Heart Monitoring," *Journal of Obstetric, Gynecologic & Neonatal Nursing*, vol. 47, no. 6, pp. 874–877, Nov. 2018, doi: 10.1016/j.jogn.2018.09.007.

[2] S. Jain and N. Acharya, "Fetal Wellbeing Monitoring – A Review Article," *Cureus*, Sep. 2022, doi: 10.7759/cureus.29039.

[3] B. H. Al Wattar *et al.*, "Effectiveness of intrapartum fetal surveillance to improve maternal and neonatal outcomes: a systematic review and network meta-analysis," *Can Med Assoc J*, vol. 193, no. 14, pp. E468–E477, Apr. 2021, doi: 10.1503/cmaj.202538.

[4] B. S. Schifrin, "Electronic Fetal Monitoring–Prevention or Rescue?," *Front Pediatr*, vol. 8, Aug. 2020, doi: 10.3389/fped.2020.00503.

[5] A. M. Ponsiglione, C. Cosentino, G. Cesarelli, F. Amato, and M. Romano, "A comprehensive review of techniques for processing and analyzing fetal heart rate signals," *Sensors*, vol. 21, no. 18, Sep. 2021, doi: 10.3390/s21186136.

[6] J. Li *et al.*, "Automatic classification of fetal heart rate based on convolutional neural network," *IEEE Internet Things J*, vol. 6, no. 2, pp. 1394–1401, Apr. 2019, doi: 10.1109/JIOT.2018.2845128.

[7] Z. Zhao, Y. Deng, Y. Zhang, Y. Zhang, X. Zhang, and L. Shao, "DeepFHR: Intelligent prediction of fetal Acidemia using fetal heart rate signals based on convolutional neural network," *BMC Med Inform Decis Mak*, vol. 19, no. 1, Dec. 2019, doi: 10.1186/s12911-019-1007-5.

[8] A. Subasi, B. Kadasa, and E. Kremic, "Classification of the cardiotocogram data for anticipation of fetal risks using bagging ensemble classifier," in *Procedia Computer Science*, Elsevier B.V., 2020, pp. 34–39. doi: 10.1016/j.procs.2020.02.248.

[9] N. Rahmayanti, H. Pradani, M. Pahlawan, and R. Vinarti, "Comparison of machine learning algorithms to classify fetal health using cardiotocogram data," in *Procedia Computer Science*, Elsevier B.V., 2021, pp. 162–171. doi: 10.1016/j.procs.2021.12.130.

[10] Y. Yin and Y. Bingi, "Using Machine Learning to Classify Human Fetal Health and Analyze Feature Importance," *BioMedInformatics*, vol. 3, no. 2, pp. 280–298, Apr. 2023, doi: 10.3390/biomedinformatics3020019.

[11] 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, Aug. 2023, doi: 10.3390/diagnostics13152471.

[12] N. S. Hassan, A. Mohsin Abdulazeez, J. N. Saeed, D. Qader Zeebaree, A. Al-Zebari, and F. Y. H. Ahmed, "A Compassion of Three Data Miming Algorithms for Heart Disease Prediction," in *2021 IEEE Symposium on Industrial Electronics & Applications (ISIEA)*, IEEE, Jul. 2021, pp. 1–6. doi: 10.1109/ISIEA51897.2021.9509985.

[13] S. Das, H. Mukherjee, K. Roy, and C. K. Saha, "Fetal Health Classification from Cardiotocograph for Both Stages of Labor—A Soft-Computing-Based Approach," *Diagnostics*, vol. 13, no. 5, Mar. 2023, doi: 10.3390/diagnostics13050858.

[14] A. Mehbodniya *et al.*, "Fetal health classification from cardiotocographic data using machine learning," *Expert Syst*, vol. 39, no. 6, Jul. 2022, doi: 10.1111/exsy.12899.

[15] J. Piri, P. Mohapatra, and R. Dey, *Fetal Health Status Classification Using MOGA - CD Based Feature Selection Approach*.

[16] W. Gao and Y. Lu, "Fetal Heart Baseline Extraction and Classification based on Deep Learning," in *Proceedings - 2019 International Conference on Information Technology and Computer Application, ITCA 2019*, Institute of Electrical and Electronics Engineers Inc., Dec. 2019, pp. 211–216. doi: 10.1109/ITCA49981.2019.00053.
[17] S. Boudet, A. Houzé De L'aulnoit, R. Demailly, L. Peyrodie, R. Beuscart, and D. Houzé De L'aulnoit, "Fetal heart rate baseline computation with a weighted median filter," 2019. [Online]. Available: http://utsb.univ-catholille.fr/fhr-review].
[18] S. Dash, J. G. Quirk, and P. M. Djuric, "Fetal heart rate classification using generative models," *IEEE Trans Biomed Eng*, vol. 61, no. 11, pp. 2796–2805, Nov. 2014, doi: 10.1109/TBME.2014.2330556.
[19] A. Petrozziello, C. W. G. Redman, A. T. Papageorghiou, I. Jordanov, and A. Georgieva, "Multimodal Convolutional Neural Networks to Detect Fetal Compromise during Labor and Delivery," *IEEE Access*, vol. 7, pp. 112026–112036, 2019, doi: 10.1109/ACCESS.2019.2933368.
[20] S. Das, H. Mukherjee, K. C. Santosh, C. K. Saha, and K. Roy, "Periodic change detection in fetal heart rate using cardiotocograph," in *Proceedings - IEEE Symposium on Computer-Based Medical Systems*, Institute of Electrical and Electronics Engineers Inc., Jul. 2020, pp. 104–109. doi: 10.1109/CBMS49503.2020.00027.
[21] J. Spilka, J. Frecon, R. Leonarduzzi, N. Pustelnik, P. Abry, and M. Doret, "Sparse Support Vector Machine for Intrapartum Fetal Heart Rate Classification," *IEEE J Biomed Health Inform*, vol. 21, no. 3, pp. 664–671, May 2017, doi: 10.1109/JBHI.2016.2546312.
[22] Z. Hoodbhoy, M. Noman, A. Shafique, A. Nasim, D. Chowdhury, and B. Hasan, "Use of machine learning algorithms for prediction of fetal risk using cardiotocographic data," *Int J Appl Basic Med Res*, vol. 9, no. 4, p. 226, 2019, doi: 10.4103/ijabmr.ijabmr_370_18.
[23] J. Spilka *et al.*, "Using nonlinear features for fetal heart rate classification."
[24] H. T. Sadeeq and A. M. Abdulazeez, "Metaheuristics: A Review of Algorithms," *International Journal of Online and Biomedical Engineering (iJOE)*, vol. 19, no. 09, pp. 142–164, Jul. 2023, doi: 10.3991/ijoe.v19i09.39683.
[25] N. Asaad Zebari, A. A. H. Alkurdi, R. B. Marqas, and M. Shamal Salih, "Enhancing Brain Tumor Classification with Data Augmentation and DenseNet121," *Academic Journal of Nawroz University*, vol. 12, no. 4, pp. 323–334, Oct. 2023, doi: 10.25007/ajnu.v12n4a1985.
[26] D. Singh and B. Singh, "Investigating the impact of data normalization on classification performance," *Appl Soft Comput*, vol. 97, Dec. 2020, doi: 10.1016/j.asoc.2019.105524.
[27] Z. Arif Ali, Z. H. Abduljabbar, H. A. Tahir, A. Bibo Sallow, and S. M. Almufti, "eXtreme Gradient Boosting Algorithm with Machine Learning: a Review," *Academic Journal of Nawroz University*, vol. 12, no. 2, pp. 320–334, May 2023, doi: 10.25007/ajnu.v12n2a1612.
[28] V. N. G. Raju, K. P. Lakshmi, V. M. Jain, A. Kalidindi, and V. Padma, "Study the Influence of Normalization/Transformation process on the Accuracy of Supervised Classification," in *Proceedings of the 3rd International Conference*

on Smart Systems and Inventive Technology, ICSSIT 2020, Institute of Electrical and Electronics Engineers Inc., Aug. 2020, pp. 729–735. doi: 10.1109/ICSSIT48917.2020.9214160.

[29] R. Rajab Asaad, "Review on Deep Learning and Neural Network Implementation for Emotions Recognition," *Qubahan Academic Journal*, vol. 1, no. 1, pp. 1–4, Feb. 2021, doi: 10.48161/qaj.v1n1a25.

[30] J. N. Saeed, A. M. Abdulazeez, and D. A. Ibrahim, "FIAC-Net: Facial Image Attractiveness Classification Based on Light Deep Convolutional Neural Network," in *2022 Second International Conference on Computer Science, Engineering and Applications (ICCSEA)*, IEEE, Sep. 2022, pp. 1–6. doi: 10.1109/ICCSEA54677.2022.9936582.

[31] A. Hasan Bdair Aighuraibawi *et al.*, "Feature Selection for Detecting ICMPv6-Based DDoS Attacks Using Binary Flower Pollination Algorithm," *Computer Systems Science and Engineering*, vol. 47, no. 1, pp. 553–574, 2023, doi: 10.32604/csse.2023.037948.

[32] H. T. Sadeeq and A. M. Abdulazeez, "Giant Trevally Optimizer (GTO): A Novel Metaheuristic Algorithm for Global Optimization and Challenging Engineering Problems," *IEEE Access*, vol. 10, pp. 121615–121640, 2022, doi: 10.1109/ACCESS.2022.3223388.

[33] H. Utama, "Sentiment Analysis in Airline Tweets Using Mutual Information for Feature Selection," in *2019 4th International Conference on Information Technology, Information Systems and Electrical Engineering (ICITISEE)*, IEEE, Nov. 2019, pp. 295–300. doi: 10.1109/ICITISEE48480.2019.9003903.

[34] R. E. A. Armya, A. M. Abdulazeez, A. B. Sallow, and D. Q. Zeebaree, "Leukemia Diagnosis using Machine Learning Classifiers Based on Correlation Attribute Eval Feature Selection," *Asian Journal of Research in Computer Science*, pp. 52–65, Jun. 2021, doi: 10.9734/ajrcos/2021/v9i330225.

[35] J. Ircio, A. Lojo, U. Mori, and J. A. Lozano, "Mutual information based feature subset selection in multivariate time series classification," *Pattern Recognit*, vol. 108, p. 107525, Dec. 2020, doi: 10.1016/j.patcog.2020.107525.

[36] N. H. M. Khalid, A. R. Ismail, N. A. Aziz, and A. A. A. Hussin, "Performance Comparison of Feature Selection Methods for Prediction in Medical Data," in *Communications in Computer and Information Science*, Springer Science and Business Media Deutschland GmbH, 2023, pp. 92–106. doi: 10.1007/978-981-99-0405-1_7.

[37] S. Rukhsar *et al.*, "Artificial Intelligence Based Sentence Level Sentiment Analysis of COVID-19," *Computer Systems Science and Engineering*, vol. 47, no. 1, pp. 791–807, 2023, doi: 10.32604/csse.2023.038384.

[38] A. A. H. Alkurdi, "Enhancing Heart Disease Diagnosis Using Machine Learning Classifiers," *Fusion: Practice and Applications*, vol. 13, no. 1, pp. 08–18, 2023, doi: 10.54216/FPA.130101.

[39] A. Jenifa, R. Ramalakshmi, and V. Ramachandran, "Classification of Cotton Leaf Disease Using Multi-Support Vector Machine," in *IEEE International Conference on Intelligent Techniques in Control, Optimization and Signal Processing, INCOS 2019*, Institute of Electrical and Electronics Engineers Inc., Apr. 2019. doi: 10.1109/INCOS45849.2019.8951356.

[40] L. Chen *et al.*, "Multi-SVM based Dempster–Shafer theory for gesture intention understanding using sparse coding feature," *Applied Soft Computing Journal*, vol. 85, Dec. 2019, doi: 10.1016/j.asoc.2019.105787.
[41] R. Muzzammel and A. Raza, "A Support Vector Machine Learning-Based Protection Technique for MT-HVDC Systems," *Energies (Basel)*, vol. 13, no. 24, p. 6668, Dec. 2020, doi: 10.3390/en13246668.
[42] H. T. Sadeeq and A. M. Abdulazeez, "Improved Northern Goshawk Optimization Algorithm for Global Optimization," in *2022 4th International Conference on Advanced Science and Engineering (ICOASE)*, IEEE, Sep. 2022, pp. 89–94. doi: 10.1109/ICOASE56293.2022.10075576.
[43] F. P. Pramadi, C. Atika Sari, E. H. Rachmawanto, and D. Rosal Ignatius Moses Setiadi, "Flowers identification using first-order feature extraction and multi-SVM Classifier," in *Proceedings - 2020 International Seminar on Application for Technology of Information and Communication: IT Challenges for Sustainability, Scalability, and Security in the Age of Digital Disruption, iSemantic 2020*, Institute of Electrical and Electronics Engineers Inc., Sep. 2020, pp. 22–27. doi: 10.1109/iSemantic50169.2020.9234260.
[44] S. CHALO and İ. Berkan AYDİLEK, "A New Preprocessing Method for Diabetes and Biomedical Data Classification," *Qubahan Academic Journal*, vol. 2, no. 4, pp. 6–18, Jan. 2023, doi: 10.48161/qaj.v2n4a135.
[45] A. Das, P. Das, S. S. Panda, and S. Sabut, "Detection of Liver Cancer Using Modified Fuzzy Clustering and Decision Tree Classifier in CT Images," *Pattern Recognition and Image Analysis*, vol. 29, no. 2, pp. 201–211, Apr. 2019, doi: 10.1134/S1054661819020056.
[46] H. T. Sadeeq and A. M. Abdulazeez, "Giant Trevally Optimizer (GTO): A Novel Metaheuristic Algorithm for Global Optimization and Challenging Engineering Problems," *IEEE Access*, vol. 10, pp. 121615–121640, 2022, doi: 10.1109/ACCESS.2022.3223388.
[47] H. Rashid Abdulqadir, A. Mohsin Abdulazeez, and D. Assad Zebari, "Data Mining Classification Techniques for Diabetes Prediction," *Qubahan Academic Journal*, vol. 1, no. 2, pp. 125–133, May 2021, doi: 10.48161/qaj.v1n2a55.
[48] J. Liang, Z. Qin, S. Xiao, L. Ou, and X. Lin, "Efficient and Secure Decision Tree Classification for Cloud-Assisted Online Diagnosis Services," *IEEE Trans Dependable Secure Comput*, vol. 18, no. 4, pp. 1632–1644, Jul. 2021, doi: 10.1109/TDSC.2019.2922958.
[49] T. Noi Phan, V. Kuch, and L. W. Lehnert, "Land cover classification using google earth engine and random forest classifier-the role of image composition," *Remote Sens (Basel)*, vol. 12, no. 15, Aug. 2020, doi: 10.3390/RS12152411.
[50] K. I. Taher, A. M. Abdulazeez, and D. A. Zebari, "Data Mining Classification Algorithms for Analyzing Soil Data," *Asian Journal of Research in Computer Science*, pp. 17–28, May 2021, doi: 10.9734/ajrcos/2021/v8i230196.
[51] A. Parmar, R. Katariya, and V. Patel, "A Review on Random Forest: An Ensemble Classifier," in *Lecture Notes on Data Engineering and Communications Technologies*, vol. 26, Springer Science and Business Media Deutschland GmbH, 2019, pp. 758–763. doi: 10.1007/978-3-030-03146-6_86.

[52] J. N. Saeed, A. M. Abdulazeez, and D. A. Ibrahim, "Automatic Facial Aesthetic Prediction Based on Deep Learning with Loss Ensembles," *Applied Sciences*, vol. 13, no. 17, p. 9728, Aug. 2023, doi: 10.3390/app13179728.
[53] P. Schratz, J. Muenchow, E. Iturritxa, J. Richter, and A. Brenning, "Hyperparameter tuning and performance assessment of statistical and machine-learning algorithms using spatial data," *Ecol Modell*, vol. 406, pp. 109–120, Aug. 2019, doi: 10.1016/j.ecolmodel.2019.06.002.
[54] S. M. Almufti, "Hybridizing Ant Colony Optimization Algorithm for Optimizing Edge-Detector Techniques," *Academic Journal of Nawroz University*, vol. 11, no. 2, pp. 135–145, May 2022, doi: 10.25007/ajnu.v11n2a1320.
[55] S. M. Almufti, A. A. H. Alkurdi, and E. A. Khoursheed, "Artificial Bee Colony Algorithm Performances in Solving Constraint-Based Optimization Problem," *TELEMATIQUE*, vol. 21, no. 1, p. 2022, 2023.
[56] H. T. Sadeeq, S. Y. Ameen, and A. M. Abdulazeez, "Cancer Diagnosis based on Artificial Intelligence, Machine Learning, and Deep Learning," in *2022 International Conference on Innovation and Intelligence for Informatics, Computing, and Technologies (3ICT)*, IEEE, Nov. 2022, pp. 656–661. doi: 10.1109/3ICT56508.2022.9990784.
[57] A. bin Asad, R. Mansur, S. Zawad, N. Evan, and M. I. Hossain, "Analysis of Malware Prediction Based on Infection Rate Using Machine Learning Techniques," in *2020 IEEE Region 10 Symposium (TENSYMP)*, IEEE, 2020, pp. 706–709. doi: 10.1109/TENSYMP50017.2020.9230624.
[58] D. Upadhyay, J. Manero, M. Zaman, and S. Sampalli, "Gradient Boosting Feature Selection with Machine Learning Classifiers for Intrusion Detection on Power Grids," *IEEE Transactions on Network and Service Management*, vol. 18, no. 1, pp. 1104–1116, Mar. 2021, doi: 10.1109/TNSM.2020.3032618.
[59] C.-C. Wang, P.-H. Kuo, and G.-Y. Chen, "Machine Learning Prediction of Turning Precision Using Optimized XGBoost Model," *Applied Sciences*, vol. 12, no. 15, p. 7739, Aug. 2022, doi: 10.3390/app12157739.
[60] F. Didi, H. Pallathadka, S. Abdullaev, R. R. Asaad, S. Askar, and N. H. Haroon, "Probing the impact of process variables in laser-welded aluminum alloys: A machine learning study," *Mater Today Commun*, vol. 38, p. 107660, Mar. 2024, doi: 10.1016/j.mtcomm.2023.107660.
[61] A. A. Taha and S. J. Malebary, "An Intelligent Approach to Credit Card Fraud Detection Using an Optimized Light Gradient Boosting Machine," *IEEE Access*, vol. 8, pp. 25579–25587, 2020, doi: 10.1109/ACCESS.2020.2971354.
[62] N. R. Kapoor *et al.*, "Event-Specific Transmission Forecasting of SARS-CoV-2 in a Mixed-Mode Ventilated Office Room Using an ANN," *Int J Environ Res Public Health*, vol. 19, no. 24, p. 16862, Dec. 2022, doi: 10.3390/ijerph192416862.
[63] R. R. Asaad and R. I. Ali, "Back Propagation Neural Network(BPNN) and Sigmoid Activation Function in Multi-Layer Networks," *Academic Journal of Nawroz University*, vol. 8, no. 4, pp. 216–221, Nov. 2019, doi: 10.25007/ajnu.v8n4a464.
[64] A. M. Hemeida *et al.*, "Nature-inspired algorithms for feed-forward neural network classifiers: A survey of one decade of research," *Ain Shams*

Engineering Journal, vol. 11, no. 3. Ain Shams University, pp. 659–675, Sep. 01, 2020. doi: 10.1016/j.asej.2020.01.007.

[65] N. A. Kako and A. M. Abdulazeez, "Peripapillary Atrophy Segmentation and Classification Methodologies for Glaucoma Image Detection: A Review," *Current Medical Imaging Formerly Current Medical Imaging Reviews*, vol. 18, no. 11, pp. 1140–1159, Nov. 2022, doi: 10.2174/1573405618666220308112732.

[66] A. M. Alsugair, K. S. Al-Gahtani, N. M. Alsanabani, A. A. Alabduljabbar, and A. S. Almohsen, "Artificial Neural Network Model to Predict Final Construction Contract Duration," *Applied Sciences*, vol. 13, no. 14, p. 8078, Jul. 2023, doi: 10.3390/app13148078.

[67] J. N. Saeed, A. M. Abdulazeez, and D. A. Ibrahim, "An Ensemble DCNNs-Based Regression Model for Automatic Facial Beauty Prediction and Analyzation," *Traitement du Signal*, vol. 40, no. 1, pp. 55–63, Feb. 2023, doi: 10.18280/ts.400105.

[68] A. E. Mehyadin, A. M. Abdulazeez, D. A. Hasan, and J. N. Saeed, "Birds Sound Classification Based on Machine Learning Algorithms," *Asian Journal of Research in Computer Science*, pp. 1–11, Jun. 2021, doi: 10.9734/ajrcos/2021/v9i430227.

[69] B. M. K. P, S. P. R, N. R K, and A. K, "Type 2: Diabetes mellitus prediction using Deep Neural Networks classifier," *International Journal of Cognitive Computing in Engineering*, vol. 1, pp. 55–61, Jun. 2020, doi: 10.1016/j.ijcce.2020.10.002.

[70] D. Ayres-de-Campos, J. Bernardes, A. Garrido, J. Marques-de-S�, and L. Pereira-Leite, "Sisporto 2.0: A program for automated analysis of cardiotocograms," *J Matern Fetal Med*, vol. 9, no. 5, pp. 311–318, Sep. 2000, doi: 10.1002/1520-6661(200009/10)9:5<311::AID-MFM12>3.0.CO;2-9.

[71] J. Li and X. Liu, "Fetal Health Classification Based on Machine Learning," in *2021 IEEE 2nd International Conference on Big Data, Artificial Intelligence and Internet of Things Engineering, ICBAIE 2021*, Institute of Electrical and Electronics Engineers Inc., Mar. 2021, pp. 899–902. doi: 10.1109/ICBAIE52039.2021.9389902.