[1,2] A. Venkata Sriram

P. Rajesh Kumar[3]

L. Alekhya[4]

Cardiotocogram Data classification using Machine Learning Algorithms along with Swarm-Based Metaheuristic optimizations for Autonomous Fetal distress detection.

Journal of Electrical Systems

Abstract: - Fetal distress is a critical condition that requires timely and accurate detection to warrant the safety of birth mother and fetus. Uterine contractions of birth mother and foetus heart rate are generally monitored by a non-invasive technique called Cardiotography. However, the manual interpretation of CTG data is often subjective and prone to errors. This study presents the autonomous detection of fetal distress using various metahuriestic optimization techniques for machine learning algorithms. The dataset of CTG recordings from UCI repository is utilized and pre-processed to apply for various machine learning algorithms including Naïve Bayes (NB), k-Nearest Neighbour (k-NN), Support Vector Machines (SVM) and Random Forest (RF) that are employed to classify the CTG data into various categories such as normal, suspect, and pathological. To further improve the classification accuracy, we integrate nature-inspired metaheuristic optimization methods such as Firefly, Grasshopper and Greywolf algorithms to fine tune the hyperparameters select appropriate features. Greywolf optimized model with Random Forest classifier algorithm outperformed other models with overall accuracy of around 93.65%, Weighted F1 score of 94.12% with Mean MCC and Mean Kappa score of about 82.86% and 81.70% respectively and an average ROC-AUC of 92.78% suggests that Greywolf optimized model with Random Forest algorithm is a reliable tool for the early detection of fetal distress. The results indicate that the integration of machine learning with metaheuristic optimization not only enhances the predictive capabilities but also provides a robust framework for autonomous fetal monitoring systems.

Keywords: Cardiotocography, Support Vector Machine, Random Forest, k-Nearest Neighbour, Firefly, Grasshopper, GreyWolf.

I Introduction

Fetal distress, characterized by signs that indicate the fetus is not well during pregnancy or labor, possess significant risks to both the mother and the unborn child. Cardiotocography (CTG) is most commonly used in clinical practice for fetal monitoring with the help of fetal heart rate (FHR) and uterine contractions. Despite its prevalence, the interpretation of CTG data remains challenging due to the variability in fetal heart rate patterns and the subjective nature of clinical assessments.

Traditional methods of observing and analysing the CTG data involve manual inspection by healthcare professionals, which can be defective and prone to human error. Henceforth there is an eventual need for automated systems that can provide reliable and objective assessments of fetal health. Advancements in Artificial intelligence with machine learning algorithms offers a promising solution to this problem by increasing the computational power to identify patterns and make accurate predictions from complex datasets.

In recent years, machine learning algorithms have shown considerable potential in medical diagnostics, including the classification of CTG data. Techniques such as Naïve Bayes, Support Vector Machines (SVM), k-Nearest Neighbour, Random Forests and Neural Networks have been employed to differentiate between normal, suspect and pathological states. However, the performance of these algorithms can be significantly affected by the choice of hyperparameters and feature selection, which are crucial for achieving high classification accuracy.

To address these challenges, this study explores the integration of swarm-based metaheuristic optimization techniques with machine learning algorithms for the autonomous detection of fetal distress. Swarm-Based Metaheuristic algorithms like Firefly Algorithm (FA), Grasshopper Optimisation Algorithm (GOA) and Greywolf

[1] Lecturer, Electronics and Communication Department, Government Polytechnic, Parvathi Puram, Andhra Pradesh, India

Electronics and Communication Department, Andhra University College of Engineering, Visakhapatnam, India[2]

Professor, Electronics and Communication Department, Andhra University College of Engineering, Visakhapatnam, India[3]

Assistant Professor, Electronics and Communication Department, Lendi Institute of Engineering and Technology, Vizianagaram, Andhra Pradesh, India[4]

Algorithm (GW) are employed to optimize the hyper-parameters and enhance the feature selection process which thereby improve the predictive performance of the classifiers.

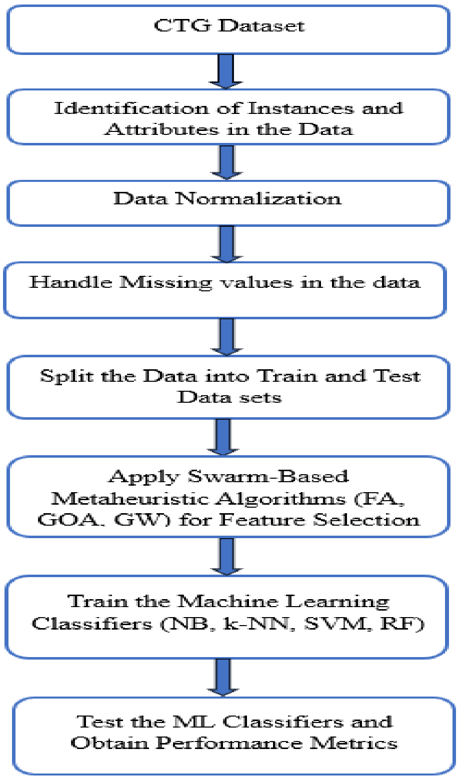


Figure 1. Block Diagram of Proposed Methodology

The primary aim of this paper is to develop a precise and resilient artificial intelligence system capable of identifying and assessing fetal distress during labor, thereby categorizing the health status as normal, suspect, or pathological using Machine learning algorithms with swarm-based optimization techniques. Figure 1 illustrates the block diagram depicting the proposed methodology. The objectives of this research are given below:

1. Implement and assess the machine learning models for the classification of CTG data in UCI repository into categories such as normal, suspect, and pathological.

2. Develop swarm-based metaheuristic optimization methods to fine-tune model parameters and select the most relevant features.

3. Evaluate and compare the efficacy of the optimized models in accurately detecting fetal distress with the obtained performance metrics like Overall Accuracy, Weighted F_1 score, Mean MCC, Mean Kappa score and ROC-AUC.

By combining machine learning algorithms with swarm-based metaheuristic optimization, this study aims to create a robust and reliable system for autonomous fetal distress detection. The proposed approach not only enhances the accuracy and consistency of CTG data interpretation but also paves the way for improved prenatal care and outcomes.

In the following sections, we discuss the methodology, including data preprocessing, feature extraction, and the implementation of machine learning algorithms and swarm-based metaheuristic optimization techniques. We then present the results and analyze the performance of the optimized models, followed by a discussion on the implications of our findings and potential directions for future research.

II Literature Work

The classification of Cardiotocogram (CTG) data for fetal distress detection has been an area of active research for several decades. This review provides an overview of the key studies and methodologies that have contributed to advancements in this field, particularly focusing on machine learning algorithms and the use of swarm-based metaheuristic optimization techniques.

Traditional analysis of CTG data has largely been manual, relying on clinical expertise to interpret the fetal heart rate patterns and uterine contractions. These methods are often subjective and vary significantly between clinicians, leading to inconsistent diagnoses and potential misinterpretation of fetal distress signals. Early automated systems attempted to standardize interpretation using rule-based approaches, but these systems lacked flexibility and adaptability to individual variations in CTG recordings.

Machine learning (ML) has emerged as a powerful tool for the analysis of medical data, including CTG recordings. Several studies have demonstrated the potential of various ML algorithms to classify CTG data with higher accuracy and reliability compared to traditional methods. Georgoulas et al. [1] employed SVMs to classify CTG recordings, achieving promising results in differentiating between normal and pathological cases. Their approach highlighted the capability of SVMs to handle the high-dimensional and non-linear nature of CTG data. Jezewski et al. [2] explored the use of neural networks for CTG classification, leveraging their ability to learn complex patterns in data. They reported significant improvements in classification accuracy, particularly when deep learning architectures were used. A study by Ocak [3] applied Random Forests to classify CTG data, demonstrating the algorithm's robustness and interpretability. The ensemble nature of Random Forests helped in managing the variability in CTG patterns.

Improta et al. [4] utilized a custom-developed CTG automatic analysis software to extract feature data from CTG signals and predict childbirth using various algorithms, including J48, AdaBoost, random forests, and gradient boosting trees. The random forests classifier achieved the highest performance, with an accuracy of 87.6% and an AUC-ROC of 93.0Ricciardi et al. [5] extracted 17 features from existing CTG signals using custom software and applied machine learning algorithms, including J48, random forests (RF), and decision tree AdaBoost (AdA-B), for classification. Among these, RF and AdA-B produced superior results, with AUC-ROC values exceeding 94.9%. In a separate study, Chudáček et al. [6] developed an algorithm to classify CTG signals as normal, suspicious, or pathological based on FHR and UC patterns, achieving a classification accuracy of 87.3%, highlighting its clinical potential. In another study by Chudáček et al. [7], an algorithm was developed to classify CTG signals based on FHR variability and acceleration, resulting in a classification accuracy of 88.8%, underscoring the potential for more targeted CTG signal analysis to enhance classification accuracy.

Feature selection and engineering play a critical role in enhancing the performance of ML algorithms for CTG classification. Several studies have focused on identifying relevant features from CTG recordings that are indicative of fetal distress. Diab et al. [8] utilized wavelet transform techniques to extract time-frequency features from CTG data. This approach provided a more comprehensive representation of the signals, improving the classification accuracy of ML models. Chudáček et al. [9] applied PCA to reduce the dimensionality of CTG data while preserving the most informative features. This method enhanced the efficiency and performance of subsequent classification algorithms

Swarm-Based Metaheuristic optimization techniques have been increasingly integrated with ML algorithms to optimize hyperparameters and improve feature selection, leading to better classification performance. Recent research has focused on combining multiple ML algorithms and optimization techniques to develop hybrid models for CTG classification. For instance, Silva et al. [10] proposed a hybrid model combining SVM, Random Forest, and GA for feature selection and hyperparameter optimization. Their hybrid approach achieved state-of-the-art performance in CTG classification. S. Öztürk et al. [11] performed Empirical mode Decomposition for spectral entropy feature extraction ReliefF algorithm for feature selection and support vector machine for classification of Normal and Abnormal classes and achieved an average accuracy of around 90.0%. The integration of deep learning with metaheuristic optimization is also gaining traction. Studies are exploring the use of deep neural networks (DNN) [12] and convolutional neural networks (CNN) combined with optimization techniques like GA and PSO [13] to handle the complexity and variability of CTG data more effectively. Also for the Imbalanced data classification model, combining the concepts of Generative AI into the data generation further increase the performance of the model. Future research should focus on real-time implementation and validation in clinical settings to fully realize the potential of these technologies in improving the prenatal care.

III Research Methodology

In this study, the analysis focuses on the CTG dataset sourced from the UCI Machine Learning Repository, originally obtained from the SisPorto 2.0 software adhering to FIGO guidelines. The objective is to evaluate the effectiveness of machine learning models. The dataset undergoes preprocessing, encompassing the handling of missing values and data normalization. Subsequently, the data is split into training and testing datasets, as depicted in Figure 1. The training data is utilized to train four distinct machine learning algorithms. Various performance metrics, including Accuracy, F_1 score, MCC and kappa score along with ROC-AUC curves obtained through Confusion Matrix analysis, are then examined to address the multiclass classification problem of fetal distress detection using CT dataset.

A. Preprocessing

The CTG data set considered from UCI machine learning repository contains 2126 instances with 21 related features. The features values of this feature set are in different ranges and some values may be missed or may contain NaN values. The steps involved in the data preprocessing stage is shown in the Figure 2. To apply these values to next stage i.e., classification stage the missing values must be handled by either removing them or replacing them with mean of the feature column where that value is present.

Once the missing values are handled, the feature scaling is applied using z-score normalisation to improve convergence speed of gradient descent algorithms and enhances the potential of the machine learning models. After proper preprocessing, the data set can now be applied to classification stage.

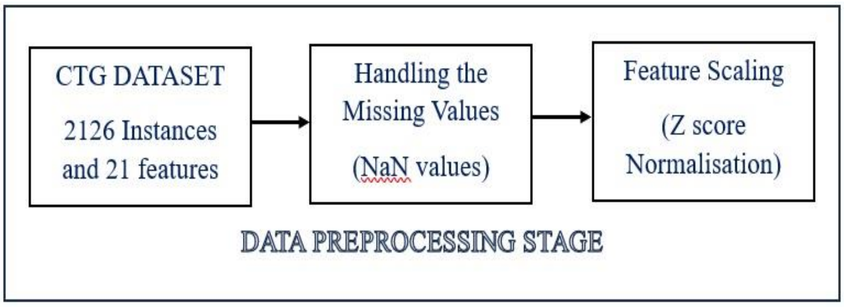


Figure 2. Steps involved in Data Preprocessing Stage

B. Dataset Description

Cardiotocography serves as a method for monitoring both fetal heart activity and uterine contractions. The dataset utilized in this study is sourced from the UCI Machine Learning Repository, recorded via the Sisporto version 2.0 program [14], which can be installed on any personal computer for signal acquisition. This automated program adheres closely to FIGO guidelines for the analysis of fetal distress detection. The dataset comprises 2126 instances and 23 attributes [15] of which 11 are discrete 8 are continuous features and 2 are 3-class and 10-class labels, primarily derived from fetal heart rate (FHR) baseline, uterine contractions per second (UC), and fetal movements per second (FM). Additionally, several other attributes contribute to the recognition of fetal health status. Among these attributes, four are deemed fundamental and critical in CTG data analysis: Fetal Heart Rate Baseline (BL), Accelerations (ACC), Decelerations (DCL), and Variability. If we observe the dataset instances there are 1655 Normal, 295 Suspect and 176 Pathologic 3-class samples which is an imbalanced dataset. For analysing the performance of the model with such data Mean MCC and Kappa score metrics play a crucial role.

C. Machine Learning Algorithms

1. Naive Bayes classifier (NB)

The Naive Bayes classifier is the probability-based classifier which uses the Bayes Theorem that is simple and most efficient. It is relevant to Bayesian network in which all attributes are independent of given class variable. This conditional independency of the attributes of Bayes theorem can be called as Naive bayes [16].

2. k-Nearest Neighbour classifier (k-NN)

The k-Nearest Neighbor algorithm is a straightforward supervised machine learning technique that falls under the category of example-based learning. It classifies data by assessing the similarity of each data point to others [17] by computing the k neighbors using the Euclidean distance.

3. Support Vector Machine (SVM)

SVMs are widely recognized for its effectiveness in solving classification problems, particularly for datasets with high dimensions (i.e., numerous features) [18]. The fundamental principle of SVM revolves around identifying the optimal maximum margin hyperplane (MMH). In constructing this hyperplane, SVM selects the most extreme points or vectors, termed support vectors, hence lending its name to the algorithm. Error-Correcting Output Codes (ECOC) is a technique employed to address multi-class classification problems by transforming them into multiple binary classification tasks [19]. Therefore, SVM can be adapted to handle multiclass classification problems through the utilization of the ECOC method.

4. Random Forest (RF)

Random Forest is a classifier comprising multiple decision trees. A training forest which is composed of multiple classification trees is used to create a random forest. It is more accurate and stable prediction. It trains the model with the bagging method. The classification result of the test data is obtained by the score formed by the classification tree voting [20].

D. Swarm-Based Metaheuristic Optimization

1. Firefly Algorithm

The Firefly Optimization Algorithm (FA) is a nature-inspired metaheuristic optimisation algorithm [21]. It mimics the behavior of fireflies, particularly their flashing patterns used for attracting mates or prey. The algorithm's core idea is to use the intensity of the flashes as an analogy for the objective function to be optimized. Brighter fireflies attract others, guiding the swarm towards optimal solutions. Figure 3 given below depicts the flow chart of Firefly optimisation algorithm. FA has been applied successfully across various domains due to its simplicity and effectiveness in solving complex optimization problems. FA can majorly be used as feature selection or reduction technique to optimize the number attributes which further reduces the redundancy in the data and properly trains the classification models and increases the performance.

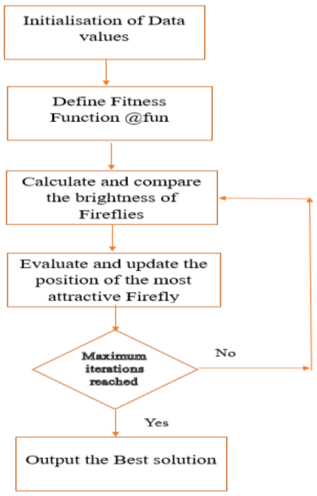


Figure 3. Flow chart of Firefly Optimization Algorithm

2. Grasshopper Algorithm

The Grasshopper Optimization Algorithm (GOA) [22] is a population-based optimization technique is stimulated by the joint attitude of grasshoppers. This algorithm has gained attention in the field of feature selection due to its ability to effectively navigate through complex solution spaces. GOA can be hired to identify the optimal subset of features that contribute most significantly to predictive modeling, effectively enhancing the performance of classifiers applied to CTG data. Figure 4 depicts the flow chart of Grasshopper optimisation algorithm in the selection of features of the dataset.

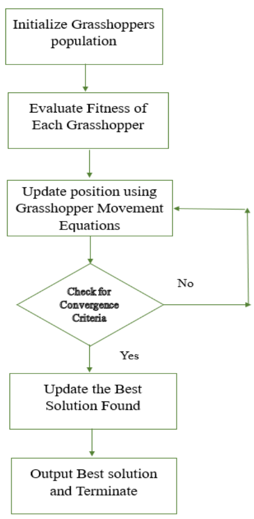


Figure 4. Flow chart of Grasshopper Optimization Algorithm

3. Greywolf Algorithm

The Grey Wolf Optimization (GWO) [23] algorithm is encouraged by the social hierarchy and chasing process of grey wolves. It has emerged as a powerful swarm-based metaheuristic algorithm for solving optimization problems, including feature selection tasks. GWO effectively models the dynamics of wolf packs, which allows it to explore and exploit the solution space efficiently, making it suitable for high-dimensional data like CTG. Figure 5 given below depicts the flow chart of grey wolf optimisation algorithm in feature selection process of CTG dataset.

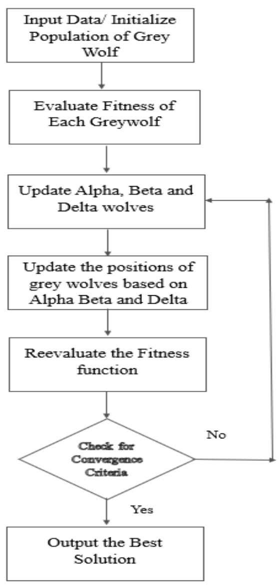


Figure 5. Flow chart of Grey wolf optimization

E. Multiclass classification Metrics

Multiclass classification is such scenario where the number of output units/classes are greater than two. In this classification problem we need evaluate the parameter based on One vs Rest algorithm. One vs Rest is based on considering one class and evaluate the parameters of that one class with respect to the rest of all class parameters from the confusion matrix only. In assessing the performance of machine learning models, the Confusion Matrix serves as a pivotal tool, comprising an N x N matrix where N represents the number of target classes. In this study, N equals 3. Unlike binary class classification, parameters are evaluated individually for each class in a multiclass scenario. Consequently, their measurement varies and is contingent upon the specific class. These parameters are calculated separately for each class, and subsequently, the relevant performance metrics are determined using the formulas outlined in Table 1 below.

Table 1. Performance Metrics Formulae for the Evaluation of the ML models

Performance Metrics	Formulae
Accuracy of the Class	$\frac{(TP_c + TN_c)}{(TP_c + TN_c + FP_c + FN_c)}$
Sensitivity	$\frac{TP_c}{(TP_c + FN_c)}$
Specificity	$\frac{TN_c}{(TN_c + FP_c)}$
Precision	$\frac{TP_c}{(TP_c + FP_c)}$
F_1 Score	$\frac{2 * TP_c}{(2 * TP_c + FP_c + FN_c)}$
Mathews Correlation Coefficient	$\frac{(TP_c * TN_c - FP_c * FN_c)}{\sqrt{(TP_c + FP_c)(TP_c + FN_c)(TN_c + FP_c)(TN_c + FN_c)}}$
Cohen Kappa Score	$\frac{2 * (TP_c * TN_c - FP_c * FN_c)}{(TP_c + FP_c) * (TN_c + FP_c) + \ (TP_c + FN_c) * (TN_c + FN_c)}$
Classification Error	1-Accuracy

Since the study focuses on multiclass classification problem with three classes, the value of c is 3 in the above metrics and are calculated for each class individually. The overall accuracy of the model is obtained from Equation (1) given below.

$$\text{Overall Accuracy of the model} = \frac{Correctly\ predicted\ samples}{Total\ samples} \quad (1)$$

Another essential metric frequently utilized in analysing false positive rates is the Receiver Operating Characteristic (ROC) curve, which plots the false positive rate (FPR) against the true positive rate (TPR), where TPR corresponds to sensitivity and FPR is given as 1-specificity. This curve aids in selecting an optimal cut-off value for determining the classes utilized in the study. Additionally, the Area Under the Curve (AUC) is calculated from the ROC curve, with its value ranging between 0 and 1. A higher AUC value, closer to 1, signifies superior performance of the classifier model.

In addition to overall accuracy, mean Matthews Correlation Coefficient (MCC), weighted F1 score, mean Cohen Kappa score, and misclassification error were computed for analysing the machine learning models. For unbalanced data, Mathew's correlation coefficient (MCC) and Cohen Kappa score are preferred over accuracy and F1 score, as they provide more appropriate measurements of the model's ability to classify the problem. These metrics are calculated and tabulated in the Results and Discussions section.

IV Experimental Setup

Experiments were conducted on a computer having an Intel i5 microprocessor and 16 GB Ram and Nvidia GPU of 2 GB, while coding were carried out on MATLAB 2023b version software. The study was conducted on the CTG dataset of size 2126 x 23 which include 21 attributes of which 11 are discrete and 8 are continuous, 1 for 3 class and the other 1 for 10 class classification labels together constitute 23 columns in the dataset. Out of 2126 instances 1655 belongs to Normal, 295 to Suspect and 176 to pathologic classes respectively. Hence the data size becomes 2126 x 21 with 3 class classification columns of the data taken as labels in this study. This data along with labels are divided into train and test dataset in the ratio of 80 and 20 percent respectively. Meanwhile Swarm based Metaheuristic optimization techniques namely Firefly, Grasshopper and Greywolf techniques are used to select the optimized features from the available 21 feature set.

By applying Firefly optimization technique on the CTG data 15 optimized feature indexes were selected and the elapsed time taken for the algorithms is 20.15 seconds. With Grasshopper optimization technique 7 optimized feature indexes were selected and around 8.68 seconds were taken to execute this technique. On the other hand, with Grey wolf optimization technique 15 best optimised feature indexes were selected with elapsed time of 22.84 seconds. The graph in the Figure 6 shows the comparison of elapsed time and number of features obtained from each Swarm based Metaheuristic feature optimization methods.

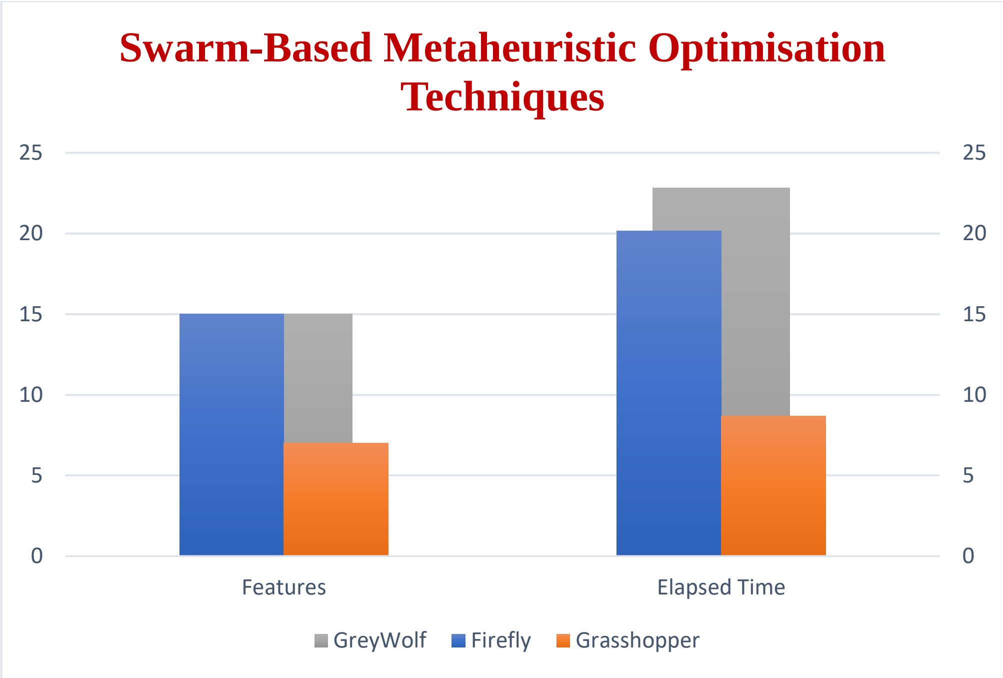


Figure 6. Comparison of three Swarm based Metaheuristic Optimization Techniques

Both the train and test dataset with labels are reformed with respect to the obtained feature indexes from the swarm-based metaheuristic optimization techniques. The train dataset is now applied as input to four Machine learning algorithms namely Naïve bayes, SVM, k-NN and Random Forest to train the models and the test dataset is used evaluate these models and calculate various performance metrics using Confusion plot as discussed in Section III. The experimental analysis with appropriate Confusion plots and ROC-AUC curves for four Machine learning models with three Swarm based Metaheuristic optimization techniques are given below. Figure 7 shows the confusion plots of Firefly optimization techniques with four machine learning algorithms where as Figure 8 shows the confusion plots of Grasshopper optimization technique with four machine learning algorithms used to classify the three classes of CTG dataset namely Normal, Pathologic and Suspect.

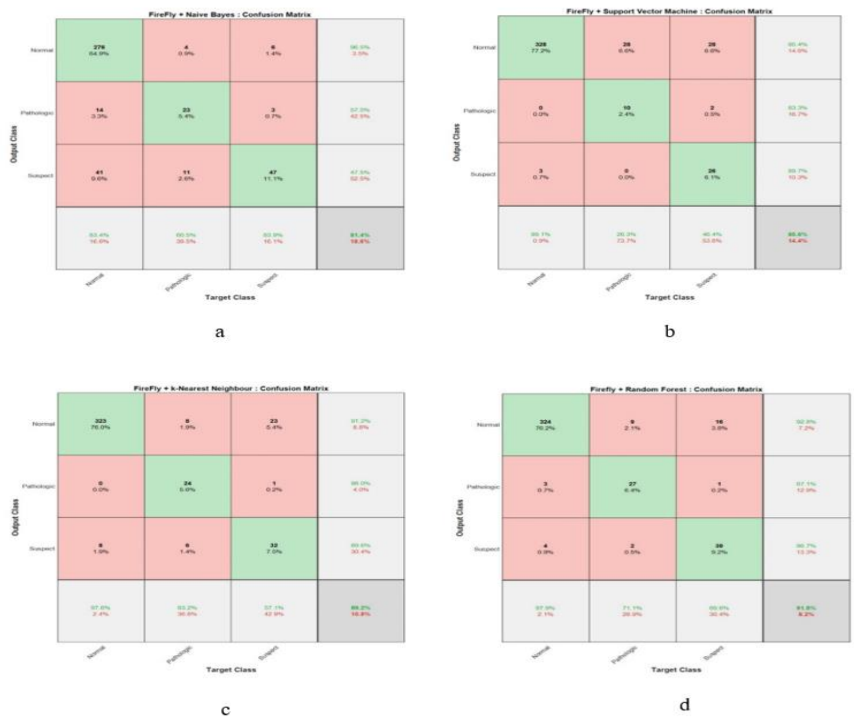


Figure 7. Confusion plot of Firefly Optimization with a) Naive Bayes b) Support Vector Machine c) k-Nearest Neighbor d) Random Forest.

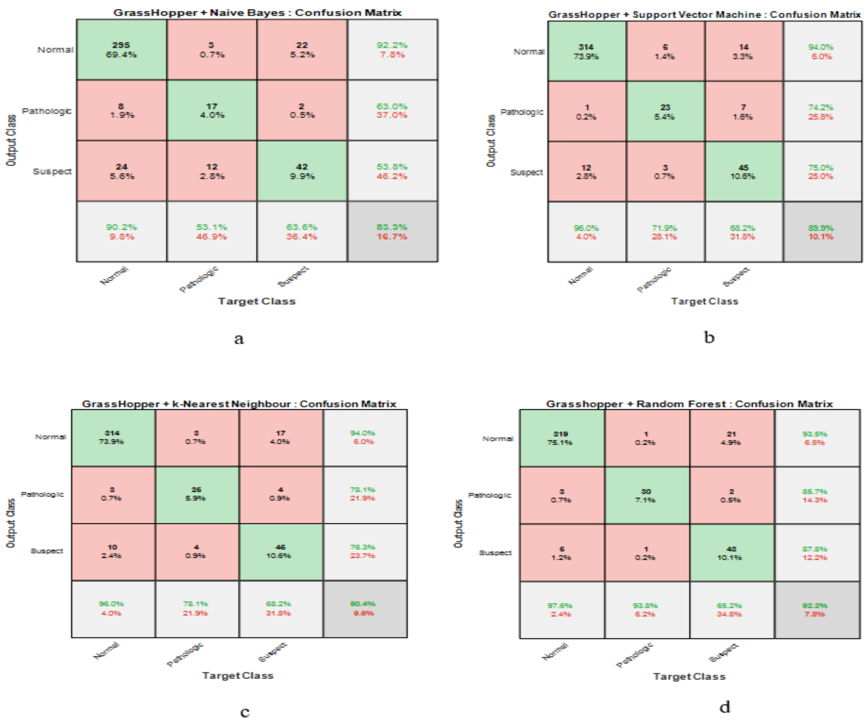


Figure 8. Confusion plot of Grasshopper Optimization with a) Naïve Bayes b) Support Vector Machine c) k-Nearest Neighbor d) Random Forest.

The confusion plots of Grey wolf Optimization technique with four machine learning models is depicted in the Figure 9 below. This plot gives the overall accuracy of the model with appropriate TP, TN, FP, FN values for the considered CTG dataset with three classes. Figure 10 below gives the Macro Averaged ROC-AUC curves of Firefly optimization while Figure 11 gives the same for Grasshopper and Figure 12 for Grey Wolf optimization techniques with four machine learning models.

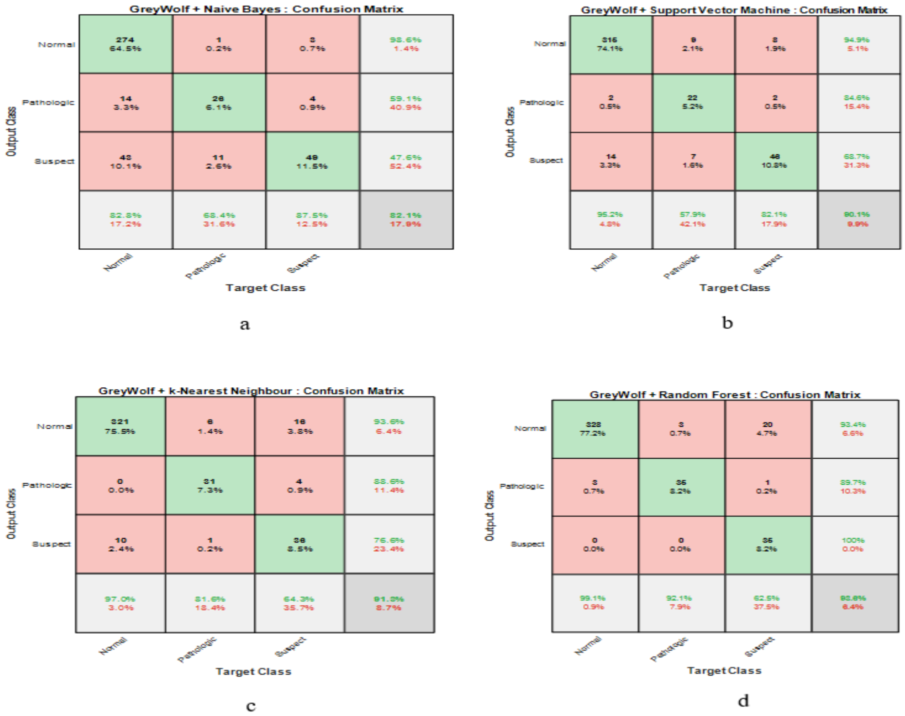


Figure 9. Confusion plot of Grey Wolf Optimization with a) Naïve Bayes b) Support Vector Machine c) k-Nearest Neighbor d) Random Forest.

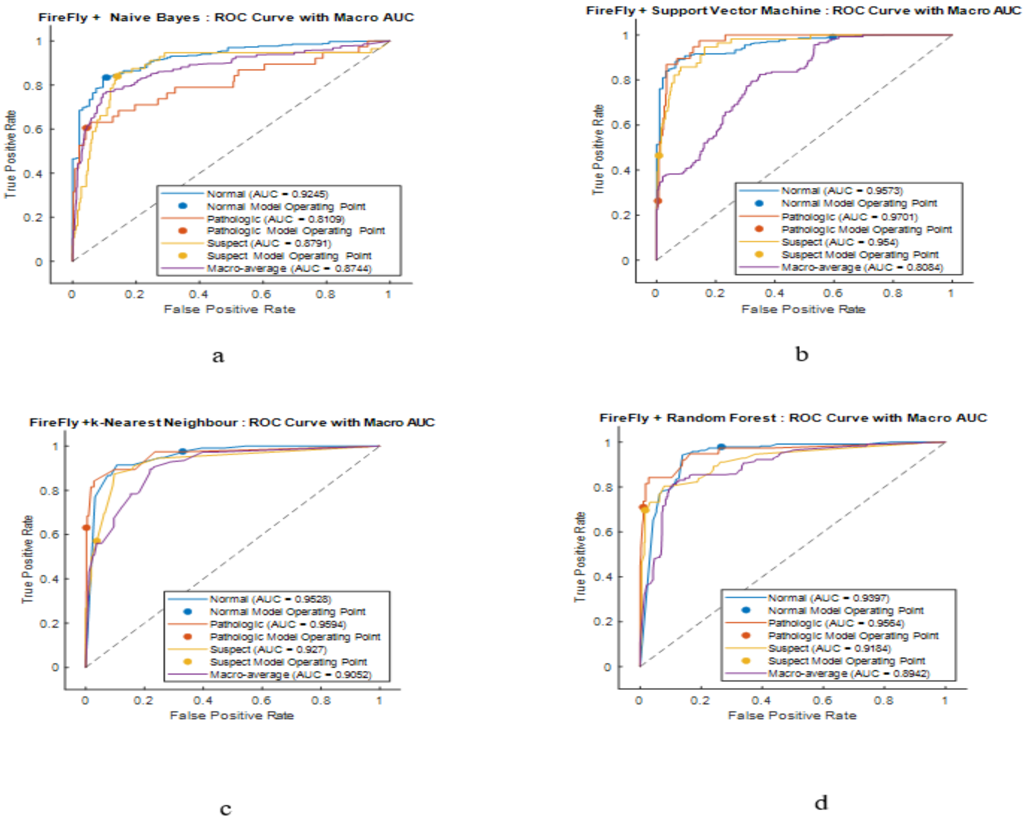


Figure 10. Macro Averaged ROC-AUC curves of Firefly Optimization with a) Naïve Bayes b) Support Vector Machine c) k-Nearest Neighbor d) Random Forest.

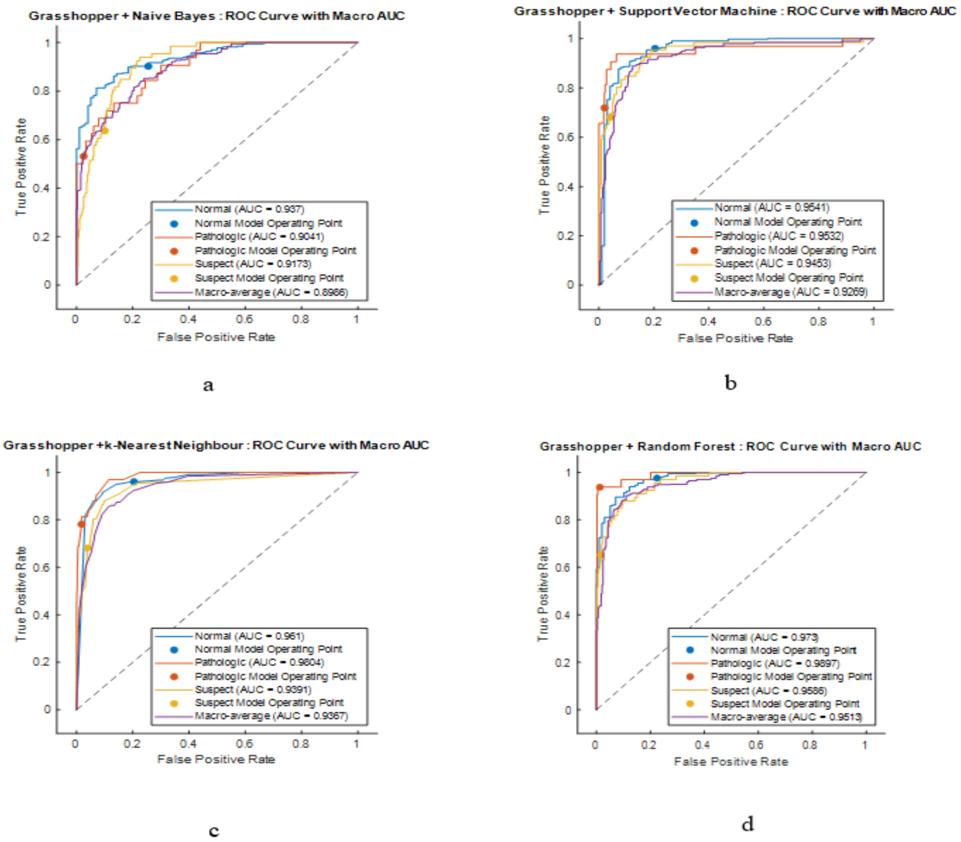


Figure 11. Macro Averaged ROC-AUC curves of Grasshopper Optimization with a) Naïve Bayes b) Support Vector Machine c) k-Nearest Neighbor d) Random Forest.

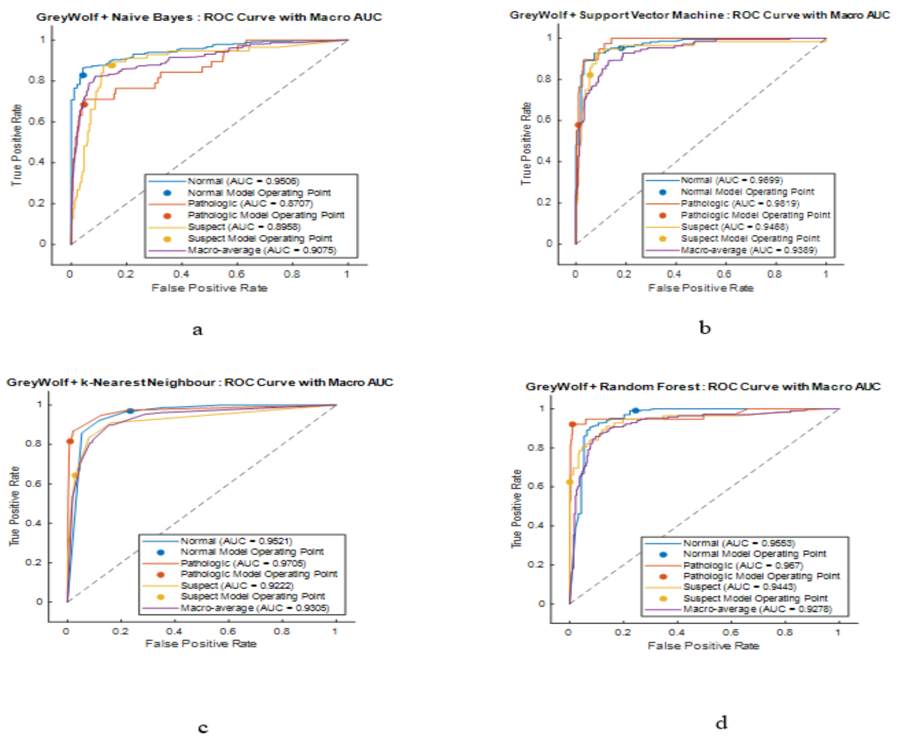


Figure 12. Macro Averaged ROC-AUC curves of Grey Wolf Optimization with a) Naïve Bayes b) Support Vector Machine c) k-Nearest Neighbor d) Random Forest.

V Results and Discussion

The overall performance metrics of the three Swarm-based Metaheuristic Optimization techniques with four considered machine learning algorithms are given below. Table 2 depicts the Overall Performance metrics of Firefly Optimization with ML models. Overall Accuracy is high for Random Forest classifier with 91.8%, mean MCC of 76.27% and mean Kappa score of 75.70%.

Table 2. Overall Performance Metrics of the Firefly optimization + Machine Learning Algorithms

Performance Metrics	Naïve Bayes	SVM	k-NN	RF
Overall Accuracy	81.41%	85.65%	89.18%	91.8%
Classification Error	18.59%	14.35%	10.82%	8.2%
Weighted F_1 score	79.88%	88.20%	89.82%	92.14%
Mean MCC	58.36%	53.73%	68.75%	76.27%
Mean Kappa Score	56.54%	48.05%	67.63%	75.70%
Averaged AUC	87.44%	80.84%	90.52%	89.42%
Elapsed Time	5.14s	5.13s	4.6s	5.9s

Table 3 depicts the Overall performance metrics of Grasshopper optimization with four ML models. Of all the four Random Forest classifier gave better Overall Accuracy of 92.24% with better mean MCC and mean Kappa score comparative with Random Forest of firefly optimization. Table 4 below depicts the Overall performance metrics of Grey wolf optimization with four ML models of which Random Forest model outperforms the rest with overall accuracy of 93.65%, mean MCC of 82.86% and mean Kappa score of 81.7%. Comparing Elapsed time of all the 12 hybrid models, 4.6 seconds in Table 2 is the least which is obtained for Firefly optimization with k-NN algorithm. The rest all models have on an average elapsed time of 6 seconds which is considerable.

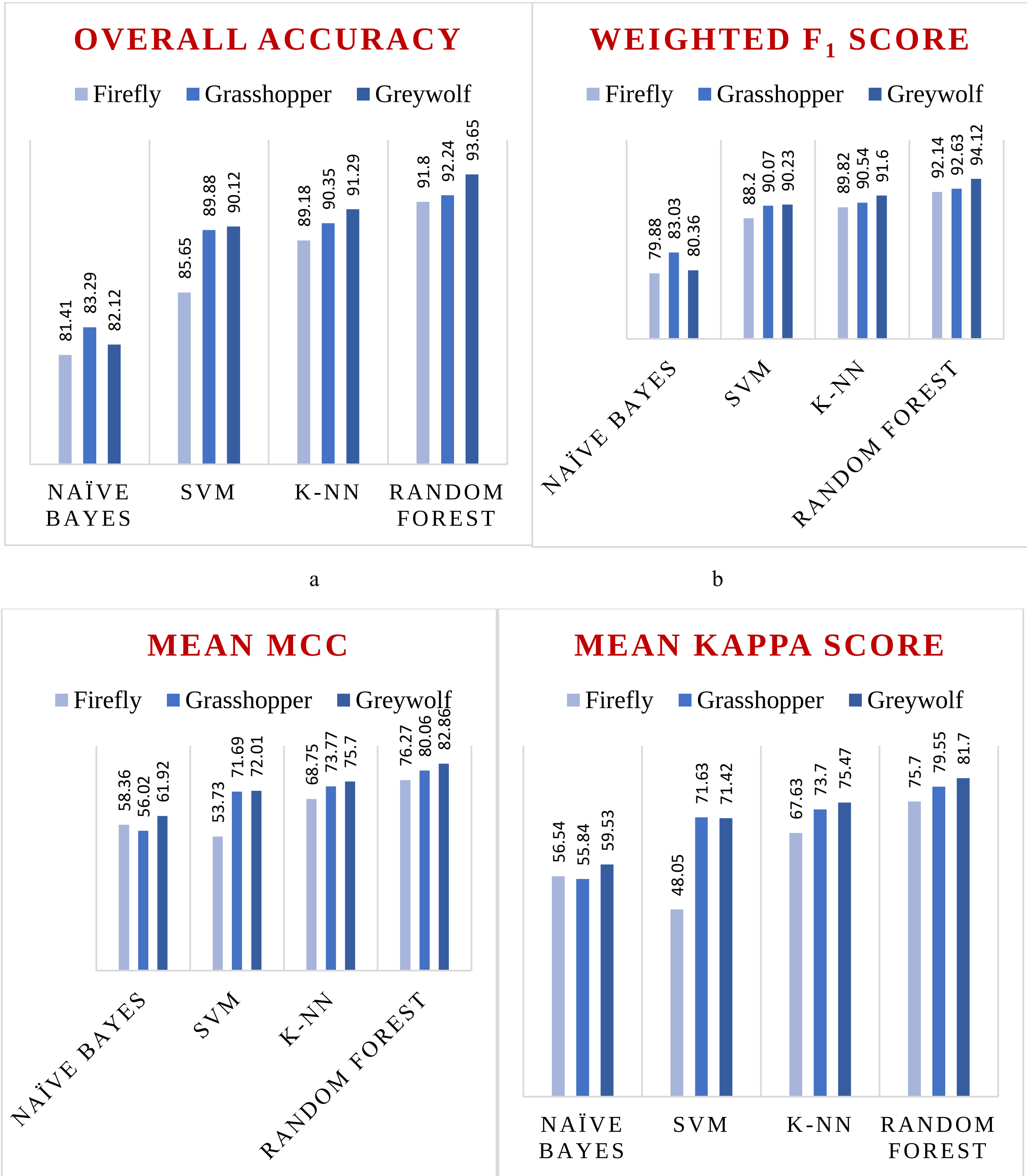
Table 3. Overall Performance Metrics of the Grasshopper optimization + Machine Learning Algorithms

Performance Metrics	Naïve Bayes	SVM	k-NN	RF
Overall Accuracy	83.29%	89.88%	90.35%	92.24%
Classification Error	16.71%	10.12%	9.65%	7.76%
Weighted F_1 score	83.03%	90.07%	90.54%	92.63%
Mean MCC	56.02%	71.69%	73.77%	80.06%
Mean Kappa Score	55.84%	71.63%	73.70%	79.55%
Averaged AUC	89.86%	92.69%	93.67%	95.13%
Elapsed Time	6.8s	4.7s	4.8s	5.8s

Table 4. Overall Performance Metrics of the Grey Wolf optimization + Machine Learning Algorithms

Performance Metrics	Naïve Bayes	SVM	k-NN	RF
Overall Accuracy	82.12%	90.12%	91.29%	93.65%
Classification Error	17.88%	9.9%	8.71%	6.35%
Weighted F_1 score	80.36%	90.23%	91.60%	94.12%
Mean MCC	61.92%	72.01%	75.70%	82.86%

Mean Kappa Score	59.53%	71.42%	75.47%	81.70%
Averaged AUC	90.75%	93.89%	93.05%	92.78%
Elapsed Time	4.83s	4.83s	5.06s	6.3s



c d

Figure 13. Comparison graphs of Overall performance metrics a) Overall Accuracy b) Weighted F_1 score c) Mean MCC d) Mean Kappa score

Since the CTG dataset considered from UCI machine learning repository is imbalanced dataset, the values of mean MCC and mean Kappa score is not overwhelming while overall accuracy and weighted F_1 scores are outperforming when compared with other state of the art techniques. Figure 13 gives the comparison graph of the overall performance metrics of the 12 used models of which Grey wolf with Random Forest model outperforms the rest of the models. Figure 14 below depicts the performance of the Random Forest model with considered three Swarm based Metaheuristic optimization techniques. It shows clearly that Overall accuracy is high for Greywolf optimization technique while the average AUCROC is high for Grasshopper optimization technique. From Figure 6 number of features and elapsed time is less for Grasshopper optimization and ROCAUC is more for Random Forest with Grasshopper optimization, but overall accuracy of classification is not high. This shows that when a model is required which is economical, better to prefer Grasshopper optimization, at the same time which is more accurate prefer Grey wolf optimization with Random Forest classifier.

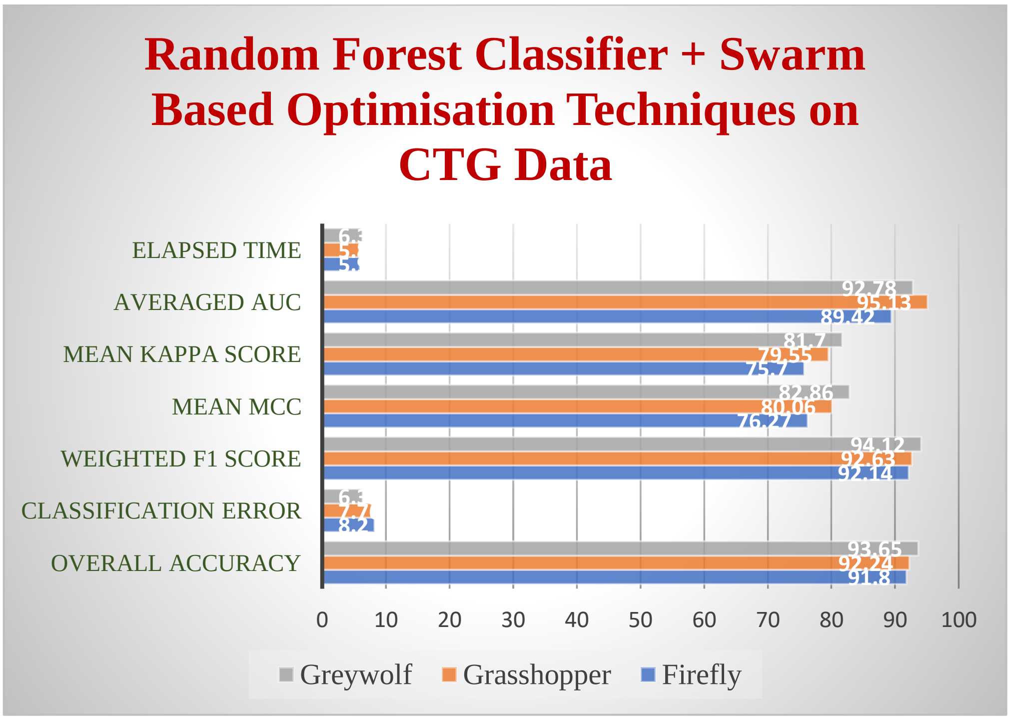


Figure 14. Comparison of Overall Performance Metrics of Random Forest Classifier with three Swarm based Metaheuristic Optimisation Techniques on CTG dataset.

In this study, three Swarm based Metaheuristic optimization techniques are used to select the features set of CTG dataset taken from UCI machine learning repository. After feature selection the data split into train and test data with 80 and 20 percent ratio respectively and is applied to four machine learning models to analyse the results. Confusion matrix for each model is obtained which is multiclass classification problem. From this confusion matrix various performance metrics were evaluated and a satisfied performance of accuracy was obtained for the proposed Grey wolf optimization with Random Forest classifier and best value for AUCROC metric is obtained for Grasshopper optimization with Random Forest classifier.

The Table 5 shows the comparison of the proposed metho with other state of the art models. Improta G et al. [4] used Random Forest classifier to recognize normal and abnormal samples and achieved an overall accuracy of 87.6%, precision of 87.9% and AUCROC value of 93%. Ricciardi et al. [5] used Random Forest classifier only with Ada Boosting technique as hybrid model and achieved AUCROC value greater than 94.9% to classify normal and abnormal samples. Öztürk et al. [11] used Empirical Mode Decomposition with ReliefF optimization technique and SVM classifier on binary classes which achieved an overall accuracy of 90.0%. N. Chamidah *et al.* [24] used Hybrid k-means and SVM classifier on CTG multiclass data and achieved an overall accuracy of 90.64%.

The proposed method used Greywolf optimization with Random Forest classifier which outperforms in terms of overall accuracy with **93.65%** and weighted F_1 score of **94.12%** and Grasshopper optimization with RF classifier gave a n AUCROC value of **95.13%** which is high when compared to other state of the art techniques as mentioned in the Table 5.

Table 5. Summary of Comparison with State-of-the-art models.

References	Method	Performance	Class
Improta G et al. [4]	Random Forest	A = 87.6%, Precision = 87.9%, AUCROC = 93%	Normal or Abnormal (Binary Class)
Ricciardi et al. [5]	Random Forest + AdA-B	AUCROC > 94.9%	Normal or Abnormal (Binary Class)
Öztürk et al. [11]	EMD + ReliefF + SVM	A = 90.0%	Normal or Abnormal (Binary Class)
N. Chamidah *et al.* [24]	Hybrid k-Means + SVM	A = 90.64%	Normal, Suspect, Pathological (Multi class)
Proposed Method*	Grey Wolf Optimization + Random Forest (DT with Bagging) Grasshopper Optimization + Random Forest (DT with Bagging)	**A = 93.65%** **MCC = 82.86%** **Kappa = 81.7%** **AUC = 92.78%** **F_1 score=94.12%** A = 92.24% MCC = 80.06% Kappa = 79.55% **AUC = 95.13%** F_1 score=92.63%	Normal, Pathological and Suspect (Multi class)

VI Conclusions and Future Work

The application of machine learning algorithms combined with swarm-based metaheuristic optimization for classifying Cardiotocogram (CTG) data signifies advancement in the field of autonomous fetal distress detection. The proposed method uses firefly, grasshopper and grey wolf optimization techniques to feature select the dataset and apply them to machine learning algorithms which compares the twelve hybrid models of the classification. Of these models, Greywolf optimization with Random Forest gave best overall accuracy of 93.65% weighted F_1 score of 94.12%, mean MCC of 82.86%, mean Kappa score of 81.7% with elapsed time of 6.3 seconds. While Grasshopper optimization of the proposed model achieved a better AUCROC value of 95.13%. This hybrid approach enhances the accuracy, reliability, and efficiency of fetal monitoring systems. The mean MCC and mean Kappa score values can be further improved by making the data as balanced data using Generative AI techniques. Hence the application of machine learning and swarm-based metaheuristic optimizations in CTG data classification can significantly contribute to the advancement of autonomous fetal distress detection, ultimately improving maternal and fetal health outcomes.

References

[1] Georgoulas, G., Stylios, C., & Groumpos, P. “Predicting the risk of metabolic acidosis for newborns based on fetal heart rate signal classification using support vector machines”. *IEEE Transactions on Biomedical Engineering*, 53(5), 875-884, 2006.

[2] Jezewski, J., Horoba, K., Matonia, A., & Wrobel, J, “The use of artificial neural networks for fetal monitoring”. *Advances in Intelligent and Soft Computing*, 84, 321-332, 2011.

[3] Ocak, H, "A medical decision support system based on support vector machines and the genetic algorithm for the evaluation of fetal well-being. *Journal of Medical Systems*, 37(2), 9912, 2013.
[4] Improta G., Ricciardi C., Amato F., D'Addio G., Cesarelli M., Romano M. "Efficacy of Machine Learning in Predicting the Kind of Delivery by Cardiotocography". *XV Mediterr. Conf. Med. Biol. Eng. Computing-MEDICON* 76, 793–799. 10.1007/978-3-030-31635-8_95, 2019.
[5] Ricciardi C., Improta G., Amato F., Cesarelli G., Romano M. "Classifying the Type of Delivery from Cardiotocographic Signals: A Machine Learning Approach". *Comput. Methods Programs Biomed.* 196, 105712, 2020.
[6] V. Chudáèek, J. Andén, S. Mallat, P. Abry and M. Doret, "Scattering transform for intrapartum fetal heart rate characterization and acidosis detection", *Proc. 35th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, pp. 2898-2901, 2013.
[7] V. Chudáèek et al., "Low dimensional manifold embedding for scattering coefficients of intrapartum fetale heart rate variability", *Proc. 36th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, pp. 6373-6376, 2014.
[8] Diab, A., et al. "Cardiotocography feature extraction using wavelet transform". *Medical & Biological Engineering & Computing*, 51, 1131-1143, 2013.
[9] Chudáček, V., et al. "Automatic evaluation of intrapartum fetal heart rate recordings: A comprehensive analysis of useful features" *Physiological Measurement*, 35, 1143-1161, 2014.
[10] Silva, I., et al "Hybrid machine learning model for the classification of cardiotocograms using genetic algorithms and support vector machines". *Applied Soft Computing*, 92, 106280, 2020.
[11] S. Öztürk, S. A. Şahin, A. N. Aksoy, B. Ari and A. Akinbi, "A Novel Approach for Cardiotocography Paper Digitization and Classification for Abnormality Detection," in IEEE Access, vol. 11, pp. 42521-42533, 2023, doi: 10.1109/ACCESS.2023.3271137.
[12] Zhang, Q., et al. "Deep learning-based cardiotocography analysis for fetal hypoxia diagnosis". *IEEE Access*, 7, 159064-159074, 2019.
[13] Rattadilok, P., et al. "Feature selection and classification using PSO for intrapartum fetal heart rate monitoring". *Biomedical Signal Processing and Control*, 27, 54-61, 2016.
[14] Ayres-de Campos D, Bernardes J, Garrido A, Marques-de-Sá J, Pereira-Leite L. SisPorto 2.0: a program for automated analysis of cardiotocograms. J Matern Fetal Med. 2000 Sep-Oct;9(5):311-8. doi: 10.1002/1520-6661(200009/10)9:5<311::AID-MFM12>3.0.CO;2-9. PMID: 11132590.
[15] Campos,D. and Bernardes,J.. (2010). Cardiotocography. UCI Machine Learning Repository. https://doi.org/10.24432/C51S4N.
[16] S.-B. Kim, K.-S. Han, H.-C. Rim and S. H. Myaeng, "Some effective techniques for Naive Bayes text classification", IEEE Trans. Knowl. Data Eng., vol. 18, no. 11, pp. 1457-1466, Nov. 2006.
[17] T. Denoeux, "A k-nearest neighbor classification rule based on Dempster–Shafer theory", IEEE Trans. Syst. Man. Cybern., vol. 25, no. 5, pp. 804-813, May 1995.
[18] L. B. Marinho, N. D. M. M. Nascimento, J. W. M. Souza, M. V. Gurgel, P. P. R. Filho and V. H. C. de Albuquerque, "A novel electrocardiogram feature extraction approach for cardiac arrhythmia classification", Future Gener. Comput. Syst., vol. 97, pp. 564-577, Aug. 2019.
[19] Mikel Galar, Alberto Fernández, Edurne Barrenechea, Humberto Bustince, and Francisco Herrera. 2011. An overview of ensemble methods for binary classifiers in multi-class problems: Experimental study on one-vs-one and one-vs-all schemes. Pattern Recogn.44, 8 (August, 2011), 1761–1776. DOI: https:// doi.org/10.1016 /j. patcog.2011.01.017.
[20] A. Liaw and M. Wiener, "Classification and regression by random forest", R Newslett., vol. 2, no. 3, pp. 18-22, 2002.
[21] A. H. Gandomi, X-S. Yang, S. Talatahari and A. H. Alavi, "Firefly algorithm with chaos", *Commun. Nonlinear Sci. Numer. Simul.*, vol. 18, no. 1, pp. 89-98, Jan. 2013.
[22] Y. Meraihi, A. B. Gabis, S. Mirjalili and A. Ramdane-Cherif, "Grasshopper Optimization Algorithm: Theory, Variants, and Applications," in IEEE Access, vol. 9, pp. 50001-50024, 2021, doi: 10.1109/ACCESS.2021.3067597.
[23] S. Mirjalili, S. M. Mirjalili and A. Lewis, "Grey wolf optimizer", *Adv. Eng. Softw.*, vol. 69, pp. 46-61, Mar. 2014.
[24] N. Chamidah and I. Wasito, "Fetal state classification from cardiotocography based on feature extraction using hybrid K-Means and support vector machine," 2015 International Conference on Advanced Computer Science and Information Systems (ICACSIS), Depok, Indonesia, 2015, pp. 37-41, doi: 10.1109/ICACSIS.2015.7415166.