



Research Article

A hybrid model of random forest ensemble and resample for cardiocography data classification

R. HEPHZIBAH^{1,*}, A. Hepzibah CHRISTINAL^{1,*}, Deepthy Mary ALEX², R. JAYANTHI¹,
D. Abraham CHANDY¹, Chandrajit BAJAJ³

¹Department of Mathematics, Karunya University, Coimbatore, 641114, India

²Department of Electronics and communication Engineering, Mangalam College of Engineering, Kerala, 686631, India

³Department of Computational Applied Mathematics, University of Texas, Austin, Texas, 78712, USA

ARTICLE INFO

Article history

Received: 14 January 2024

Revised: 09 March 2024

Accepted: 06 April 2024

Keywords:

Bagging Classifier;
Cardiocography (CTG);
Classification; Fetal Health;
Random Forest Ensemble;
Resampling; SmoteTomek

ABSTRACT

Fetal health monitoring is essential as it leads to increased mortality rates in fetuses. Cardiocography is a medical technique used by obstetricians to monitor fetal health during labor, particularly in cases involving complications. Though various works have been carried out in the classification of CTG data there seems to be a need for improvement in achieving significant accuracy levels. In this work, first, we implemented the Smote Tomek sampling technique to create a balanced dataset. Then, the balanced data is employed for classification in the Random Forest ensemble with a bagging classifier. Our technique's performance is assessed using metrics including accuracy, precision, recall, and F1-score. Experimental findings reveal our method achieves an accuracy of 98.5%, outperforming not only other classifiers examined in the study but also surpassing deep learning algorithms. Hence, the findings of our study highlight the effectiveness of our approach in classifying Cardiocography data, suggesting the potential for enhancing fetal health monitoring during labor and for improved obstetric care.

Cite this article as: Hepzibah R, Christinal AH, Alex DM, Jayanthi R, Chandy DA, Bajaj C. A hybrid model of random forest ensemble and resample for cardiocography data classification. Sigma J Eng Nat Sci 2025;43(2):452–462.

INTRODUCTION

In recent years, perinatal mortality rates have notably declined, attributed to heightened awareness of perinatal physiology and pathology, the expansion of regional centers, and the implementation of intrapartum electronic monitoring for in-utero surveillance. The intrapartum electronic fetal monitoring grows progressively and it helps in reducing morbidity and mortality rates,

especially in high-risk pregnancies [1]. Fetal Heart Rate tracking would address two issues. In the beginning, it would act as a screening procedure for severe asphyxia. Second, FHR monitoring would enable the early detection of asphyxia, enabling prompt obstetric intervention to prevent asphyxia-induced brain damage or the death of the newborn [2]. CTG involves continuous monitoring of the fetal heart rate via an ultrasound transducer

*Corresponding author.

*E-mail address: hephzibahrajasekaran@gmail.com, christyhep@gmail.com

This paper was recommended for publication in revised form by
Editor-in-Chief Ahmet Selim Dalkilic



positioned on the mother's abdomen. Primarily employed throughout pregnancy, it serves to evaluate fetal well-being, particularly in cases of significant complications [3]. An essential component of antepartum and postpartum care is the routine recording of Fetal Heart Rate (FHR) using CTG monitoring. However, it became clear from several randomized trials that using CTG antenatally to improve fetal outcomes has only modest effectiveness. A comprehensive meta-analysis of studies examining intrapartum cardiotocography found a 50% decrease in perinatal mortality but a 2.5-fold rise in surgical interventions [4]. A cardiotocograph, generating a paper trace, electronically records both the newborn's heart rate and the mother's uterine contractions. This involves two transducers—one linked to a recording device and the other to a Doppler ultrasound transducer.

External CTG monitoring is commonly continuous during labor, although intermittent usage is also practiced [5]. On CTG, four crucial features require analysis: the baseline rate, variability, accelerations, and decelerations. The normal range of the baseline rate is 110-160bpm and it is analyzed between 5 to 10mins. Variability is the bandwidth variation of the baseline after eliminating the decelerations and accelerations. The Acceleration is a short time improvement in the baseline of >15 bpm, enduring for 15 s or more and restoring to the normal baseline. Decelerations are defined as a temporary decrease in the fetal heart rate of more than 15 beats per minute, lasting for over 15 seconds [6]. Machine learning algorithms primarily classify CTG recordings as either normal or pathological based on the attributes of fetal heart rate and uterine contractions, aiding clinicians in decision-making. [7]. Different algorithms including the Naïve Bayes classifier, Decision tree (DT), Radial basis function, and Multilayer perceptron networks are utilized for classifying the CTG dataset [8]. Due to the imbalance in our CTG dataset, various sampling techniques like the Synthetic Minority Oversampling Technique (SMOTE) are employed to improve the classification performance of CTG data [9].

The paper is structured as follows: Section 2 reviews related works on CTG data classification. Section 3 outlines the methodology, while Section 4 details our proposed method. Experimental results are presented in Section 5, followed by a discussion of the findings in Section 6, and concluding remarks in Section 7. The key innovation of our method lies in integrating the Smote Tomek sampling technique with a Random Forest ensemble and bagging classifier for CTG data classification. This hybrid approach, utilizing ensemble learning with Random Forest as the base classifier and employing the bagging technique, enhances model performance by reducing variance, improving generalization, and offering insights into feature importance. It effectively tackles the challenge of imbalanced datasets often encountered in medical data analysis, particularly in fetal health monitoring during labor.

Related Works

A hybrid model is implemented by integrating the Least square support vector machine with binary decision tree and the particle swarm optimization and it achieved an accuracy of 91.62% [10]. A new approach is proposed by using the genetic algorithm and Support Vector Machine (SVM) in evaluating fetal well-being and has a very good evaluating performance [11]. The bagging approach is used with decision tree algorithms such as random forest, REPTree, and J48 for the categorization of CTG data and achieves an accuracy of more than 90% by all classifiers [12]. The decision tree, discriminant analysis, and artificial neural networks are used in predicting fetal distress in CTG data and achieved 86.36%, 82.1%, and 97.78% accuracy respectively [13]. A sub-adaptive neuro-fuzzy inference system, deep-adaptive Neuro-Fuzzy Inference Systems (ANFIS) models, and Deep Stacked Sparse AutoEncoders (DSSAE) implemented in a CTG dataset and it is found that DSSAE performs very well with greater sensitivity, specificity, and accuracy in the CTG dataset [14]. A simulation of the rough neural networks is introduced in the classification of the CTG dataset and it also achieves a very good accuracy of 92.95% [15]. The technique which uses a firefly algorithm helps in feature selection with the naïve Bayesian classifier and achieves an accuracy of 86.54%. In this technique, the firefly algorithm helps in selecting the best subset features for the dataset [16]. A new shallow architecture of 1D Convolutional Neural Network (CNN) is proposed and it consists of only one convolutional layer which helps in reducing the complexity of computation. It is very helpful to evaluate the fetal state assessment and this performs very well than the traditional CNN in the classification and achieves a higher accuracy [17].

The boost ensemble methods along with the various classifiers such as the Random Forest, AdaBoost, k-Nearest Neighbors, Support Vector Machine, and Decision Trees, are introduced and it was found that the performance of classifiers is improved in the classification of CTG data [18]. The ensemble methods such as bagging and boosting are implemented in the various classifiers such as naïve Bayes, Decision tree, random forest, and k-nearest neighbors and achieved the best accuracy of 96.175 with the help of the random forest classifier [19]. The R programming is utilized for classifying fetal data which employs a random forest classifier and achieves 99.94% training accuracy and 93.57% testing accuracy [20]. This study introduces a novel ensemble approach, combining sampling techniques like SMOTE and ENN with Random Forest classification, achieving 93% accuracy in CTG data classification. Addressing data imbalance through oversampling and undersampling, enhances classification performance effectively [21]. A novel ensemble classifier merging XGBoost and Random Forest to predict fetal state from CTG data maintains 96% accuracy [22].

MATERIALS AND METHODS

Dataset Description

The CTG dataset, extracted from the University of California Irvine Machine Learning (ML) repository, consists of 2126 instances featuring 23 distinct attributes. Expert clinicians have classified the dataset, containing measurements of uterine contractions and fetal heart rates obtained through CTG, into three classes: Suspect, Pathology, and Normal, with 1655, 295, and 176 instances respectively. This publicly accessible dataset is partitioned into training and testing sets in a 70:30 ratio for classification purposes [23].

Resolving Data Imbalance Using Sampling Techniques

The complication in considering the clinical dataset for experimentation is the imbalance in the dataset. The imbalance occurs when the dataset comprises more classes in the majority and very few classes in the minority [24]. These imbalances in data decrease the performance of the algorithms leading to incorrect classification performance [25]. This is because the classifiers have a bias toward the majority data while the minority class is considered noise and is avoided. It leads to greater misclassification in minority datasets. To overcome all these problems

sampling techniques are introduced [26]. The CTG dataset obtained from the ML repository consists of 2126 instances with 23 features. The dataset includes 1655 instances classified as normal, 295 as suspect, and 176 as pathology classes, indicating an evident imbalance in the dataset. Hence, we applied the following techniques to get balanced data for our classification [27]. The main motive of the sampling technique is to balance the dataset where there is an imbalance. Two sampling techniques, namely oversampling and under-sampling, were employed. Given the data imbalance observed in our CTG dataset, we applied the following methodologies.

Oversampling Techniques

Oversampling techniques are employed to address the issue of dataset imbalance by producing a greater count of training samples for the minority classes. These techniques encompass Random Oversampling, Synthetic Minority Oversampling Technique (SMOTE), Borderline SMOTE, ADASYN, and Safe-level SMOTE, among others. In this study, we applied three oversampling techniques: Random Oversampling, SMOTE, and Borderline SMOTE. [28].

Random Oversampling

In this Random oversampling technique, the randomly selected samples from the minority class are duplicated to balance the dataset. The benefit of this technique is that there is no loss of details but may cause overfitting since it makes a similar copy of minority class samples [29].

Synthetic minority oversampling technique (SMOTE)

SMOTE is an Oversampling technique that increases the minority class by generating synthetic samples. These additional data points are created based on existing real data, with synthetic samples inserted along lines connecting the nearest neighbors of the minority class. Neighbors are selected randomly as needed [30]. While SMOTE's primary advantage lies in its ability to generate diverse synthetic instances of the minority class, differing from substitution methods in other techniques, its main drawback is its disregard for neighboring class examples and vulnerability to data intricacy [31].

Borderline-smote

Contrary to SMOTE, this technique generates minority samples besides the borderline and its k-nearest neighbors [29,32]. It is also classified as borderline-SMOTE1 and borderline-SMOTE2. It is also classified as borderline-SMOTE1 and borderline-SMOTE2. It has the advantage of sampling only the minority class's boundary instances, which helps minimize operation time, but it lacks customized policies for dealing with other types of border instances [31]

Under-sampling technique

Under-sampling techniques aim to reduce the number of samples in the majority class. By decreasing the size of the majority class set A, these techniques help alleviate the

Table 1. Description of dataset features

S.NO	Attributes	Details
1.	B	Initiate instant
2.	E	Terminate instant
3.	LB	Baseline Value
4.	AC	Acceleration Count
5.	FM	Fetal Movement
6.	UC	Uterine contractions
7.	DL	Mild decelerations
8.	DS	Severe decelerations
9.	DP	Prolongued decelerations
10.	ASTV	Abnormal short-term variation percentage
11.	MSTV	Short-term variation Mean
12.	ALTV	Abnormal long-term variation percentage
13.	MLTV	Long-term variation Mean
14.	Width	Histogram Width
15.	Min	Low frequency
16.	Max	High frequency
17.	Nmax	Histogram peak count
18.	Nzeros	Histogram Zero Count
19.	Mode	Histogram mode
20.	Mean	Histogram mean
21.	Median	Histogram median
22.	Variance	Histogram variance
23.	Tendency	Histogram Tendency

skewed distribution between A and the minority class set I. In this study, we applied three under-sampling techniques: Random Under-sampling, Tomek Links Under-sampling, and Edited Nearest Neighbors (ENN) Rule.

Random under-sampling

It is the simplest technique where the randomly selected samples are eliminated from the majority class which may lead to information loss [34]. It has the advantage of avoiding excess majority samples and shortening the prediction model’s training time, but the main limitation of this technique is that it removes the useful data that is needed for the induction process [35].

Tomek links under-sampling

It is a method of under-sampling developed by Tomek which is an improvement of the Nearest-Neighbor Rule. Let u be an instance of class P and v an instance of class Q. Let $d(u, v)$ be the distance between u and v . (u, v) is a T-Link if for any instance t , $d(u, v) < d(u, t)$ or $d(u, v) < d(v, t)$. If any two examples are T-Link then one will be noise and the other will be the examples that are located on the boundary of the classes [36]. It also improves the data

balance by eliminating large samples that form “Tomek link pairs” with each other in the dataset, but it ignores small samples and abnormal points in the data, and the sample distribution balancing is also limited [37].

Edited Nearest Neighbors Rule (ENN)

If the minority class has three nearest neighbors ($k=3$) belonging to the majority class, these neighbors are excluded, along with the majority class samples found in the boundary [38]. Despite removing more instances compared to Tomek links, ENN is efficient in eliminating misclassified cases from all classes based on the consensus of the three nearest neighbors. However, undersampling with ENN may sometimes discard potentially valuable instances from clinical datasets [39].

Figure 1 and Figure 2 represent the illustration of over-sampling and undersampling techniques on a given dataset. The first box represents the distribution of the original dataset where each color represents the data in each class. Blue represents the normal, white represents the suspect and the red represents the pathology. The next three boxes represent the significant change in the number of samples

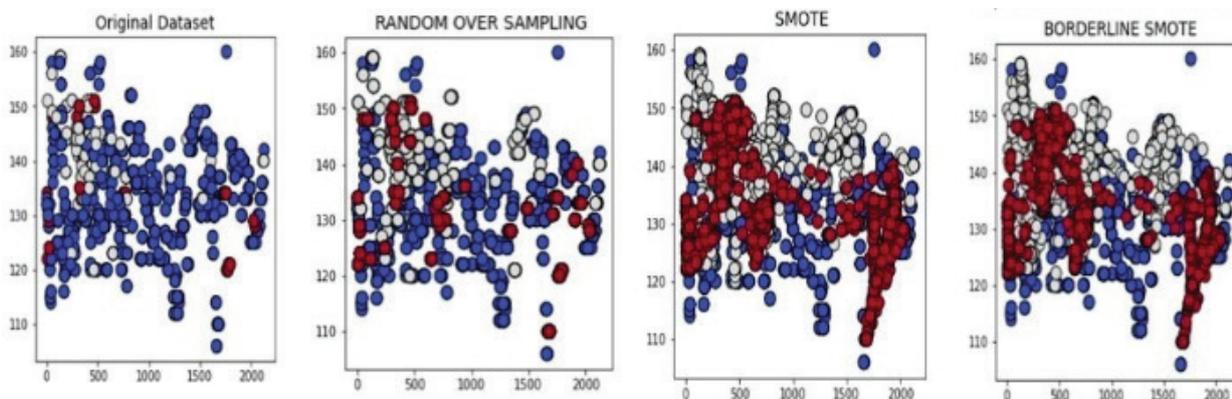


Figure 1. Visualizing over-sampling techniques for dataset representation.

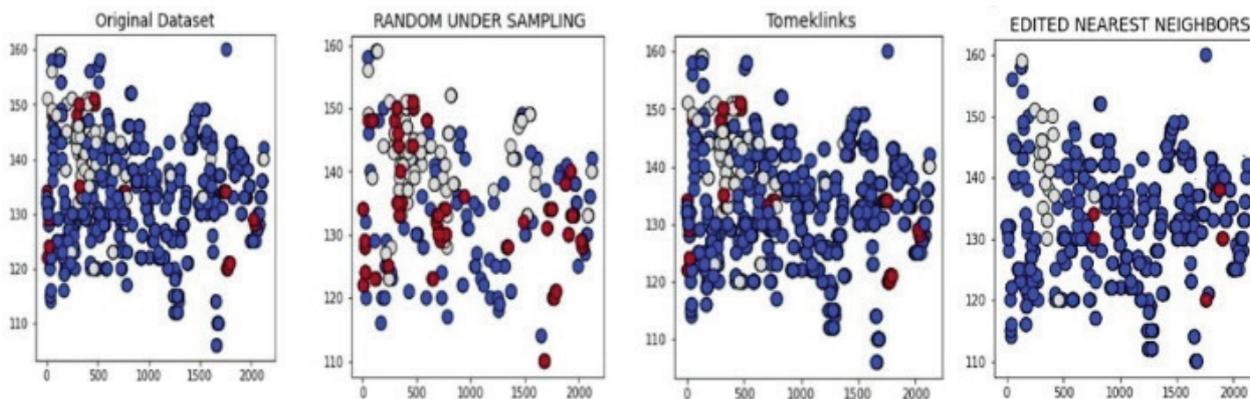


Figure 2. Visualizing under-sampling techniques for dataset representation.

in each class after applying the oversampling and under-sampling techniques.

Proposed Technique

Our proposed approach integrates sampling techniques by combining SMOTE from Oversampling with TOMEK LINKS from Undersampling. This combination is implemented within the Bagging framework using the Random Forest classifier for classification.

Some of the key assumptions include Class Imbalance, Feature Importance, and Model Generalization.

To effectively apply the proposed method for CTG data classification:

1. Applying Smote Tomek sampling for class imbalance.
2. Training a Random Forest ensemble with a bagging classifier, tuning hyperparameters for optimal performance.
3. Evaluating the model using metrics like accuracy and interpreting feature importance to enhance fetal health monitoring outcomes. Fig 3 represents the overall flow-chart of our method.

SMOTETomek for Data Imbalance

From our sampling techniques, we have combined Smote from oversampling and Tomek Links from the undersampling technique into SMOTETomek for our data imbalance [40]. We found that SMOTETomek performs very well than other techniques. We have implemented this technique in the Bagging with Random Forest classifier for classification.

Random Forest Classifier

A classifier containing the multitude of tree structure classifiers used in the process of regression and classification is the Random Forest. The multitude of tree structures is represented as $\{g(x, \phi_t), t=1 \dots\}$ where $\{\phi_t\}$ are unique random vectors that are distributed independently. Each tree votes for the desired class at input x . The selection is done by the forest for the classification based on the increased votes over all the trees in the forest. Every tree grows in the following procedure:

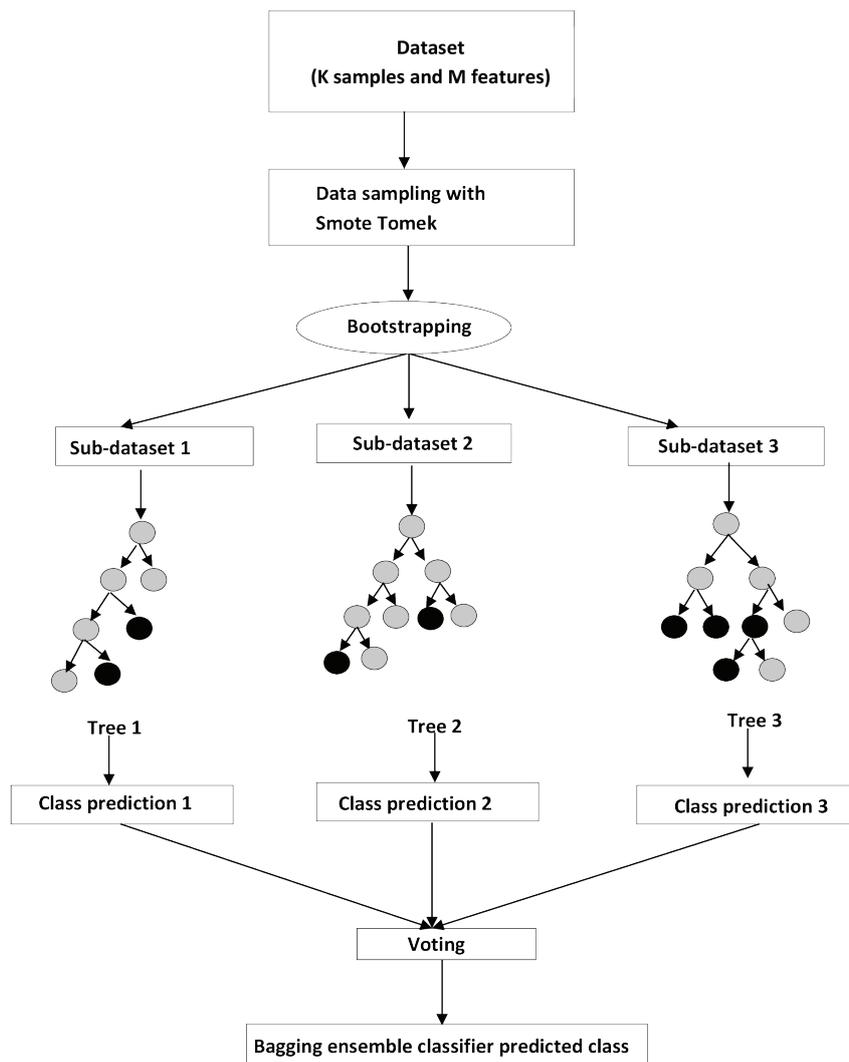


Figure 3. Illustration of our technique

1. If M is the number of cases in the training set and if it samples randomly M cases with replacement from original data then it is the training data for the tree.
2. If we give N input variables, the value of $n \ll N$ will be represented at the individual node, and random selection of m variables out of M is done and the best split helps in splitting the node. The value of m is the constant during the growth.
3. Without any pruning, the tree grows to the maximum extent [41].

RF introduces modifications to the way classification or regression trees are built and incorporates additional randomness through the use of diverse bootstrap samples from the dataset for each tree's construction. In the context of our CTG dataset, which is used for classifying data into 'normal', 'suspect', or 'pathology' categories, the Random Forest method is described as follows:

1. Multiple bootstrap samples are drawn from the actual dataset, creating a collection of n_{tree} bootstrapped datasets.
2. For each of these bootstrap samples, an unpruned tree is developed for classification or regression, with a key difference:
3. In each tree-building node, instead of selecting the best split among all available predictors, a random subset of predictors is sampled, and the split is chosen from this subset. When the number of predictors sampled (m_{try}) equals the total number of predictors (p), it resembles the conventional bagging approach with no randomness.
4. To make predictions on new data, the results from the trees are combined. For classification, majority votes are used, while for regression, the mean is employed [42].

Bagging Ensemble with Random Forest Classifier

In our study, we employed a bagging classifier wherein the random forest serves as the base classifier. This ensemble learning technique combines bagging and the random forest algorithm to create a more resilient and precise model. It generates and consolidates multiple models from bootstrapped samples of the training set, with each model using a Random Forest Classifier as the base estimator. This classifier constructs numerous decision trees using random subsets of the training data and their respective features. The predictions from these trees are then amalgamated to produce the final prediction, enhancing accuracy while mitigating overfitting. The "fit()" function of the bag models is utilized to train the Bagging classifier on the training data. The "max_samples" parameter specifies the number of samples to be drawn from the training set. In a Bagged Random Forest (BRF), a distinct random subset of features is selected for each split within every tree, introducing an additional layer of diversity beyond bootstrap sampling. The overall summary of the proposed method is as follows:

1. Install the necessary libraries.

2. Resample the training data using the SMOTE and TomekLinks sampling strategy to produce two distinct datasets.
3. Construct a pipeline that includes oversampling, under-sampling, and a random forest classifier with 1000 estimators.
4. Create a repeated stratified k-fold cross-validation approach comprising 10 splits and 3 repetitions to evaluate pipeline performance using f1 micro-scoring. Form the pipeline to the training data and compute the average f1 score.
5. Create a bagging classifier with 1000 estimators, a sample size of 0.8, and a True Out-of-Bag score.
6. Apply the bagging classifier to the training data and compute the Out-of-Bag score.

Random Forest, as a standalone algorithm, already combines the predictions of multiple decision trees within the Random Forest model. In contrast, a bagging classifier with Random Forest as the base estimator combines the predictions of multiple base classifiers, with Random Forest. This approach leverages the diversity and strengths of various base classifiers to enhance overall performance and robustness.

RESULTS AND DISCUSSION

Metrics for Assessment

Accuracy (ACC)

Accuracy is the measure of the ratio of precise predictions over the total number of samples evaluated.

$$\text{Acc} = \frac{\text{TN} + \text{TP}}{\text{TP} + \text{FN} + \text{FP} + \text{TN}} \quad (1)$$

Precision (P)

Precision represents the accuracy of positive predictions among all positive samples predicted within a positive class.

$$P = \frac{\text{TP}}{\text{FP} + \text{TP}} \quad (2)$$

Recall (R)

The recall is used to calculate the proportion of correctly classified positive samples.

$$R = \frac{\text{TP}}{\text{TP} + \text{TN}} \quad (3)$$

F1-score

This F1-score is the measure of the harmonic mean between the values of recall and precision

$$\text{F1-score} = \frac{2PR}{P+R} \quad (4)$$

where TP and TN represent the True Positives and Negatives and FP and FN represent the False Positives and Negatives

Performance Evaluation

The fetal CTG dataset is utilized for experimentation, and the method is implemented using Python on a system featuring an 11th Gen Intel(R) Core (TM) i5 processor running a 64-bit operating system, Windows 11. Google Colab with GPU backend is employed for executing the experimentation.

The experimental results for basic Random Forest are given in Table 2. The overall accuracy of the Basic random forest is 93.57%. It is found that good results are produced in the classification of normal images with 95% of precision, 100% of recall, and 97% of F1-score.

The evaluation performance of the under-sampling and over-sampling techniques implemented in the unbalanced data to enhance the classification process is given in Table 3 and Table 4. In the oversampling techniques, it is found that SMOTE with Random Forest produces a greater accuracy of 95% than the Random oversampling and Borderline SMOTE techniques. In the under-sampling technique, the Tomek links produce a very good accuracy of 93% than the Random under-sampling and edited nearest neighbor technique. The Under-sampling techniques other than Tomek Links produce a very low accuracy in comparison with basic RF. It is also seen that the oversampling techniques perform very well compared to the undersampling techniques in this dataset. The method in which we combine the oversampling and under-sampling techniques improves the performance of classification by removing the errors

Table 2. Performance of basic random forest

Technique	Class	Accuracy (%)	Precision	Recall	F1-score
Basic RF	Normal	93.57%	0.95	1.00	0.97
	Suspect		0.88	0.70	0.78
	Pathology		0.91	0.83	0.87

Table 3. Performance of oversampling techniques with RF

Technique	Class	Precision	Recall	F1-score	Accuracy (%)
RF with random oversampling	N	0.97	0.97	0.97	94%
	S	0.80	0.80	0.80	
	P	0.96	0.93	0.95	
RF with SMOTE	N	0.97	0.98	0.98	95%
	S	0.88	0.88	0.84	
	P	0.88	0.88	0.91	
RF with borderline SMOTE	N	0.98	0.97	0.97	93%
	S	0.78	0.86	0.82	
	P	0.94	0.88	0.91	

Table 4. Performance of undersampling techniques with RF

Technique	Class	Precision	Recall	F1-score	Accuracy
RF with Random Undersampling	N	0.99	0.89	0.94	0.89
	S	0.68	0.90	0.77	
	P	0.67	0.90	0.77	
RF with Tomeklinks	N	0.94	1.00	0.97	0.93
	S	0.90	0.68	0.77	
	P	0.92	0.85	0.88	
RF with ENN	N	0.84	1.00	0.91	0.83
	S	0.72	0.38	0.49	
	P	1.00	0.17	0.30	

in both samplings. This method yields better results and improves accuracy greatly. Here, in our results, we found that the random forest with SMOTE performs well in comparison to other techniques, and in under-sampling, it is found that the Random Forest in Tomek Links performs well and achieves the best results.

Hence, we combine these two techniques as SMOTE from oversampling and Tomeklinks from Undersampling as SmoteTomek. Table 5 gives the results of SmoteTomek implemented in the RF classifier and also the results of the combined sampling technique (SmoteTomek) in the bagging ensemble method with the random forest classifier. The proposed technique is the combination of sampling techniques for balancing data and its implementation in the Bagging classifier in which random forest acts as the base classifier. The proposed approach attains a very high accuracy of 98.5%. It also achieves 100% of Precision and Recall in normal and pathology classes.

Table 6 gives the accuracy of our dataset when implemented with other classifiers such as SVM, DT, and K-Nearest Neighbor (KNN). Among the compared deep learning classifiers in Table 7, the proposed model gets the best accuracy (98.5%), suggesting its superior performance. Deep Forest, Deep Neural Network, MKNet, and MKRNN had lesser accuracies ranging from 88.02% to 95.07%, indicating varied degrees of competence in classifying the dataset.

Key Findings

The CTG data classification of FHR signals using the Bagging technique with a Random Forest classifier yielded an impressive accuracy of 96.8%. Further enhancements, including the implementation of over-sampling (SMOTE)

and under-sampling (Tomek links) techniques, resulted in a remarkable 98.5% accuracy in classification. This combined approach effectively addressed class imbalance, demonstrating the technical significance of the method. From a theoretical viewpoint, this method creatively combines resampling techniques and ensemble learning to handle imbalanced CTG datasets. It balances class distribution using Smote Tomek sampling and boosts model performance with a Random Forest ensemble using bagging. From an experimental viewpoint, this method notably enhances classification accuracy, achieving an impressive 98.5% accuracy in CTG data classification. It surpasses traditional classifiers and deep learning models, showcasing its effectiveness in precisely categorizing fetal states based on CTG signals.

Strengths and Limitations

The technical significance of our method lies in its ability to achieve a high level of accuracy and effectively manage class imbalance using the hybrid resampling technique of SMOTE and Tomek links. This approach leverages the Bagging Classifier with a Random Forest base estimator, which not only reduces variance and overfitting but also enhances model generalization, provides valuable insights into feature importance, and benefits from parallel processing capabilities. However, a potential limitation of the Bagging ensemble with Random Forest is that it may increase computational complexity and resource requirements. The limitation of our method is the computational complexity compared to conventional Random Forest. This is because, in a Bagged Random Forest, a distinct random subset of features is chosen for each split within each tree,

Table 5. Performance of combined Techniques with RF

Technique	Class	Precision	Recall	F1-score	Accuracy
SMOTETomek	N	0.98	0.97	0.98	96.8%
With RF	S	0.79	0.89	0.84	
	P	0.94	0.86	0.90	
SMOTETomek in Bagging with RF (Proposed method)	N	1.00	0.98	0.99	98.5%
	S	0.93	0.99	0.96	
	P	0.94	1.00	0.97	

Table 6. Performance of evaluation with competent classifiers

Classifier	Accuracy (%)
SVM	92%
Decision tree	93%
KNN	81%

Table 7. Comparison with other deep learning classifier

Deep learning classifiers	Accuracy (%)
Proposed model	98.5%
Deep Forest algorithm [43]	95.07 %
Deep Neural network architecture [44]	88.02%
MKNet [45]	94.70%
MKRNN [45]	90.30%

adding an extra level of variation alongside the bootstrap sampling process.

Comparison with Similar Approaches

Our proposed method was compared to other classifiers, and it outperformed them, achieving greater accuracy. A bar chart in Figure 6 provides a visual representation of the comparison of accuracies among different classifiers. Moreover, the performance details from relevant tables, including Table 2, which demonstrates that the basic Random Forest (RF) did not meet the expected accuracy for the data, prompted us to implement over-sampling and under-sampling techniques. The results in Table 3 and Table 4 indicated that SMOTE for oversampling and Tomek links for undersampling outperformed other techniques. Consequently, we decided to oversample our data with SMOTE and undersample with Tomek links, using a basic RF classifier resulting in the enhanced accuracy of 96.8% as demonstrated in Table 5. Tables 6 and 7 give the comparative performance with other ML and DL classifiers. The confusion matrix for our proposed Rf is illustrated in Fig:5.

CONCLUSION

The assessment of fetal state using the CTG is essential due to its critical importance, as it has the potential to result in fatal outcomes if not managed appropriately. The classification of fetal state from CTG using Machine Learning algorithms is presented. Though there are various machine learning classifiers implemented in this classification of CTG data it is found that there is an obvious need for the upgrade of accuracy. In this, we use the combined sampling technique in a Random Forest classifier with Bagging. We first implemented the oversampling techniques as smote, Random oversampling, and borderline smote. Then we implemented the under-sampling techniques such as Random under-sampling, Tomek Links, and Edited nearest neighbors. Based on the results, we combined the Smote from oversampling and Tomek Links from under-sampling as they performed well among the other techniques. Then we used the ensemble method of bagging with random forest as the base classifier. Our method achieves astonishing results of 98.5 % accuracy in the classification of this CTG data. Future works can be done by applying class weight after bootstrapping and improvement in feature selection can also be done.

AUTHORSHIP CONTRIBUTIONS

Authors equally contributed to this work.

DATA AVAILABILITY STATEMENT

The authors confirm that the data that supports the findings of this study are available within the article. Raw data that support the finding of this study are available from the corresponding author, upon reasonable request.

CONFLICT OF INTEREST

The author declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

ETHICS

There are no ethical issues with the publication of this manuscript.

REFERENCES

- [1] Jenkins HM. Thirty years of electronic intrapartum fetal heart rate monitoring: a discussion paper. *J R Soc Med* 1989;82:210–214. [\[CrossRef\]](#)
- [2] Parer JT, King T. Fetal heart rate monitoring: is it salvageable? *Am J Obstet Gynecol* 2000;182:982–987. [\[CrossRef\]](#)
- [3] Grivell RM, Alfirevic Z, Gyte GM, Devane D. Antenatal cardiotocography for fetal assessment. *Cochrane Database Syst Rev* 2015;2015:CD007863. [\[CrossRef\]](#)
- [4] Von Steinburg SP, Boulesteix AL, Lederer C, Grunow S, Schiermeier S, Hatzmann W, et al. What is the “normal” fetal heart rate? *PeerJ* 2013;1:e82. [\[CrossRef\]](#)
- [5] Alfirevic Z, Gyte GM, Cuthbert A, Devane D. Continuous cardiotocography (CTG) as a form of electronic fetal monitoring (EFM) for fetal assessment during labour. *Cochrane Database Syst Rev* 2013:CD006066. [\[CrossRef\]](#)
- [6] Kelso IM, Parsons RJ, Lawrence GF, Arora SS, Edmonds DK, Cooke ID. An assessment of continuous fetal heart rate monitoring in labor: a randomized trial. *Am J Obstet Gynecol* 1978;131:526–532. [\[CrossRef\]](#)
- [7] Zhang Y, Zhao Z. Fetal state assessment based on cardiotocography parameters using PCA and AdaBoost. In: 2017 10th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI). *IEEE*; 2017. p. 1–6. [\[CrossRef\]](#)
- [8] Subha V, Murugan D, Rani J, Rajalakshmi K, Tirunelveli T. Comparative analysis of classification techniques using Cardiotocography dataset. *Int J Res Inf Technol* 2013;1:274–280.
- [9] Potharaju SP, Sreedevi M, Ande VK, Tirandasu RK. Data mining approach for accelerating the classification accuracy of cardiotocography. *Clin Epidemiol Glob Health* 2019;7:160–164. [\[CrossRef\]](#)
- [10] Yılmaz E, Kılıkçier Ç. Determination of fetal state from cardiotocogram using LS-SVM with particle swarm optimization and binary decision tree. *Comput Math Methods Med* 2013;2013:1–9. [\[CrossRef\]](#)
- [11] Ocağ H. A medical decision support system based on support vector machines and the genetic algorithm for the evaluation of fetal well-being. *J Med Syst* 2013;37:1–9. [\[CrossRef\]](#)

- [12] Shah SA, Aziz W, Arif M, Nadeem MS. Decision trees-based classification of cardiocograms using bagging approach. In: 2015 13th International Conference on Frontiers of Information Technology (FIT). IEEE; 2015. p. 12–17. [\[CrossRef\]](#)
- [13] Cömert Z, Kocamaz A. A study of artificial neural network training algorithms for classification of cardiocography signals. Bitlis Eren Univ J Sci Technol 2017;7:93–103. [\[CrossRef\]](#)
- [14] Iraj MS. Prediction of fetal state from the cardiocogram recordings using neural network models. Artif Intell Med 2019;96:33–44. [\[CrossRef\]](#)
- [15] Amin B, Gamal M, Salama AA, El-Henawy IM, Mahfouz K. Classifying cardiocography data based on rough neural network. Int J Adv Comput Sci Appl 2019;10:352–356. [\[CrossRef\]](#)
- [16] Kadhim NJ, Abed JK. Enhancing the prediction accuracy for cardiocography (CTG) using fire-fly algorithm and naive Bayesian classifier. In: IOP Conference Series: Materials Science and Engineering. Vol. 745, No. 1. IOP Publishing; 2020. p. 012101. [\[CrossRef\]](#)
- [17] Fasihi M, Nadimi-Shahraki MH, Jannesari A. A shallow 1-D convolution neural network for fetal state assessment based on cardiocogram. SN Comput Sci 2021;2:287. [\[CrossRef\]](#)
- [18] Ajirak M, Heiselman C, Quirk JG, Djurić PM. Boost ensemble learning for classification of CTG signals. In: ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE; 2022. p. 1316–1320. [\[CrossRef\]](#)
- [19] Manikandan M, Vijayakumar P. Improving the performance of classifiers by ensemble techniques for the premature finding of unusual birth outcomes from cardiocography. IETE J Res 2023;69:1734–1744. [\[CrossRef\]](#)
- [20] Hephzibah R, Christinal AH, Jayanthi R, Jebasingh S, Chandy DA, Bajaj C. A classification approach to fetal cardiocography dataset using R. CEUR Workshop Proceedings. Workshop on Computer Vision and Machine Learning for Healthcare, April 22-24, 2022, Chennai, India.
- [21] Hephzibah R, Christinal AH, Jayanthi R, Chandy DA, Bajaj C. An ensemble technique of SmoteENN and random forest for classification of the cardiocography data. International Conference on Computer Vision and Internet of Things 2023 (ICCVIoT'23). [\[CrossRef\]](#)
- [22] Hephzibah R, Christinal AH, Jayanthi R, Chandy DA, Bajaj C. A novel ensemble classifier framework for accurate fetal heart rate classification. In: 2023 4th International Conference on Signal Processing and Communication (ICSPC). IEEE; 2023. p. 321–324. [\[CrossRef\]](#)
- [23] Junsomboon N, Phienthrakul T. Combining over-sampling and under-sampling techniques for imbalance dataset. In: Proceedings of the 9th International Conference on Machine Learning and Computing. 2017. p. 243–247. [\[CrossRef\]](#)
- [24] Belarouci S, Chikh MA. Medical imbalanced data classification. Adv Sci Technol Eng Syst 2017;2:116–124. [\[CrossRef\]](#)
- [25] Sun Y, Wong AK, Kamel MS. Classification of imbalanced data: A review. Int J Pattern Recognit Artif Intell 2009;23:687–719. [\[CrossRef\]](#)
- [26] Ramyachitra D, Manikandan P. Imbalanced dataset classification and solutions: a review. Int J Comput Bus Res 2014;5:1–29.
- [27] Dutta P, Paul S, Majumder M. Intelligent SMOTE based machine learning classification for fetal state on cardiocography dataset. Research Square 2021. Preprint. doi: Doi: 10.21203/rs.3.rs-1040799/v1. [\[CrossRef\]](#)
- [28] Kovács G. An empirical comparison and evaluation of minority oversampling techniques on a large number of imbalanced datasets. Appl Soft Comput 2019;83:105662. [\[CrossRef\]](#)
- [29] Chawla NV, Japkowicz N, Kotcz A. Special issue on learning from imbalanced data sets. ACM SIGKDD Explor Newsl 2004;6:1–6. [\[CrossRef\]](#)
- [30] Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: synthetic minority over-sampling technique. J Artif Intell Res 2002;16:321–357. [\[CrossRef\]](#)
- [31] Wang L, Han M, Li X, Zhang N, Cheng H. Review of classification methods on unbalanced data sets. IEEE Access 2021;9:64606–64628. [\[CrossRef\]](#)
- [32] Han H, Wang WY, Mao BH. Borderline-SMOTE: a new over-sampling method in imbalanced data sets learning. In: International Conference on Intelligent Computing. Springer Berlin Heidelberg; 2005. p. 878–887. [\[CrossRef\]](#)
- [33] Yen SJ, Lee YS. Under-sampling approaches for improving prediction of the minority class in an imbalanced dataset. In: Intelligent Control and Automation: International Conference on Intelligent Computing, ICIC 2006 Kunming, China, August 16-19, 2006. Springer Berlin Heidelberg; 2006. p. 731–740. [\[CrossRef\]](#)
- [34] Chawla NV. Data mining for imbalanced datasets: an overview. The Data Mining and Knowledge Discovery Handbook. Berlin, Heidelberg, Dordrecht, and New York City: Springer Nature; 2010. p. 875–886. [\[CrossRef\]](#)
- [35] Wongvorachan T, He S, Bulut O. A comparison of undersampling, oversampling, and SMOTE methods for dealing with imbalanced classification in educational data mining. Information 2023;14:54. [\[CrossRef\]](#)
- [36] Elhassan T, Aljurf M. Classification of imbalance data using Tomek link (T-link) combined with random under-sampling (RUS) as a data reduction method. Glob J Technol Optim S 2016;1. [\[CrossRef\]](#)

- [37] Ai-jun L, Peng Z. Research on unbalanced data processing algorithm base Tomeklings-Smote. In: Proceedings of the 2020 3rd International Conference on Artificial Intelligence and Pattern Recognition. 2020. p. 13–17. [\[CrossRef\]](#)
- [38] Tomek I. An experiment with the edited nearest-neighbor rule. *IEEE Trans Syst Man Cybern* 1976;448–452. [\[CrossRef\]](#)
- [39] Kumar V, Lalotra GS, Sasikala P, Rajput DS, Kaluri R, Lakshmana K, et al. Addressing binary classification over class imbalanced clinical datasets using computationally intelligent techniques. *Healthcare* 2022;10:1293. [\[CrossRef\]](#)
- [40] Alex DM, Chandy DA, Christinal AH, Singh A, Pushkaran M. A hybrid random forest classifier for chronic kidney disease prediction from 2D ultrasound kidney images. *Int J Pattern Recognit Artif Intell* 2022;2256010. [\[CrossRef\]](#)
- [41] Kulkarni VY, Sinha PK. Random forest classifiers: a survey and future research directions. *Int J Adv Comput* 2013;36:1144–1153.
- [42] Liaw A, Wiener M. Classification and regression by randomForest. *R News* 2002;2:18–22.
- [43] Chen Y, Guo A, Chen Q, Quan B, Liu G, Li L, et al. Intelligent classification of antepartum cardiotocography model based on deep forest. *Biomed Signal Process Control* 2021;67:102555. [\[CrossRef\]](#)
- [44] Miao JH, Miao KH. Cardiotocographic diagnosis of fetal health based on multiclass morphologic pattern predictions using deep learning classification. *Int J Adv Comput Sci Appl* 2018;9. [\[CrossRef\]](#)
- [45] Tang H, Wang T, Li M, Yang X. The design and implementation of cardiotocography signals classification algorithm based on neural network. *Comput Math Methods Med* 2018;2018. [\[CrossRef\]](#)