International Journal of Computing and Digital Systems
ISSN 2210-142X
Int. J. Com. Dig. Sys. 16, No.1 (Jul-24)

http://dx.doi.org/10.12785/ijcds/160137

CFCM-SMOTE: A Robust Fetal Health Classification to Improve Precision Modeling in Multiclass Scenarios

Ahmad Ilham[1,3], Asdani Kindarto[1], Akhmad Fathurohman[1], Laelatul Khikmah[2,3], Rima Dias Ramadhani[1], Syafrie Abdunnasir Jawad[1], Dhewi April Liana[1,3], Aura Amylia. AR[1,3], Ahmed Kareem Oleiwi[4] and Astri Mutiar[5]

[1]*Department of Informatics, Universitas Muhammadiyah Semarang, Semarang, Indonesia*
[2]*Department of Statistics, Institut Teknologi Statistika dan Bisnis Muhammadiyah Semarang, Semarang, Indonesia*
[3]*Intelligent Data Science Research Group, Semarang, Indonesia*
[4]*Department of Computer Technical Engineering, Islamic University of Najaf, Najaf, Iraq*
[5]*Department of Nursing, Sekolah Tinggi Ilmu Keperawatan PPNI Jawa Barat, Bandung, Indonesia*

Received 11 Dec. 2023, Revised 19 Apr. 2024, Accepted 21 Apr. 2024, Published 1 Jul. 2024

Abstract: The advent of cardiotocography (CTG) has radically transformed prenatal care, facilitating in-depth evaluations of fetal health. Despite this, the reliability of CTG is frequently undermined by data-related issues, such as outliers and class imbalanced data. To address these challenges, our study introduces an innovative integrated methodology that combines cluster-based fuzzy C-means (CFCM) with the synthetic minority oversampling technique (SMOTE) to improve the precision of classification of fetal health status classification in multiclass scenarios. We used a considerable dataset from the UCI Machine Learning Repository, employing CFCM to manage outliers and SMOTE to rectify class imbalanced data. This approach has significantly improved the performance of the classification algorithm, a fact that is corroborated by the comprehensive experimental validation that can be found in the study in Ref. [1]. We observed notable improvements in several evaluation metrics, including precision (PRC), sensitivity (SNS), specificity (SPC), F1 score (F1-S), and accuracy (ACC), surpassing the capabilities of prior methodologies. Specifically, the deployment of our algorithm amplified the precision (PRC: from 98.16% to 99.58%), sensitivity (SNS: from 95.82% to 100%), specificity (SPC: from 85.81% to 99.75%), F1 score (F1-Score: from 96.98% to 99.79%), and accuracy (ACC: from 94.20% to 99.84%) of the Classification and Regression Tree (CART) algorithm for the 'normal' class, while also improving the precision and accuracy of the Random Forest (RF) algorithm from PRC: 94.77% to 95.89% and ACC: 90.60% to 97.45%. These results confirm the potential of CFCM-SMOTE as a robust model for fetal health diagnostics and as a basic strategy for the development of predictive analyzes in prenatal healthcare.

Keywords: Fetal Health Classification, Cardiotocography, Outlier, Class Imbalanced Data, Fuzzy C-Means, SMOTE

1. INTRODUCTION

Medical diagnosis, a meticulous process, requires healthcare professionals to assess various factors for accurate patient health status determination, involving a comprehensive review of medical history, physical exams, lab tests, and imaging studies [2]. The complexity escalates when evaluating indirect evidence like the risk of fetal death during labor [3]. Such critical assessment necessitates a profound understanding of mother-fetus physiological dynamics and data interpretation skills. Healthcare providers must merge this knowledge with advanced diagnostic techniques to accurately evaluate and lessen fetal death risk, ensuring maternal and fetal well-being [4].

Ultrasound waves monitor the Fetal Heart Rate (FHR) by detecting frequency changes from the fetus's heartbeat. A transducer in the mother's abdomen sends signals to the heart of the fetus, and these reflected frequencies are processed by cardiotocography (CTG) to obtain the FHR. Uterine contraction pressure (UC) is monitored similarly, and both measurements are displayed on a screen (see Figure 1).

Several studies reported that the visual interpretation of CTG is highly sensitive in detecting the risk of hypoxia, particularly with pathological traces. This helps medical professionals assess fetal well-being during labor and act quickly. However, the specificity for suspect cases is low,

E-mail address: ahmadilham@unimus.ac.id, asdani@gmail.com, akhmadfathur@unimus.ac.id

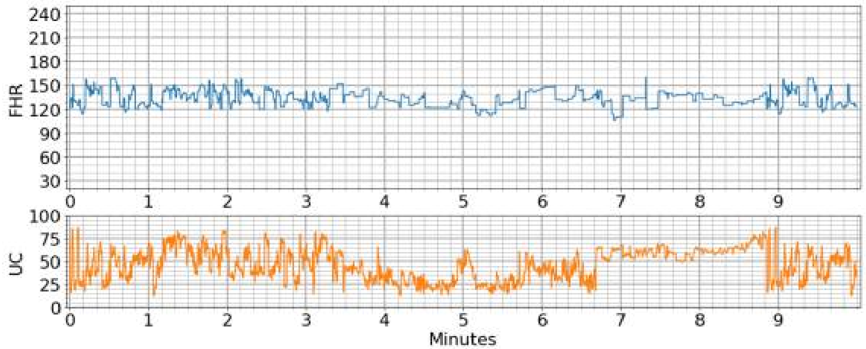

Figure 1. Fetal Heart Rate (FHR) and Uterine Contraction (UC) Patterns Over A 9-Minute Interval. The top panel displays the FHR, indicating the beats per minute, while the bottom panel shows the UC pressure measured in units over time (extracted from [5]).

often necessitating complex and impractical confirmatory procedures such as pH testing of the fetal scalp.

Despite the aim of fetal monitoring to lower mortality rates, perinatal deaths are on the rise, partly due to unnecessary caesarean sections from CTG use. To mitigate this, obstetric bodies like FIGO, NICHD, and NICE have created CTG interpretation guidelines. Yet, misinterpretation and observer variability continue, leading to birth-related brain injuries.

In the initial stages of automation, conventional algorithms were used to classify cardiotocography (CTG) signals into different classes: normal, suspect, and pathological. This was achieved by identifying and defining different features in fetal heart rate (FHR) signals. However, this approach had limitations and has improved with advances in AI and machine learning techniques. These algorithms often did not account for the inherent uncertainty in medical diagnoses. With advances in obstetric practice, machine learning has proven to be effective in the classification and diagnosis of FHR patterns, with various systems described in Table I.

CTG data analysis is a promising approach to improve prenatal care. However, it is important to consider the challenges faced by this technology when working with imbalanced data classes and outliers in the CTG dataset. The authors of Ref. [7] have highlighted these challenges, indicating a decrease in diagnostic precision due to data set problems. To address these challenges, we must explore ways to improve the quality of data, including algorithms to address unbalanced data classes and outliers, and develop algorithms to handle these issues more effectively. In doing this, the improved accuracy, classification precision, and reliability of machine learning algorithms for analyzing CTG data ultimately improve prenatal care for patients.

Our proposed algorithm aims to address two problems in CTG datasets: class imbalanced and outliers in fetal health status classification, which we call CFCM-SMOTE. By integrating cluster-based fuzzy c-means (CFCM) and synthetic minority sampling techniques (SMOTE), this novel approach in the field aims to improve the quality and representativeness of the data set significantly. The application of this strategy can significantly improve the accuracy of fetal health status classification and ultimately provide better health outcomes for the baby.

The primary motivation behind developing CFCM-SMOTE is to improve prenatal care by offering a more reliable and precise instrument for assessing fetal health risks. Addressing key challenges such as outliers and imbalanced data, the goal is to decrease erroneous diagnoses and guarantee prompt medical interventions for pregnant women. This research aims to reduce maternal and infant mortality by utilizing the potential of CTG technology in prenatal care.

Specifically, this paper has three main objectives:

1 Evaluate the effectiveness of the CFCM-SMOTE approach in addressing the problems caused by outliers and class imbalanced data in CTG datasets.
2 Compare the performance of this new approach with existing classification algorithms and state-of-the-art techniques.
3 Emphasize the importance of adopting advanced data preprocessing techniques to improve the diagnostic accuracy of the fetal health assessment.

The remainder of this paper is structured as follows. Section 2 provides an overview of the relevant literature. Section 3 introduces the materials and the algorithm suggested. The experimental results from the comparison of the proposed approach with others are provided, along with detailed analysis discussions in Sections 4 and 5. Finally, the final section is dedicated to the conclusion of this paper's work.

2. Related Works

The classification of fetal health risks using cardiotocography (CTG) datasets has witnessed significant advances in recent years. Numerous studies have introduced various

TABLE I. Comparative summary of various analytical systems for FHR and UC monitoring (extracted from [6])

Years	Systems	Details
1991	System 8000	Based on the algorithm proposed by Dawes and Redman in 1981. Uses crisp logic.
1995	NST-EXPERT	Developed by Alonso-Betanzos. Capable of making diagnoses and suggesting treatments.
2002	CAFE	An upgraded version of NST-EXPERT developed by Alonso-Betanzos.
2008	Omniview SisPorto 3.5	Developed by Ayres-de-Campos. Analyzes fetal status using standard CTG parameters and ST analysis of FHR signals.
2015	TOITU	Developed by Maeda and Noguchi. Based on ANN.

methodologies to address the challenges of classifying fetal health status, with a primary focus on the early detection of conditions that can pose a risk to the fetus. This section aims to provide a comprehensive overview of the methodologies used in this field and to compare the proposed CFCM-SMOTE approach with existing techniques.

A. Existing Algorithms

In their pivotal work, in Ref. [3] introduced a novel algorithm designed to classify cardiotocography (CTG) data, improving fetal health assessment during pregnancy. This algorithm integrates the Apriori algorithm with a multi-ensemble approach combining AdaBoost and Random Forest, specifically targeted at improving classification accuracy for ambiguous CTG data classes. The algorithm achieved a significant classification accuracy of 97.6% and an AUC of 0.98, able to differentiate between normal, suspect, and pathological fetal conditions despite challenges such as the non-stationary nature and unbalanced class distribution of the CTG data. The validity and reliability of the algorithm were affirmed by a 10-fold cross-validation, providing a comprehensive evaluation across various data samples. Although ensemble algorithms generally mitigate overfitting risks better than single-model approaches, their complexity can still precipitate overfitting in scenarios involving diverse or noisy data.

Research conducted by [8] investigated the classification of fetal health using CTG data using several machine learning algorithms, including the support vector machine, random forest, multilayer perceptron, and K-nearest neighbors. The study classified health conditions as normal, questionable, or pathological and used 10-fold cross-validation for its evaluations. Among the algorithms tested, Random Forest performed the best, achieving an accuracy of 94.5%, a precision of 93%, a recall of 94%, and a 93% F1 score. This research highlights the importance of regression and correlation analyzes in elucidating the effects of CTG attributes on fetal health, although it acknowledges limitations such as insufficient analysis of individual feature contributions and potential biases from data class imbalances that could impact the generalizability of the findings.

The authors in Ref. [9] conducted a comprehensive exploration of resilient machine learning classification algorithms, applied effectively to various stages of labor. Using standard classifiers such as support vector machine (SVM), random forest (RF), multilayer perceptron (MLP), and bagging, these were utilized for CTG classification. The results were rigorously validated through a 5-fold cross-validation, with the ROC-AUC measure evaluating the model's efficacy. SVM and RF demonstrated strong performance; for cases classified as suspicious, SVM and RF showed accuracies of 97.4% and 98% respectively, with sensitivity around 96.4% and specificity approximately 98%. Despite these promising results, the study acknowledges limitations, including inadequate handling of class imbalance and the potential for elevated validation results not reflecting true clinical conditions due to reliance on limited and homogeneous datasets. Furthermore, the implementation of these models requires considerable computational resources, posing challenges in resource-limited clinical settings. Lastly, while ROC-AUC was used primarily, other metrics such as precision, recall, and F1-score are also vital for a comprehensive evaluation, a factor that may not have been fully integrated into the study.

Other research is described in Ref. [10] also introduced a groundbreaking algorithm that skillfully employs the Apriori algorithm in conjunction with a multi-ensemble algorithm that merges AdaBoost and Random Forest. This innovative approach was carefully designed to improve the precision of the fetal health classification using CTG, directly addressing the challenges posed by ambiguous data classes often seen in pregnancy. The performance of this algorithm was impressive, achieving a classification accuracy of 97.6% and an AUC of 0.98, which proved to be highly effective in distinguishing between normal, suspect, and pathological fetal conditions. Despite the nonstationary and imbalanced nature of CTG data, using 10-fold cross-validation has provided a robust and comprehensive evaluation of the algorithm across diverse data samples. Although the ensemble methodology typically reduces the risks of overfitting seen in single algorithm applications, it is essential to note that the complexity inherent in these algorithms can still result in overfitting, particularly in varied or noisy data scenarios.

In Ref. [11] they embarked on a detailed exploration of various machine learning classification algorithms to assess the complex risks associated with maternal health. Their focused study analyzed key parameters such as maternal age, heart rate, blood oxygen level, blood pressure, and body temperature. The study used robust algorithms, including LDA, QDA, KNN, decision tree, random forest, bagging, and support vector machine, to unravel the complex web of health risks of patients. Through the application of split validation and 10-fold cross-validation, the Support Vector Machine emerged as the most reliable tool, achieving an accuracy rate of 86.13%. The primary objective of this research was to quantitatively assess the severity of maternal health problems using machine learning, providing a comprehensive analysis of risk factors. However, the researchers recognized potential shortcomings, such as outliers and class imbalanced data in the CTG, which might skew algorithm performance and distort the recognition of true risk patterns.

In another significant research effort, Salini et al. [12] thoroughly evaluated various machine learning algorithms to improve the precision and effectiveness of prenatal care diagnosis through the classification of CTG data. The preprocessing phase was executed meticulously, incorporating normalization and managing missing data to ensure data quality. This preparation included sophisticated techniques such as missing value filling, outlier removal, and data scaling. Feature engineering was crucial in extracting relevant features from the CTG data, significantly enhancing the prediction accuracy. The algorithms tested included random forest, logistic regression, decision trees, support vector classifiers, and k-nearest neighbor. Among these, Random Forest distinguished itself by demonstrating outstanding accuracy, with a success rate of 93%. To ensure the reliability of its findings, the study implemented 10-fold cross-validation. Despite its successes, the study recognized some limitations, particularly in dealing with class imbalances and outliers, which could potentially affect the reliability of the algorithmic findings.

B. CFCM-SMOTE Approach

In contrast, our proposed CFCM-SMOTE methodology integrates two key techniques: Conditional fuzzy c-means (CFCM) and synthetic minority oversampling technique (SMOTE). CFCM is used to identify and address outliers within the dataset, while SMOTE is used to balance the class distribution through synthetic sampling of minority classes. This dual approach facilitates a more effective mitigation of outliers and class imbalanced data, significant challenges in fetal health classification using CTG datasets.

The primary advantage of the CFCM-SMOTE algorithm lies in its ability to improve classification accuracy by improving the handling of outlier and class imbalanced data. Unlike existing techniques that often overlook one of these critical aspects, the CFCM-SMOTE approach provides a more holistic and robust solution. The innovation of this algorithm is the unique combination of CFCM and SMOTE, which not only improves dataset quality through preprocessing, but also supports more accurate machine learning algorithms in fetal health risk classification.

3. Materials and Algorithms

A. CTG Dataset

The dataset used in this study was obtained from the University of California, Irvine Machine Learning Repository (UCI ML Repository) [13]. Known as the Cardiotocography (CTG) dataset, it consists of 2.126 records. The data include features generated from SISSPorto 2.0 and was collected from pregnant women in Portugal between 29 and 42 years of age. Evaluation of fetal health status was performed by obstetricians, following the standards of the International Federation of Gynecology and Obstetrics (FIGO) [14], as shown in Table II.

Within this dataset, the fetal health status is divided into three distinct classes: normal, suspect, and pathological. The distribution of these classes in the dataset is as follows: 'normal' with 1,655 records (70.8%), 'suspect' with 295 records (13.9%) and 'pathological' with 176 records (8.3%), as shown in Figure 2.

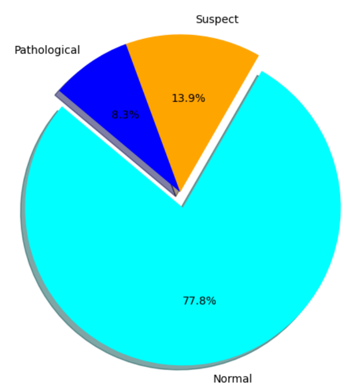

Figure 2. Class Distribution in Cardiotocography (CTG) Dataset: Percentage of Classes 'Normal', 'Suspect', and 'Pathological

B. Outlier in CTG Dataset

By applying the interquartile range (IQR) method, the preliminary analysis of the dataset has successfully identified the presence of outliers. A total of 449 outlier observations were detected across nine variables, specifically: V1, V19, V16, V15, V10, V2, V8, V6 and V5. Figure 3 illustrates the distribution of feature values in relation to class classification before data preprocessing, showing variations in distribution patterns between different classes.

Figure 3 indicates that certain variables display skewed distributions, which could potentially introduce bias into classification results if left unaddressed. This problem is particularly pronounced in variables V16 and V19. In contrast, variable V2, although it does not show a perfectly normal distribution, presents a distribution pattern that is comparatively more normal than those of the other variables. This observation underscores the need for data preprocessing to normalize distributions and mitigate the

TABLE II. Features description on CTG dataset

Features name	Label	Description
baseline_value	V0	Baseline Fetal Heart Rate (FHR) (beats per minute)
accelerations	V1	Number of accelerations per second
fetal_movement	V2	Number of fetal movements per second
uterine_contractions	V3	Number of uterine contractions per second
light_decelerations	V4	Number of light decelerations (LDs) per second
severe_decelerations	V5	Number of severe decelerations (SDs) per second
prolonged_decelerations	V6	Number of prolonged decelerations (PDs) per second
abnormal_short_term_variability	V7	Percentage of time with abnormal short-term variability
mean_value_of_short_term_variability	V8	Mean value of short-term variability
percentage_of_time_with_abnormal _long_term_variability	V9	Percentage of time with abnormal long-term variability
mean_value_of_long_term_variability	V10	Mean value of long-term variability
histogram_width	V11	Width of histogram made using all values from a record
histogram_min	V12	Histogram minimum value
histogram_max	V13	Histogram maximum value
histogram_number_of_peaks	V14	Number of peaks in the exam histogram
histogram_number_of_zeroes	V15	Number of zeros in the exam histogram
histogram_mode	V16	Histogram mode
histogram_mean	V17	Histogram mean
histogram_median	V18	Histogram median
histogram_variance	V19	Histogram variance
histogram_tendency	V20	Histogram tendency: -1 = left asymmetric, 0 = symmetric, 1 = right asymmetric
fetal_health	Class	1-Normal; 2-Suspect; 3-Pathological

influence of outliers on the effectiveness of the classification algorithm. This conclusion is based on the assumption that the integrity of the CTG data significantly affects the precision of the subsequent classification algorithm. Therefore, detecting and managing outliers is an essential aspect of data preprocessing, aimed at reducing data distortions and enhancing the performance of the forthcoming fetal health classification algorithm.

C. *Feature Selection and Resampling*

Feature Selection. Feature selection involves utilizing the Conditional Fuzzy C-means (CFCM) algorithm, a variant of the Fuzzy C-means (FCM), which categorizes data based on degrees of membership. CFCM reduces ambiguity via an objective function, incorporating constraints for directed cluster formation. In this study, CFCM processes outliers from the CTG dataset by reducing the distance of outlying data from the center and selects features based on their similarity. These stages are illustrated in Algorithm 1.

Resampling. The class imbalanced data issue will be tackled through the resampling algorithm in machine learning, employing synthetic sample creation via the Synthetic Minority Oversampling Technique (SMOTE) to prevent bias towards the majority class. Table III provides a comprehensive definition and a visual depiction of the input parameters of SMOTE.

The procedural steps of SMOTE are the following Algorithm 2.

D. *Proposed algorithm*

Figure 4 illustrates the proposed algorithm, which uses CFCM to handle outliers and SMOTE to tackle class imbalanced data through synthetic points. The performance of CART, RF, SVM, and k-NN was evaluated using cross-validation.

E. *Algorithm Validation*

We implemented the 10-fold cross-validation method, as recommended by Kaliappan et al. [15], to ensure a stable evaluation. This technique partitions the dataset into ten equal segments. Nine out of these ten segments are utilized as training data for the algorithm, and the remaining segment acts as validation data. The selection of the validation data segment is rotated iteratively, allowing each segment to serve as validation data. This strategy is instrumental in mitigating the risk of overfitting within the Cardiotocography (CTG) dataset.

F. *Matrix Algorithms Evaluation*

The confusion matrix is an essential evaluation tool to analyze the performance of classification algorithms, especially in scenarios where the class distribution is unbalanced. It offers detailed information on the number of

TABLE III. Input definition in SMOTE

Input	Description
x	The dataset contains n samples and m features
y	Class labels indicate the minority and majority class
k	The number of nearest neighbors to be used for interpolation
N	The number of synthetic samples to be generated for the minority class

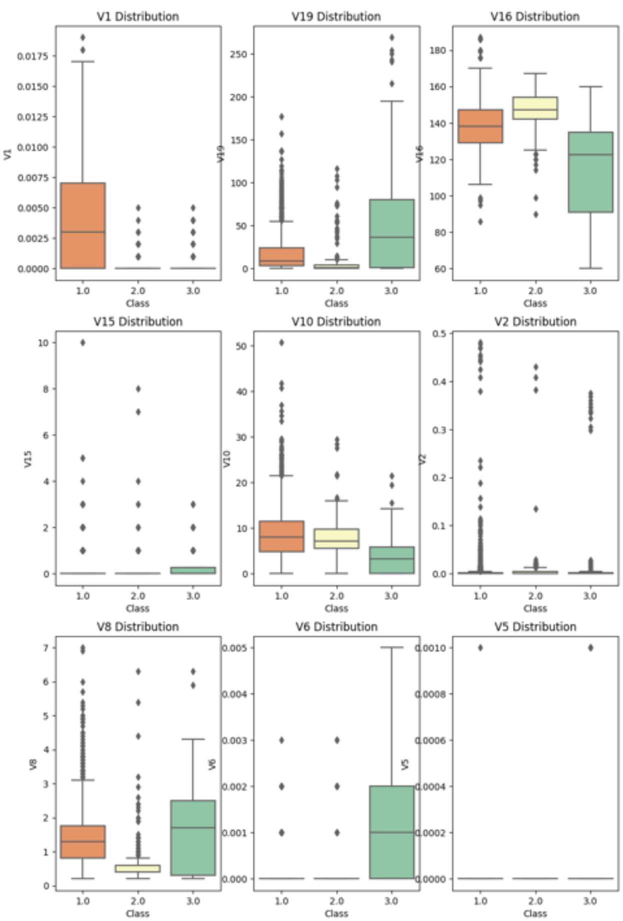

Figure 3. Features Distribution of V1, V19, V16, V15, V10, V2, V8, V6, and V5 in CTG Dataset before Data Preprocessing

true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN), which are crucial to understanding how the algorithm differentiates between classes. The data provided by the confusion matrix enable the calculation of important metrics such as precision (PRC), which measures the proportion of correctly predicted positive outcomes out of all positive predictions; sensitivity (SNS), or the rate at which the algorithm successfully identifies positive cases; specificity (SPC), assessing the algorithm's capability to recognize negative cases; the F1 score, representing the harmonic mean of precision and sensitivity, thus offering a balance between them; and accuracy (ACC), which calculates the proportion of correctly predicted cases for both

Algorithm 1 Algorithm for Updating Membership and Feature Relevance Matrices

Initialize membership matrix $U = [u_{ij}]$ randomly
Input: Data points x_i, cluster count c, feature count p, fuzziness m
Output: Updated matrices U and V
repeat
  Calculate the cluster centers $C = \{c_1, c_2, \ldots, c_n\}$ using current U
  Compute the distance matrix $D = [d_{ij}]$ for each feature to each cluster center
  for each feature i and cluster j **do**
    Update v_{ij} using
    $v_{ij} = \frac{1}{\sum_{k=1}^{p}\left(\frac{d_{ik}}{d_{kj}}\right)^{\frac{2}{m-1}}}$
    Normalize each row of V so that each row sums to 1
  end for
  for each data point i and cluster j **do**
    Update u_{ij} using
    $u_{ij} = \frac{1}{\sum_{k=1}^{c}\left(\frac{v_{ik}}{v_{jk}}\right)^{\frac{2}{m-1}}}$
    Normalize each row of U so that each row sums to 1
  end for
until Convergence of V or maximum iterations are reached
Select the top p features based on the final matrix V

positive and negative classes out of all predictions made. The equations for the metrics (1, 2, 3, 4, 5) in this document are designed to calculate these metrics, providing valuable information on various aspects of the performance of the classification algorithm under evaluation.

$$\text{Precision (PRC)} = \frac{TP}{TP + FP} \tag{1}$$

$$\text{Sensitivity (SNS)} = \frac{TP}{TP + FN} \tag{2}$$

$$\text{Specificity (SPC)} = \frac{TN}{TN + FP} \tag{3}$$

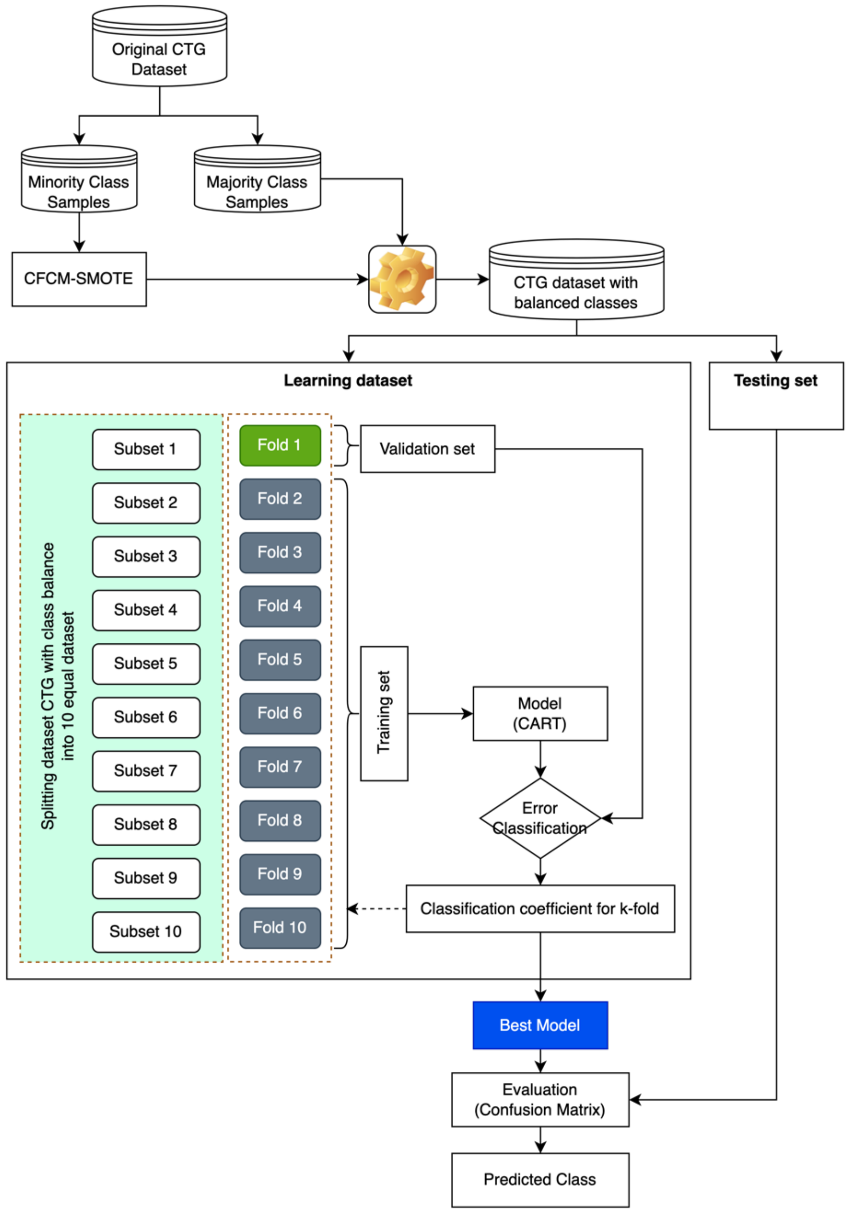

Figure 4. Blockdiagram of the Proposed Algorithm

$$\text{F1 Score} = \frac{2 \times \text{SNS} \times \text{PRC}}{\text{SNS} + \text{PRC}} \tag{4}$$

$$\text{Accuracy (ACC)} = \frac{TP + TN}{TP + TN + FP + FN} \tag{5}$$

G. Statistical Validation

Friedman and pairwise comparison tests using Dunn-Bonferroni methods are key to validating experimental results when comparing algorithm performance [16], [17], [18]. The Friedman test, a non-parametric method, is advantageous when parametric test assumptions like normality are unmet. It is used in computational studies to evaluate multiple algorithms across datasets, focusing on their ranking rather than averages, thereby minimizing outlier impact. This approach is crucial when evaluating algorithm enhancements or modifications, including those involving specific sampling or data preprocessing techniques. Ref. [19] underscores the importance of non-parametric tests like the Friedman test in machine learning algorithm comparisons, given its suitability for typical scenarios in this field where data distributions often defy parametric test assumptions. The Friedman test formula is provided in

Algorithm 2 Synthetic Minority Over-sampling Technique (SMOTE)

Input: Dataset X, minority class samples in X, number of nearest neighbors k, number of synthetic samples to generate N
Output: Dataset X augmented with synthetic samples
Identify the samples belonging to the minority class in X
for each sample x_i from the minority class in X **do**
  Determine k nearest neighbors (randomly chosen) from the minority class, excluding x_i, using Euclidean distance
  for each nearest neighbor x_{nn} **do**
    Generate a new synthetic sample x_{new} by interpolating between x_i and x_{nn} using:
    $x_{new} = x_i + \text{rand}(0, 1) \times (x_{nn} - x_i)$
    Incorporate the synthetic sample x_{new} into the dataset X
  end for
end for
Repeat the process for N iterations to produce N synthetic samples

Equation (6).

$$chi_F^2 = \frac{12N}{k(k+1)}\left[\sum R_j^2 - \frac{k(k+1)^2}{4}\right] \tag{6}$$

where N is the number of subjects, k represents the number of groups or treatments, and R_j denotes the total rank for group j. The test statistic (χ_F^2) thus calculated is then compared to the critical value of the chi-square distribution with $k-1$ degrees of freedom to determine the statistical significance of the observed differences.

When significant differences are identified using the Friedman test, the Paired Comparison Test is applied using the Dunn-Bonferroni method for a more detailed analysis of differences between groups on a pairwise basis. This test facilitates pairwise comparisons between groups by adjusting the value p with the Bonferroni correction to mitigate the risk of type I error, which increases with the number of comparisons. The z-statistic in the Dunn test is calculated using Equation (7).

$$z = \frac{R_i - R_j}{\sqrt{\frac{k(k+1)}{6N}}} \tag{7}$$

where R_i and R_j are the average ranks of the two compared groups. The adjusted p value takes into account the total number of comparisons. This method ensures that the conclusions derived from the statistical analysis are robust and credible, laying a strong foundation for the interpretation of the research results.

4. EXPERIMENTAL WORKS AND RESULTS

The research was carried out using the Python version 3.8.8 environment and KNIME version 4.7.5. For evaluation purposes, various metrics were used, including accuracy, F1 score, sensitivity, specificity, precision, error classification (EC), and execution time (ET). Excel, MS Office, was used to perform statistical comparisons between algorithms. The specific parameters of the classification algorithms used are detailed in Table IV.

TABLE IV. Parameters configuration

Algorithm	Parameters and Values
CFCM	addClusterFeatures: TRUE; Cluster: 3; Iterations: 50; Fuzziness: 2; Measure Types: Numerical Measures; Numerical Measures: Euclidean Distance.
SMOTE	Neighbors: 5; Normalize: TRUE; Equalize Classes: TRUE; Detect Minority Class: TRUE; Nominal Change Rate: 0.5.
CART	Heuristic: YES; Min Num Obj: 2; Num Folds Pruning: 10; Seed: 1; Size Per: 1; Use One SE: FALSE; Use Prune: TRUE.
RF	Max Depth: 0; Num Execution Slots: 1; Num Features: 21; Num Trees: 10.
SVM	Cache Size: 40; Coef: 0.1; Cost: 1.0; Eps: 0.001; Gamma: 0.0; Kernel Type: Radial Basis Function; Loss: 0.1.
k-NN	Num of Neighbors: 3; Distance Metric: Euclidean; Nearest-Neighbor Search Algorithm: Linear NNSearch; Mean Squared: FALSE; Window Size: 0.

A. Experimental Scheme

We conducted three main stages of experiments to evaluate the effectiveness of the CFCM-SMOTE algorithm in classifying CTG dataset:

1. Initial Data Evaluation
2. Classification algorithms Comparison
3. Comparison with Existing algorithms

B. Preprocessing: Initial Data Evaluation (Stage 1)

In the initial data evaluation stage, the Conditional Fuzzy C-means (CFCM) algorithm was implemented to identify and mitigate the impact of outlier data. The results of this data preprocessing process are shown in Figure 6, which shows the management of nine features in the Cardiotocography (CTG) dataset.

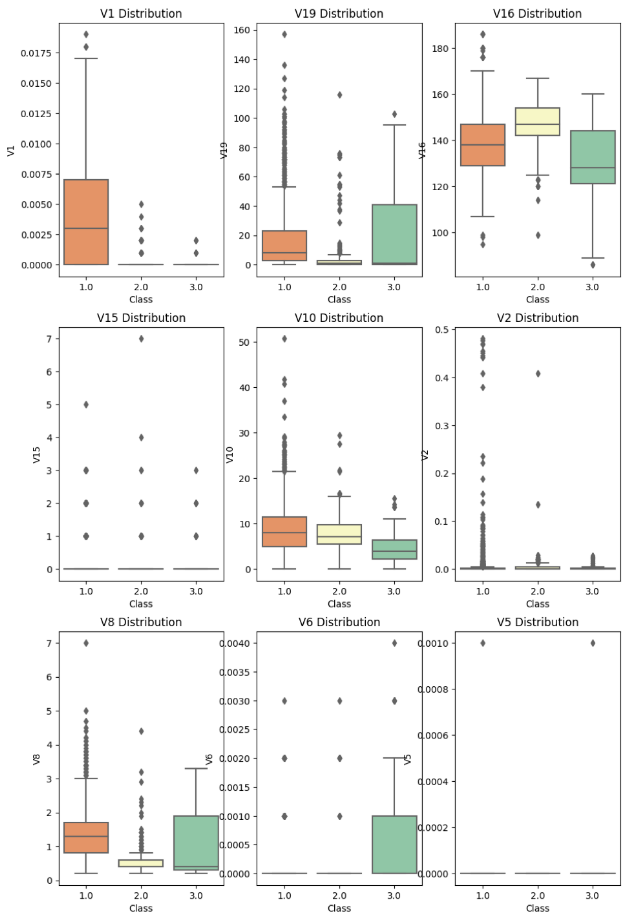

Figure 5. Improvement of CTG Data Feature Distribution After CFCM Preprocessing

Figure 6 shows that features V1, V19, V16, V15, V10, V2, V8, V6, and V5 experienced a significant reduction in the number of outliers after CFCM application. This reduction is evident from the decreased number of points that lie outside the 'whiskers' boundary in the box-and-whisker plot, indicating outliers, compared to before the treatment. Features such as V16 and V19, which previously exhibited asymmetrical distributions, now display distributions that are more normal or symmetrical. This alteration underscores the effectiveness of CFCM in filtering and minimizing outliers, leading to a more consistent and controlled data distribution.

Furthermore, the enhanced uniformity in the median's position (center line in the box) suggests that after applying CFCM, the data are less influenced by extreme values on either side of the distribution. The reduction in the Interquartile Range (IQR) for some features also signifies a decrease in data variability, indicating that the data have become more centralized around the median. This validates the improvement in the quality of the CTG dataset after preprocessing with CFCM, facilitating further classification analysis with increased accuracy.

To address the issue of class imbalanced data in the dataset, the synthetic minority sampling technique (SMOTE) was used following the application of CFCM. Figure 6 visualizes the class distribution before and after the joint application of CFCM and SMOTE. Figure 5(a) illustrates the initial state of the CTG dataset, characterized by a class imbalanced data distribution, where the "normal" class is predominant compared to the "suspect" and "pathological" classes. Following the implementation of CFCM and SMOTE, as depicted in Figure 5(b), there is a notable enhancement in the balance of class distribution. This approach not only reduces the number of outliers, but also introduces synthetic samples for the minority classes, achieving a more equitable distribution among the three classes.

C. Classification Algorithms Comparison and Integrated Algorithm Evaluation (Stage 2)

In this evaluation stage, we compared four classification algorithms, namely: CART, RF, SVM and k-NN, before and after the integration of CFCM-SMOTE with the original CTG dataset. The results, documented in Tables V, VI, and VII, demonstrate significant improvements in precision, sensitivity, and specificity for each algorithm after the implementation of CFCM-SMOTE.

Table V shows that the CART algorithm excels in classifying the 'normal' class with very high precision (98.16%) and impressive sensitivity (95.82%) and specificity (85.81%), showcasing its strong capability in accurately identifying this class. The performance in the pathological and suspect classes was also quite good, confirming the effectiveness of CART without the CFCM-SMOTE intervention. In contrast, while the RF algorithm performed well, it did not reach the same levels of precision and sensitivity as CART for the pathological class. The SVM and k-NN algorithms displayed lower performance than CART and RF, with SVM's performance varying the most between classes and k-NN showing more consistent precision.

After applying CFCM, as documented in Table VI, there was a decrease in performance. The CART algorithm experienced a reduction in precision from 98.16% in the 'normal' class to 94.97%, indicating a decrease of 3.19%. Meanwhile, the sensitivity for the same class increased from 95.82% to 96.69%, an increase of 0.87%. Similarly, in the RF algorithm, the precision for the normal class increased from 94.77% to 97.24%, an increase of 2.47%, but decreased for the pathological class, showing that the effect of CFCM varied by class. SVM and k-NN algorithms showed improvements in some metrics but still exhibited performance instability, especially SVM, with a decrease in average precision to 39.01%.

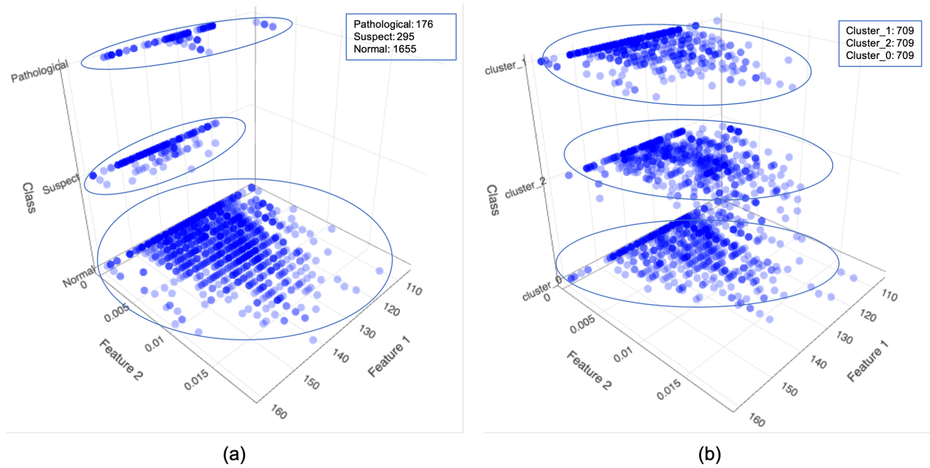

Figure 6. Distribution of (a) Classes Before and (b) After CFCM-SMOTE Implementation

TABLE V. Performance comparison of classification algorithms without CFCM-SMOTE

Algorihtms	Classes	PRC (%)	SNS (%)	SPC (%)	F1-S (%)	ACC (%)	EC	TE($_s$)
CART	Normal	98.16	95.82	85.81	96.98	**94.20**	**5.80**	**2.2598**
	Pathological	81.37	86.46	97.57	83.84			
	Suspect	80.43	92.50	99.49	86.05			
	AVERAGE	**86.65**	**91.59**	**94.29**	**88.96**			
RF	Normal	94.77	93.82	78.01	94.29	90.60	5.90	5.914
	Pathological	60.76	68.57	96.06	64.43			
	Suspect	95.16	89.39	98.78	92.19			
	AVERAGE	83.56	83.93	90.95	83.64			
SVM	Normal	97.35	88.33	57.43	92.62	85.74	14.26	13.66
	Pathological	38.24	61.90	95.52	47.27			
	Suspect	67.39	88.57	99.32	76.54			
	AVERAGE	67.66	79.60	84.09	72.14			
k-NN	Normal	97.55	92.28	72.97	94.84	90.44	9.56	4.259
	Pathological	64.71	79.52	96.83	71.35			
	Suspect	71.74	89.19	99.32	79.52			
	AVERAGE	78.00	87.00	89.71	81.90			

These results highlight that although CFCM has the potential to improve outlier identification, it does not universally enhance performance on the class imbalanced data present in the CTG dataset, thus degrading algorithm performance. Based on the results of the application of CFCM, it is crucial to adopt a holistic approach that considers outlier handling techniques and class balancing strategies to optimize performance in analyzing complex data, such as CTG datasets.

Integration of CFCM-SMOTE, as illustrated in Table VII, resulted in significant improvements in algorithm performance, notably for the CART algorithm, which achieved almost perfect precision (99.58%) and high accuracy (99.84%). This indicates that the integration of CFCM-SMOTE significantly enhanced CART's ability to identify each class with high precision and a very low error rate. The RF algorithm also showed a highly significant improvement in accuracy and precision, confirming that CFCM-SMOTE effectively strengthened the RF's classification capabilities for CTG data. Although the SVM and k-NN algorithms showed improvements in precision and accuracy, they still faced some limitations, particularly the SVM, in terms of sensitivity and specificity. On the contrary, k-NN demonstrated more stable results and lower error rates.

D. Statistical Difference Test

In this stage, we evaluate the significant performance differences between four distinct classification algorithms—CART integrated with CFCM-SMOTE, RF integrated with CFCM-SMOTE, SVM integrated with CFCM-SMOTE, and k-NN integrated with CFCM-SMOTE using the value F1 as the primary evaluation indicator. The results

TABLE VI. Impact of CFCM implementation on classification algorithm performance without SMOTE

Algorithms	Classes	PRC (%)	SNS (%)	SPC (%)	F1-S (%)	ACC (%)	EC	TE($_s$)
CART	Normal	94.97	96.69	81.91	95.82	**93.19**	**6.81**	3.171
	Pathological	84.21	91.43	98.47	87.67			
	Suspect	86.00	72.88	98.09	78.90			
	AVERAGE	**88.39**	87.00	**92.82**	87.46			
RF	Normal	97.24	93.79	97.24	79.00	92.72	10.39	13.668
	Pathological	67.19	87.76	67.19	98.34			
	Suspect	97.22	89.74	97.22	98.97			
	AVERAGE	87.22	**90.43**	87.22	**92.10**			
SVM	Normal	100.00	79.81	100.00	10.64	79.81	6.1	105.53
	Pathological	8.57	75.00	8.57	99.74			
	Suspect	8.47	83.33	8.47	99.73			
	AVERAGE	39.01	79.38	39.01	70.04			
k-NN	Normal	97.59	94.74	97.59	80.85	91.78	8.22	**2.259**
	Pathological	77.14	90.00	77.14	99.23			
	Suspect	67.80	74.07	67.80	96.19			
	AVERAGE	80.84	86.27	80.84	92.09			

TABLE VII. Performance improvement of classification algorithm after CFCM-SMOTE integration

Algorihthms	Classes	PRC (%)	SNS (%)	SPC (%)	F1-S (%)	ACC (%)	EC	TE($_s$)
CART	Normal	99.58	100	99.75	99.79	**99.84**	**0.16**	**0.783**
	Pathological	100	99.47	100.00	99.74			
	Suspect	100	100	100	100			
	AVERAGE	**99.86**	**99.82**	**99.92**	**99.84**			
RF	Normal	95.89	97.09	98.60	96.49	97.45	2.55	1.428
	Pathological	96.6	96.27	98.07	96.46			
	Suspect	99.80	99.00	99.50	99.40			
	AVERAGE	97.45	97.45	98.72	97.45			
SVM	Normal	99.79	70.95	80.16	82.94	86.44	13.56	11.016
	Pathological	76.77	99.49	99.80	86.67			
	Suspect	83.23	99.76	99.90	90.75			
	AVERAGE	86.60	90.07	93.29	86.79			
KNN	Normal	94.25	98.92	99.50	96.53	91.18	2.82	26.18
	Pathological	98.43	93.98	96.74	96.15			
	Suspect	98.79	98.99	99.50	98.89			
	AVERAGE	97.16	97.30	98.58	97.19			

of this empirical investigation are detailed in Tables VIII, IX, and X.

Implementing the Friedman test yielded a Chi-square value of 8.1, with a degree of freedom (df) of 3, and a p-value of 0.044, indicating that there is a statistically significant difference between at least one pair of the algorithms evaluated. In-depth analysis is recorded in Table VIII.

Furthermore, pairwise comparison analysis was applied using the Dunn-Bonferroni post hoc method and is shown in Tables IX and X. Both tables show that there is a significant difference between the CART algorithm with the addition of CFCM-SMOTE and the SVM and k-NN algorithms that also use CFCM-SMOTE, indicated by a p-value of 0.018. However, the comparison between the CART algorithm with CFCM-SMOTE and the RF algorithm with CFCM-SMOTE, as well as between the RF algorithm with CFCM-SMOTE and the SVM and k-NN algorithms with CFCM-

TABLE VIII. Friedman statistical test results between four different algorithms based on CFCM-SMOTE

Chi^2	df	p-value
8.1	3	0.044

SMOTE, did not show statistical significance, with the *p*-value exceeding the threshold of 0.05.

In the average rank evaluation, the CART algorithm integrated with CFCM-SMOTE received the highest rank with an average of 4, while the SVM and k-NN algorithms integrated with CFCM-SMOTE received a lower average rank of 1.5.

Overall, the findings confirm that the application of the CFCM-SMOTE strategy improves the prediction accuracy of conventional classification algorithms. However, it is crucial to emphasize that the efficiency of the SMOTE algorithm itself still requires further optimization by adjusting certain parameters, such as the number of nearest neighbors or sampling strategy, especially when faced with datasets with pronounced class imbalanced data. This conclusion is reinforced by research conducted by Q. Chen et al. [20], who highlighted the potential to improve the performance of the SMOTE algorithm with appropriate parameter adjustments.

In the context of datasets containing outliers and class imbalanced data, selecting an effective classification strategy is key, as evidenced by the performance metrics reported: precision, sensitivity, specificity, F1-score, and accuracy. Therefore, this study concludes that, for large CTG dataset that contains outliers and class imbalanced data, the proposed approach, specifically the integration of CFCM-SMOTE with the CART algorithm, shows superiority in fetal health classification compared to alternative algorithms evaluated.

E. Proposed Algorithm vs Previous Studies (Stage 3)

Utilizing a Cardiotocography (CTG) dataset identical to those used in previous studies, this research ultimately provides a comparative analysis between the proposed algorithm and several previous researches. Table XI details this comparison extensively.

Table XI provides a robust comparison of the accuracy percentages achieved by our proposed algorithm compared to those reported in previous research. Chen & Yin [3] implemented AdaBoost combined with Random Forest, achieving an accuracy of 97.6%, while Mehbodniya et al. [8] reported an accuracy of 94. 5% using normalization and random forest (Norm+RF). Duhayyim et al. [10] utilized SMOTE+AdaBoost to achieve a remarkable accuracy of 99%, and Raihen & Akter [11] achieved an accuracy of 85.98% using bootstrap aggregating (Bagging). Additionally, Salini et al. [12] documented a 93% accuracy with feature selection and Random Forest (FS+RF). In stark contrast, our study advances the integration of CFCM-SMOTE with various machine learning models, demonstrating that the combination of CFCM-SMOTE and the CART algorithm leads to the highest observed accuracy of 99.84%. The amalgamation of CFCM-SMOTE with RF yielded an accuracy of 97.45%, while integrating CFCM-SMOTE with SVM and k-NN algorithms resulted in accuracies of 86.44% and 91.18%, respectively. These findings underscore that our approach, particularly the CFCM-SMOTE with CART, significantly elevates the classification efficacy of CTG data compared to prior efforts. Although studies like that of Das et al. [9] employ similar data formats and features from the CTG dataset, their results are not directly comparable due to differences in validation techniques, specifically their use of 5-fold cross-validation and distinct data sources.

5. DISCUSSIONS

This research thoroughly investigates the effectiveness of combining conditional fuzzy C-means (CFCM) and synthetic minority sampling technique (SMOTE) methods within different classification algorithms for evaluating fetal health through cardiotocography (CTG) data sets. The significant enhancement in algorithm performance, as demonstrated in Tables IV to VI, emphasizes the advantages of employing the CFCM-SMOTE strategy and accentuates the importance of its incorporation into data analysis processes.

Initially, the performance metrics of classification algorithms without CFCM-SMOTE integration showed significant improvements after integration. The CART algorithm exemplified the success of the machine learning algorithm in addressing class imbalanced data, and its consistent performance suggests a synergistic relationship with the CFCM algorithms, which merits further investigation. In addition, the robustness of the RF algorithm was confirmed, while the results for the SVM and k-NN algorithms emphasized the need for careful parameter tuning, especially in clinical settings.

Marked class imbalanced data presented a notable challenge in algorithm optimization, particularly for underrepresented classes. However, the application of the CFCM-SMOTE algorithm has effectively improved precision, sensitivity, and specificity, especially in the CART algorithm, though the challenge of optimizing for underrepresented classes persists, as the skewed data distribution suggests.

The extensive statistical analysis conducted, particularly the significant results from the Dunn-Bonferroni test, provides strong evidence of the superiority of the CFCM-SMOTE+CART algorithm. This quantitative validation is crucial for the practical implementation of this methodological approach. According to research by Roh et al. [21] and Sun et al. [22], this study validates the effectiveness of the CFCM algorithm in managing outliers and in feature selection. It recognizes that traditional learning algorithms often do not generalize well to minority data, which advocates the generation of synthetic samples using the SMOTE technique to prevent overfitting.

The integration of CFCM-SMOTE with CART algorithms significantly improves performance in classification tasks involving CTG datasets. The findings of this study lay the groundwork for its further application and development in clinical settings, opening up new directions for future research in medical machine learning and providing

TABLE IX. Friedman statistical test results for performance comparison between four different algorithms based on CFCM-SMOTE for F1-measure results

Symbol	Algorithms	N	Mean	Median	Standard Deviation	Ranks
A	CART	3	99.84	99.79	0.14	4
B	RF	3	97.45	96.49	1.69	3
C	SVM	3	86.79	86.67	3.91	1.5
D	KNN	3	86.79	86.67	3.91	1.5

TABLE X. Paired comparison test results using dunn-bonferroni based on CFCM-SMOTE

Algorithms	Test Statistics	Standard Error	Std. Test Statistics	p-value	Difference
A – B	1	1.05	0.95	0.343	No significant
A – C	2.5	1.05	2.37	0.018	Significant
A – D	2.5	1.05	2.37	0.018	Significant
B – C	1.5	1.05	1.42	0.155	No significant
B – D	1.5	1.05	1.42	0.155	No significant
C – D	0	1.05	0	1	No significant

TABLE XI. Accuracy comparison with prior studies

Researchers	Algorithms	ACC (%)
Chen & Yin [3]	AdaBoost+RF	97.6
Mehbodniya et al. [8]	Norm+RF	94.5
Duhayyim et al. [10]	SMOTE+AdaBoost	99
Raihen & Akter [11]	Bagging	85.98
Salini et al. [12]	FS+RF	93
Our study	CFCM-SMOTE+CART	99.84
Our study	CFCM-SMOTE+RF	97.45
Our study	CFCM-SMOTE+SVM	86.44
Our study	CFCM-SMOTE+kNN	91.18

direct benefits to prenatal care, with a profound impact on maternal and fetal health.

Although our proposed algorithm outperforms others, it has several limitations that need attention:

1) **Implementation Complexity:** The integration of CFCM-SMOTE with the CART algorithm leads to improved performance but also increases implementation complexity, which can pose challenges in clinical settings where quick deployment and user-friendliness are critical.
2) **Outlier Management:** Although the CFCM algorithm has proven effective in managing outliers and selecting features, its approach to outlier management still requires strengthening. A more robust strategy is essential to ensure the algorithm's ability to handle extreme variations in CTG data that could impact predictions of fetal health.
3) **Dependence on Data Quality:** The accuracy and performance of the algorithm are highly dependent on the quality and representativeness of the data utilized. Imbalanced data can undermine the algorithm's applicability to broader populations or scenarios different from those reflected in the initial training set.
4) **Need for Broader Validation:** Despite extensive quantitative validation, additional validation is necessary to confirm that the algorithm can be implemented effectively and broadly across various clinical environments and within more diverse populations.

6. CONCLUSIONS AND FUTURE RESEARCH

In this research, we conducted a comprehensive examination of the integration of cluster-based fuzzy C-means (CFCM) and synthetic minority oversampling technique (SMOTE) to assess the categorization of fetal health in a multiclass scenario using cardiotocography (CTG) data analysis. This scientific effort addresses the most significant challenges in managing outliers and class imbalanced in CTG data analysis. The empirical evidence in this study indicates that the integration of CFCM to reduce the impact of outliers and SMOTE to address class imbalanced data significantly improves the efficacy of the proposed classification algorithms. This methodology specifically improves the precision, sensitivity, specificity, F1 score, and overall accuracy of the algorithms compared to their performance without preprocessing. These findings substantially contribute to the development of methodologies for evaluating fetal health classification, and reinforce the importance of preprocessing strategies in developing more robust and accurate predictive algorithms.

Despite these advances, future research has great potential to expand on the methodologies and findings of this study. One promising direction is to explore the integration of CFCM-SMOTE with more advanced machine learning techniques. Incorporating feature selection presents a strong alternative, where research on the challenges of feature selection in CTG data, which often exhibits diverse structures, has not yet been addressed. Although the CFCM technique effectively reduces the impact of outliers, it inadvertently generates new outliers within the majority class by selecting the maximum diversity from the majority data group. Therefore, in future works, we will evaluate and compare the proposed methodology with other clustering algorithms, such as DBSCAN and k-means, as well as different imputation algorithms, including K-Nearest Neighbor imputation and Multivariate Imputation by Chained Equations (MICE), to ascertain their relative effectiveness.

Additionally, comparative studies with alternative data balancing techniques beyond SMOTE could help refine the approach to managing class imbalances and improving the performance of classification algorithms. Another important area for future work could include developing and testing real-time CTG data analysis systems in clinical settings so that immediate feedback is provided to healthcare professionals and improves the timeliness of interventions. This practical application could benefit significantly from further exploring feature selection processes to optimize models' computational efficiency and effectiveness in varied and dynamic clinical environments. Finally, more extensive validation could also be performed in diverse demographic and geographic populations to confirm the generalizability and robustness of the CFCM-SMOTE methodology, ensuring its effective adaptation to global healthcare systems. This effort will not only elucidate the strengths of existing methods but also extend the application of these techniques more broadly in health data analysis. In summary, the significant contribution of this study to the field of fetal health classification modeling lays the foundation for future research that could revolutionize predictive analytics in prenatal care.

Disclosure Statement

The authors have declared that there are no competing interests.

Research Contributing

Research design: AI, AK; Data gathering and exploration: AF, RDR, SAJ, DAL, AA.AR; Analysis, evaluation, and selection of the best algorithms: LK, AM, AKO; Conclusion and recommendation: AI; Writing, review, and editing: AI, AKO.

Funding

This research was supported by Universitas Muhammadiyah Semarang, Grant no. 107/UNIMUS.L/PG/PJ.INT/2023.

References

[1] A. Ilham, J. Silva, N. Mercado-Caruso, D. Tapias-Ruiz, and O. B. P. Lezama, "Impact of class imbalance on convolutional neural network training in multi-class problems," in *Image Processing and Capsule Networks: ICIPCN 2020*. Springer, 2021, pp. 309–318.

[2] S. Cairo Notari, J. Sader, N. Caire Fon, J. M. Sommer, A. C. Pereira Miozzari, D. Janjic, M. Nendaz, and M.-C. Audétat, "Understanding gps' clinical reasoning processes involved in managing patients suffering from multimorbidity: a systematic review of qualitative and quantitative research," *International journal of clinical practice*, vol. 75, no. 9, p. e14187, 2021.

[3] M. Chen and Z. Yin, "Classification of cardiotocography based on the apriori algorithm and multi-model ensemble classifier," *Frontiers in Cell and Developmental Biology*, vol. 10, p. 888859, 2022.

[4] S. Blackburn, *Maternal, Fetal, Neonatal Physiology: A Clinical Perspective*. Elsevier Health Sciences (5th ed.), 2016.

[5] 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, p. 2471, 2023.

[6] S. Das, H. Mukherjee, S. M. Obaidullah, K. Roy, and C. K. Saha, "Ensemble based technique for the assessment of fetal health using cardiotocograph–a case study with standard feature reduction techniques," *Multimedia Tools and Applications*, vol. 79, no. 47, pp. 35 147–35 168, 2020.

[7] R. Kapila and S. Saleti, "Optimizing fetal health prediction: Ensemble modeling with fusion of feature selection and extraction techniques for cardiotocography data," *Computational Biology and Chemistry*, vol. 107, p. 107973, 2023.

[8] A. Mehbodniya, A. J. P. Lazar, J. Webber, D. K. Sharma, S. Jayagopalan, K. K, P. Singh, R. Rajan, S. Pandya, and S. Sengan, "Fetal health classification from cardiotocographic data using machine learning," *Expert Systems*, vol. 39, no. 6, p. e12899, 2022.

[9] 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, p. 858, 2023.

[10] M. Al Duhayyim, S. Abbas, A. Al Hejaili, N. Kryvinska, A. Almadhor, and H. Mughal, "Ensemble learning for fetal health classification." *Computer Systems Science & Engineering*, vol. 47, no. 1, 2023.

[11] M. N. Raihen and S. Akter, "Comparative assessment of several effective machine learning classification methods for maternal health risk," *Computational Journal of Mathematical and Statistical Sciences*, vol. 3, no. 1, pp. 161–176, 2024.

[12] Y. Salini, S. N. Mohanty, J. V. N. Ramesh, M. Yang, and M. M. V. Chalapathi, "Cardiotocography data analysis for fetal health classification using machine learning models," *IEEE Access*, 2024.

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

[14] D. Ayres-de Campos, J. Bernardes, A. Garrido, J. Marques-de Sa, and L. Pereira-Leite, "Sisporto 2.0: a program for automated analysis of cardiotocograms," *Journal of Maternal-Fetal Medicine*, vol. 9, no. 5, pp. 311–318, 2000.

[15] J. Kaliappan, A. R. Bagepalli, S. Almal, R. Mishra, Y.-C. Hu, and K. Srinivasan, "Impact of cross-validation on machine learning models for early detection of intrauterine fetal demise," *Diagnostics*, vol. 13, no. 10, p. 1692, 2023.

[16] A. Wijaya, T. B. Adji, and N. A. Setiawan, "Logistic regression based feature selection and two-stage detection for eeg based motor imagery classification." *International Journal of Intelligent Engineering & Systems*, vol. 14, no. 1, 2021.

[17] A. Ilham, R. S. Wahono, C. Supriyanto, and A. Wijaya, "U-control chart based differential evolution clustering for determining the number of cluster in k-means," *evolutionary computation*, vol. 307, 2019.

[18] A. Ilham, L. Assaffat, L. Khikmah, S. Safuan, and S. Suprapedi, "k-means cluster-based random undersampling and meta-learning approach for village development status classification," *JOIV: International Journal on Informatics Visualization*, vol. 7, no. 2, pp. 542–547, 2023.

[19] J. Demšar, "Statistical comparisons of classifiers over multiple data sets," *The Journal of Machine learning research*, vol. 7, pp. 1–30, 2006.

[20] Q. Chen, Z.-L. Zhang, W.-P. Huang, J. Wu, and X.-G. Luo, "Pf-smote: A novel parameter-free smote for imbalanced datasets," *Neurocomputing*, vol. 498, pp. 75–88, 2022.

[21] S.-B. Roh, S.-K. Oh, W. Pedrycz, K. Seo, and Z. Fu, "Design methodology for radial basis function neural networks classifier based on locally linear reconstruction and conditional fuzzy c-means clustering," *International Journal of Approximate Reasoning*, vol. 106, pp. 228–243, 2019.

[22] J. Sun, H. Li, H. Fujita, B. Fu, and W. Ai, "Class-imbalanced dynamic financial distress prediction based on adaboost-svm ensemble combined with smote and time weighting," *Information Fusion*, vol. 54, pp. 128–144, 2020.

Ahmad Ilham (AI) is an Indonesian citizen, with an academic background in Computer Science and Informatics Engineering. He is an alumnus of Al Assariah Mandar University (UNASMAN), where he obtained his Bachelor's degree in Information Systems, and Dian Nuswantoro University (UDINUS), where he obtained his Master's degree in Intelligent Systems. He is currently spreading his knowledge as a lecturer at the Informatics Department of Universitas Muhammadiyah Semarang, Indonesia. Beyond his academic commitments, he is an honoured member of the International Association of Engineers (IAENG) and actively contributes to Data Science Indonesia (DSI). His research interests include data science, machine learning, and intelligent information processing.

Asdani Kindarto (AK) is working as a Lecturer in the Department of Informatics at Universitas Muhammadiyah Semarang, Semarang, Indonesia. He obtained his Ph.D. from National Taiwan University of Science and Technology, Taiwan. His research interests include E-Government, Machine Learning and Social Networking Analysis.

Akhmad Fathurohman (AF) is working as a Senior Lecturer in the Department of Informatics at Universitas Muhammadiyah Semarang, Semarang, Indonesia. His research interest includes Software and Network Engineering.

Laelatul Khikmah (LK) is working as a lecturer at the Institut Teknologi Statistika dan Bisnis Muhammadiyah Semarang (ITESA). She is also an active member of the Intelligent Data Science Research Group (IDSRG) as a data scientist researcher. Her research interests includes data science, categorical data analysis, and non-parametric statistics.

Rima Dias Ramadhani (RDR is a graduate of Universitas Dian Nuswantoro majoring in Informatics Engineering. Currently, she works as a lecturer in the Informatics Department, Universitas Muhammadiyah Semarang, Semarang, Indonesia. His research interests include intelligent systems, data mining, and machine learning.

Dhewi April Liana (DAL) is a student in the Department of Informatics at Universitas Muhammadiyah Semarang. She is also a junior research assistant in the Intelligent Data Science Research Group (IDSRG) as a junior data scientist. Her current research interests are data mining and intelligent systems.

Aura Amylia. AR (AA.AR) is a student in the Department of Informatics at Universitas Muhammadiyah Semarang. She is also a junior research assistant in the Intelligent Data Science Research Group (IDSRG) as a junior data scientist. Her current research interests are data mining and intelligent systems.

Syafrie Abdunnasir Jawad (SAJ) is a student and research assistant at the Department of Informatics, Universitas Muhammadiyah Semarang. His current research interests are data mining and intelligent systems.

Ahmed Kareem Oleiwi (AKO) is working as a Lecturer in the Islamic University, Najaf, Iraq. He obtained his Ph.D. from Zhengzhou University, Zhengzhou, Henan, China. His research interests includes Machine Learning, Deep learning, Data Discovery.

Astri Mutiar (AM) received the B.N. degree in Nursing Science from Universitas Padjadjaran, Indonesia, in 2013, and the M.S. degree in 2018 from National Cheng Kung University, Taiwan. From 2019 to the present, she has been actively working as a lecturer at the Sekolah Tinggi Ilmu Keperawatan PPNI Jawa Barat, Indonesia. Her research interests includes fetal health monitoring and nursing science.