Journal of Harbin Engineering University
ISSN: 1006-7043

Vol 46 No. 6
June 2025

An Innovative Framework for Recommending Features in Cardiotocography for Prenatal Care

[*1]Pragyan Paramita Das, [2]Jayashree Piri,[3]Raghunath Dey,[4]Biswaranjan Acharya,

Department of Computer Science & Engineering, Silicon University, Bhubaneswar, Odisha, India,

Department of Computer Science & Engineering, Silicon University, Bhubaneswar, Odisha, India

School of Computer Science and Engineering, KIIT Deemed to be University, Bhubaneswar, Odisha, India, Department of Computer Engineering, Marwadi University, Rajkot, India

Abstract

Introduction: Health Recommender Systems (HRSs) have become essential instruments in medicine, assisting healthcare practitioners with the prompt detection of diseases and delivering tailored health advice. In the realm of maternal health, namely prenatal care, there is an increasing demand for intelligent technologies that can facilitate the precise evaluation of fetal well-being.

Objectives: This work seeks to create a customized health recommender system centered on prenatal care through the analysis of Cardiotocography (CTG) data. The aim is to forecast the probability of negative fetal outcomes and determine the most significant factors impacting fetal health during gestation.

Methods: A novel multi-objective Grey Wolf Optimization (MOGWO) algorithm was used for feature selection to enhance model accuracy and interpretability. The CTG dataset was processed to identify significant parameters, and various machine learning classifiers were trained and evaluated. Particular emphasis was placed on assessing the predictive performance of Random Forest and Decision Tree classifiers. The predictive performance of Random Forest and Decision Tree classifiers was specifically evaluated.

Results: The MOGWO-based feature selection revealed that metrics including Prolonged Decelerations (PD), Abnormal Short-Term Variability (ASTV), Abnormal Long-Term Variability (ALTV), Accelerations (AC), and Mean Long-Term Variability (MLTV) are important in predicting fetal outcomes. The Random Forest classifier has the highest classification accuracy of 95.61%, followed by the Decision Tree classifier at 93.46%.

Conclusions: The study highlights the usefulness of intelligent systems such as HRSs in prenatal care, particularly in interpreting CTG data to predict fetal well-being. This study helps pregnant mothers make better clinical decisions and reduce risk by identifying crucial diagnostic variables and using robust classifiers.

Keywords: Health Recommender System (HRS), MOGWO, Fetal Heart Rate (FHR), CTG

1. Introduction

The wealth of digital patient data in the healthcare domain offers a chance for efficient information extraction and disease prediction, which can greatly improve patient-focused decision-making procedures. Recommender systems are essential for giving patients clear and understandable information about their medical issues, which promotes greater comprehension and well-informed decision-making. To protect the health of the mother and fetus during pregnancy, prenatal care is essential. A crucial instrument for tracking fetal heart rate and uterine contractions, prenatal Cardiotocography (CTG) [1] enables prompt identification of any anomalies or distress. However, because of its complexity and variability, effectively interpreting CTG data is still difficult. By giving healthcare professionals immediate insights for well-informed decision-making, these systems seek to enhance the interpretation of CTG data. This work introduces a distinctive health recommendation system that is specifically designed for fetal cardiotocography. To maximize the extraction of meaningful information from CTG signals while decreasing noise, our method maximizes feature selection through the application of Multi-Objective Grey Wolf Optimization (MOGWO). Among the many benefits of this approach is its ability to simultaneously optimize various goals, like feature subset size and classification accuracy, which results in the development of strong prediction models. Additionally, our solution gives medical professionals tailored advice and tools for decision support, enabling them to make well-informed clinical decisions based on thorough CTG analysis. The motive of this work is:

-To analyze and explore various feature selection approaches already used to identify crucial CTG features.

-To suggest a multi-objective Grey Wolf Optimization-based feature recognition technique for CTG data.
-To assist pregnant mothers by educating them about traits that are critical to the fetal health prognosis during pregnancy.
-To recommend prediction models for a more precise evaluation of the probability of adverse fetal Cardiotocography (CTG) outcomes.

In the research [2], several health recommender systems (HRSs) have been proposed to cope with top N predictions and recommendations. A proposed survey focused on various techniques and assessment

procedures for the Health Recommender System. The medication response prediction was initially presented as a challenge for the recommendation system (RS) [3]. A proposed hybrid recommendation system utilizes a limited Boltzmann machine [4] and deep neural networks to gain a valuable understanding of the optimal utilization of large-scale data analytics. The RS [5] was established to provide a compilation of medications that should be avoided while using a prescription, as they have the potential to induce unfavorable drug reactions. A hybrid recommender system [6] has provided family physicians with a set of suggestions for each patient, including the temporal complexity of their encounters. A k-clique integrated deep learning classifier [7] is incorporated into the diet recommendation system to offer patients food product recommendations tailored to their weight, blood pressure, cholesterol, sugar level, and other health information. In a multi-objective optimization problem, numerous objective functions must be optimized or minimized. A concept-based location recommender system with several objectives that take into account competing factors like personal and collective preferences is recommended in the study [8]. In the research [9], a novel menu recommendation system, employing a multi-objective approach to attain the best equilibrium among nutritional characteristics, harmony, and coverage of pantry products. Using the Movielens and Jester datasets [10] as test cases, a multi-objective evolutionary method based on decomposition optimizes two competing parameters: the recommender systems' accuracy and diversity.

In machine learning and statistics, feature selection refers to the process of selecting a subset of certain variables or features to be utilized in a model. Rehman et al. [11] have observed that parallel and adaptive classification systems offer a comprehensive explanation of the various feature selection strategies to predict chronic diseases. Tang et. al. [12] have implemented an automatic feature selection mechanism for insightful content. This technique is not influenced by the specific recommendation framework or features and is demonstrated in recommendation systems from different domains. Support vector machines (SVM) are among the most unduly optimistic classifier approaches; in contrast, it has been demonstrated that the Genetic Algorithm (GA) minimizes the number of characteristics that boost classifier performance [13]. An analysis of the literature shows that the majority of the research on HRS has depended on traditional algorithms for recommendation.

Moreover, there has been a noticeable lack of utilization of feature selection and multi-objective optimization methods in HRS research.

This attempt is organized as follows throughout the remainder of the article. Section 2 describes a detailed explanation of the research technique we have proposed. Section 3 is for visual representations of the experimental results, while Section 4 describes the findings of the study and provides recommendations for future research.

2. Materials and Methods

This section provides an explanation of the comprehensive methodology applied to this study.

a. Analysis of the Dataset

The system uses the Fetal CTG dataset from the source: https://archive.ics.uci.edu/dataset/193/cardiotocography [14]. It contains data from cardiotocograms, which gauge uterine contractions (UC) and the fetal heart rate (FHR). This dataset comprises 21 features used in measuring fetal heart rate and uterine contractions on CTG. Key physiological parameters are LB (fetal heart rate baseline), AC (number of accelerations), and FM (fetal movement). Furthermore, included in the dataset are features showing transient decreases in fetal heart rate (FHR) in response to uterine stress, namely uterine contraction (UC) and four types of decelerations: moderate (DL), severe (DS), and protracted (DP). Heart rate variation is evaluated using a dataset comprising ASTV (average short-term variability), MSTV (mean short-term variability), ALTV (average long-term variability), and MLTV (mean long-term variability). Furthermore, included in the statistical study of the FHR signal are histogram-based features like width, minimum, maximum, number of maxima, number of zeros, mode, mean, median, variance, and trend. These features taken together provide a complete picture of fetal well-being and point up possible flaws. A total of 2,126 samples were used for the experimental analysis. We have considered 'NSP' (Normal/Suspected/Pathological) as the target class, which is distributed as follows: 1655 Normal, 295 Suspected, and 176 Pathological classes. The dataset is pre-processed to handle the missing values and imbalances by using SMOTE (Synthetic Minority Oversampling Technique). This

the method helps to ensure the model doesn't become biased toward the dominant class, improving overall classification performance.

b. Multi-objective Grey Wolf Optimization

In multi-objective optimization, a discipline of decision-making involving several elements, the aim is to simultaneously optimize several objective functions. Different evolutionary approaches help to maximize several goals. Introducing the Multi-Objective Grey Wolf Optimization (MOGWO) [15,16], an evolutionary algorithm for multi-objective optimization challenges inspired by the social behavior and hunting tactics of grey wolves. MOGWO efficiently explores and exploits the solution space by using cooperative hunting behavior and a hierarchy based on dominance. The algorithm balances multiple goals to produce a grey wolf population and steer their evolution.

c. Feature Selection

Machine learning uses feature selection [17] to select the most relevant input information to enhance model accuracy and speed up training. This method removes irrelevant, outdated, or redundant variables to improve model prediction. We pick features using the multi-objective Grey Wolf Optimizer (MOGWO) in the suggested model. MOGWO provides an advanced method for selecting features, utilizing the benefits of evolutionary algorithms, and simultaneously dealing with numerous objectives. Using MOGWO, our objective is to effectively choose the most relevant features while minimizing complexity compared to conventional feature selection methods. Our solution uses the wrapper technique along with the multi-objective Grey Wolf Optimizer to choose features. The decision tree (DT) classifier serves as our model for feature selection, harnessing the capabilities of both the wrapper technique and MOGWO to boost the accuracy and efficiency of our prediction model.

d. Framework of Proposed MOGWO-RS

The prenatal Cardiotocography recommender system is essential for saving the lives of women. The system suggested in Figure 1 displays the proposed system's design, which utilizes MOGWO for feature selection. The competing aims of minimizing parameters and maximizing accuracy are considered. The proposed approach is a two-step process.

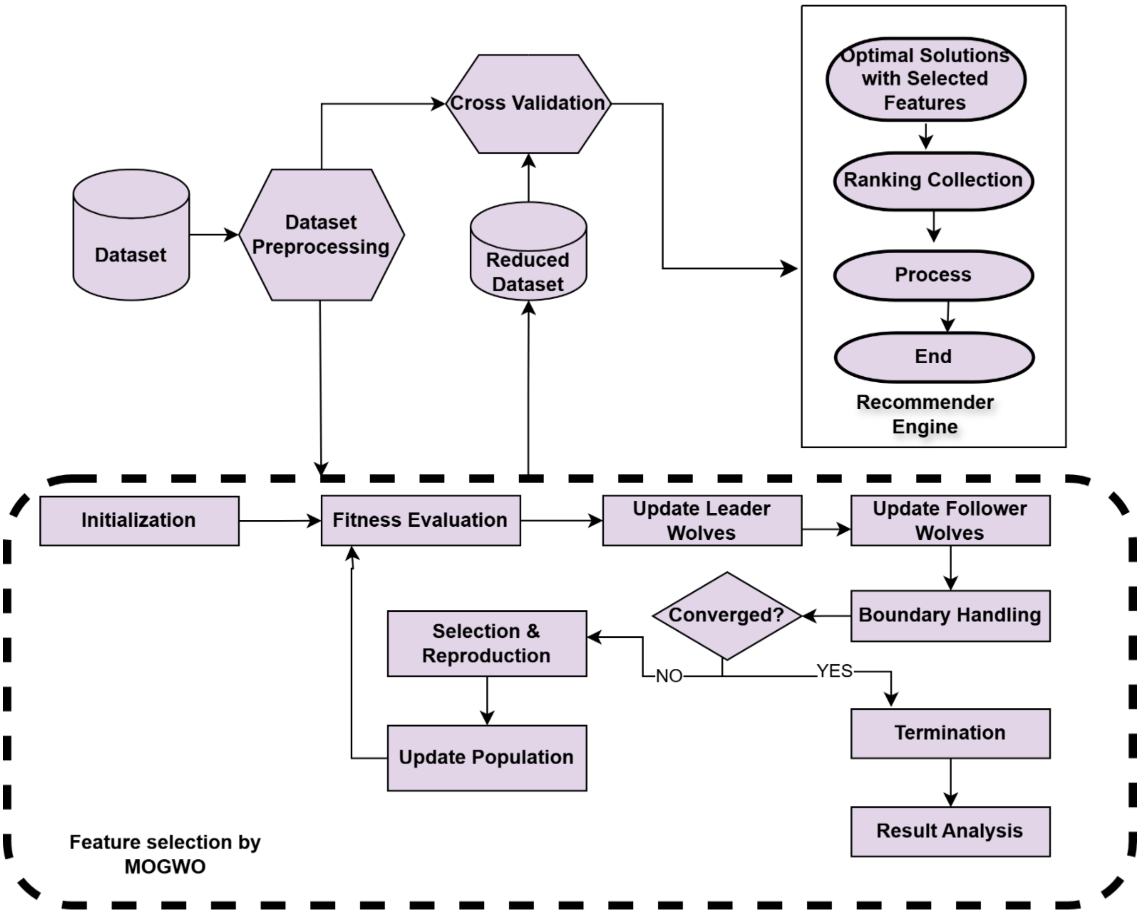

Fig. 1 Proposed System's Framework using the Multi-Objective Grey Wolf Optimization Algorithm

Step 1: Feature Selection based on MOGWO

-Initialization: Each wolf must be assigned a binary string (0 or 1) generated randomly. Wolves are randomly placed throughout the search space for exploration. The alpha (α) wolf represents the best solution among others. As such, beta (β) and delta (δ) correspondingly are the second and third best answers. Assumed to be omega (ω) are the remaining possible solutions. α, β, and δ wolves direct the hunting optimization in the GWO method. The

ω wolves copy these three wolves. By specifying the goals and constraints of the multi-objective optimization problem, the maximum iteration count as well as other algorithm parameters (like population size and convergence criterion) are initialized.

- Fitness Evaluation: The fitness of each grey wolf in the population is determined using objective functions given in equations 1 and 2, respectively.

$O1$= Minimize the *number of features*

$$\sum_{i=1}^{l} x_i \quad (1)$$

Where l is the dimension of the original dataset and x_i is the i^{th} bit value of the grey wolf's position vector X.

$O2$ = Maximize the *accuracy of the wrapper classifier*(*KNN*)

$$1/N \sum_{i=1}^{k} n_{ii} \quad (2)$$

Where N is the total number of samples in the training set, k is the number of classes (here k=3), and n_{ii} is the total number of correct predictions for the class i.

- Storing the Non-dominated Solutions: After calculating wolf fitness, Pareto solutions are saved in an external repository. None of the existing population's solutions dominate Pareto solutions.

-Update Wolves:

In GWO, wolves update their positions during the optimization process based on the leadership hierarchy (α, β, δ) and hunting behavior (encircling, hunting, attacking). The encircling behavior is mathematically modeled by (3):

$$X(t+1) = Xp(t) + A|C.Xp(t) - X(t)| \quad (3)$$

$$A = 2a.r1 - a \quad (4)$$

$$C = 2.r2 \quad (5)$$

X stands for the wolf's position in GWO; Xp is the whereabouts of the prey. A and C are coefficient vectors modeled by (4) and (5) respectively; r1 and r2 are random vectors in [0, 1]. The parameter a reduces linearly from 2 to 0 over iterations, hence controlling exploration.

To simulate the hunting procedure, the updated positions relative to α, β, and δ wolves are given in (6), (7), and (8), respectively

$$X1 = X\alpha - A1.|C1.X\alpha - X| \quad (6)$$

$$X2 = X\beta - A2.|C2.X\beta - X| \quad (7)$$

$$X3 = X\delta - A3.|C3.X\delta - X| \quad (8)$$

$$X(t+1) = (X1 + X2 + X3)/3 \quad (9)$$

The calculated values of $A1$, $A2$, and $A3$ exhibit similarities to those of A in (4), whereas the calculated values of $C1$, $C2$, and $C3$ demonstrate similarities to C in (5). Finally, the wolf's updated position is determined by (9).

The equation of updating position in (9) must be modified if the search candidates can migrate in a binary search space, shown in (10).

$$X(t+1) = \begin{cases} 1 & \text{if } sigmoid(X(t+1)) >= rand \\ 0 & \text{otherwise} \end{cases} \quad (10)$$

Where $rand$ is a random number drawn from a uniform distribution [1,0], and $sigmoid\ (a)$ is calculated by (11).

$$sigmoid\ (a) = \frac{1}{1+e^{-10(x-0.5)}} \quad (11)$$

One way to describe the grey wolves' attacking mechanism is to use a vector a, which is a random vector having elements in the range [−a, a]. The elements of a drop linearly from 2 to 0 with each iteration, and may be expressed in (12):

$$a = 2 - t.\frac{2}{maxIt} \quad (12)$$

Where t = current iteration and maxIt = maximum iterations to run.

-Termination: When the algorithm's termination conditions, such as reaching the maximum number of iterations or achieving convergence, are satisfied, then it is terminated.

-Final Output of MOGWO-based Feature

Selection: The final Pareto front that was obtained when the algorithm terminated is examined. Based on the decision-maker's preferences and trade-offs between competing goals, the best course of action is selected.

This updating mechanism allows the search agents to balance exploration and exploitation effectively, accelerating convergence toward the optimal feature subset.

Step 2: Feature Recommendation

The method recommends the top-N characteristics, where N is the number of prenatal cardiotocography prognosis parameters, protecting pregnant moms and their fetuses. It also recommends the top five prenatal cardiotocography prediction models. Features are ranked by optimal solution frequency and target attribute correlation. Building predictive models uses reduced datasets to discover the best solutions. Other classifiers are employed with these datasets to improve prediction accuracy. A smaller dataset and classifier pair are assumed. Models are ranked by accuracy. The top 5 models are used to recommend more models.

3. RESULTS AND DISCUSSIONS

As a result of applying the feature selection approach to the dataset, a set of reduced datasets is produced. It is noted that the accuracy of MOGWO with 5 features is significantly higher than that of the Multi-Objective Genetic Algorithm (MOGA) with 8 features, as demonstrated in Figure 2. The MOGWO-based feature selection strategy is implemented for the CTG dataset over 50 iterations, using

a population size of 30. Consequently, 5 sets of optimal solutions were obtained. The most efficient options are F1 (with 5 features), F2 (with 4 features), F3 (with 2 features), F4 (with 3 features), and F5 (with 7 features). The features common among many optimal solutions are listed in Table 1, selected based on their rankings. The features are ordered based on their scores, which are derived from their correlation values with the objective attribute and their frequency among the best options.

Table 1: Different possible optimal solutions

Feature Sets	Feature Names
F1(5)	AC, ASTV, MLTV, ALTV, DP
F2(4)	LB, ASTV, MLTV, DP
F3(2)	ASTV, Mean
F4(3)	ASTV, ALTV, Mean
F5(7)	AC, UC, DP, ASTV, Mean, Median

A scale from 1 to 9 is used to evaluate the features, according to the data in Table 2. The rating of the feature, denoted as Rating-1, is determined by the degree to which its correlation values align with the target characteristic. A feature that has a strong correlation value is assigned a higher rating value. On the other hand, Rating 2 indicates the ranking of the feature based on its frequency in the best solutions.

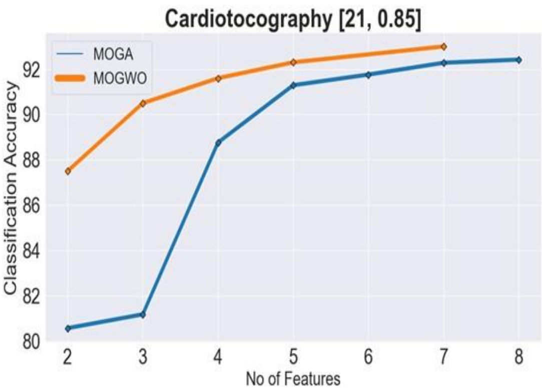

Fig. 2 MOGA Vs MOGWO

Features with high incidence values are assigned high ratings. To rank the features, the mean values of Rating-1 and Rating-2 are taken into consideration. The evaluation of the attributes can be found in Table 2. Features with a high average rating value are given a high rank, while highly ranked features are given a low rank number. If the

average rating value of the attributes is the same, the correlation value is used to resolve the tie. Out of the 22 features in the dataset, only 5 have average rating values that exceed five. Table 2 describes the top 5 recommended features based on their rank. The model's accuracy levels have determined the prescribed behavior. The model generation process employs six distinct classifiers. Figure 3 exhibits the six models with the greatest accuracy values: LR (Linear Regression), SVM (Support Vector Machine), KNN (K-Nearest Neighbor), Decision Tree (DT), Random Forest (RF), and GNB (Gaussian Naive Bayes).

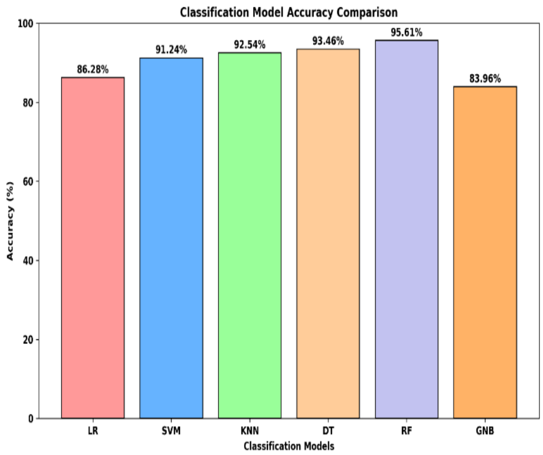

Fig. 3 Accuracy Measures of six models

The study found that PD, ASTV, ALTV, AC, and MLTV affect prenatal cardiotocography prognosis. To reduce prenatal cardiotocography risks, women must recognize these issues. According to studies, the Random Forest classifier model has the highest accuracy. Figure 3 shows that Decision Tree classifiers work well. The study proposes that women raise their knowledge of key signs that predict fetal heart rate anomalies to address this issue.

4. CONCLUSION AND FUTURE WORK

A Health Recommender System (HRS) customized for fetal cardiotocography (CTG) analysis is presented in this paper, offering a fresh approach to prenatal treatment. The use of machine learning models in conjunction with a feature selection technique based on MOGWO is the main contribution to the precise prediction of fetal health.

Table 2: Ranking of features

Feature Name	Correlation with the target	Rating-1	No. of Occurrences	Rating-2	Avg. Rating	Rank
PD	0.49	9	3	8	8.5	1
ASTV	0.47	8	5	9	8.5	2
ALTV	0.43	7	3	8	7.5	3
AC	0.36	6	2	7	6.5	4
MLTV	0.23	5	2	7	6	5
Mean	0.23	5	3	8	6.5	6
Median	0.21	4	1	6	5.5	7
UC	0.2	3	1	6	4.5	8
LB	0.15	2	1	6	4	9

PD, ASTV, ALTV, AC, and MLTV are the most important indicators in this study, which emphasizes feature relevance and interpretability in contrast to earlier research that frequently depended on large feature sets or traditional selection methods. With an accuracy of 95.61%, the suggested Random Forest-based model comes in second, followed by the Decision Tree model with 93.46%. Using the feature selection technique on the dataset generates the reduced datasets. It is observed that MOGWO with 5 features has much more accuracy than the Multi-Objective Genetic Algorithm (MOGA) with 8 features. This enhancement demonstrates how well MOGWO chooses the best attributes directly affecting model accuracy and clinical interpretability. Additionally, by integrating the prediction model into an HRS framework, this study provides a useful tool that helps physicians make diagnoses while also giving pregnant moms individualized information about the health of their fetus.By using other recent meta-heuristic algorithms and their hybrid versions for feature selection, this work can be expanded in the future. Additionally, by applying the proposed work to other datasets, it may be validated.

References

[1] Al-Yousif S, Jaenul A, Al-Dayyeni W, Alamoodi A, Najm IA, Tahir NM, Alrawi AA, Cömert Z, Al-Shareefi NA, Saleh AH. A systematic review of automated pre-processing, feature extraction and classification of cardiotocography. PeerJ Computer Science. 2021 Apr 27;7:e452.

[2] Sahoo AK, Pradhan C, Barik RK, Dubey H. DeepReco: deep learning based health recommender system using collaborative filtering. Computation. 2019 May 22;7(2):25.

[3] Alshammari A, Alshammari M. A Recommendation Engine Model for Giant Social Media Platforms using a Probabilistic Approach. Engineering, Technology & Applied Science Research. 2023 Oct 13;13(5):11904-10.

[4] Kirubahari R, Amali SM. An improved restricted Boltzmann machine using Bayesian optimization for recommender systems. Evolving Systems. 2024 Jun;15(3):1099-111..

[5] Çano E, Morisio M. Hybrid recommender systems: A systematic literature review. Intelligent data analysis. 2017 Jan 1;21(6):1487-524.

[6] Manoharan S. Patient diet recommendation system using K clique and deep learning classifiers. Journal of Artificial Intelligence. 2020 Jun 3;2(02):121-30.

[7] Kuanr M, Mohapatra P, Piri J. Health recommender system for cervical cancer prognosis in women. In2021 6th International Conference on inventive computation technologies (ICICT) 2021 Jan 20 (pp. 673-679). IEEE.

[8] Geng B, Li L, Jiao L, Gong M, Cai Q, Wu Y. NNIA-RS: A multi-objective optimization-based recommender system. Physica A: Statistical Mechanics and its Applications. 2015 Apr 15;424:383-97.

[9] Fkih F. Similarity measures for Collaborative Filtering-based Recommender Systems: Review and experimental comparison. Journal of King Saud University-Computer and Information Sciences. 2022 Oct 1;34(9):7645-69.

[10] El-Rahman SA, Saleh Alluhaidan A, AlRashed RA, AlZunaytan DN. Chronic diseases monitoring and diagnosis system based on features selection and machine learning predictive models. Soft Computing. 2022 Jul;26(13):6175-99.

[11] Zeeshan Z, Bhatti UA, Memon WH, Ali S, Nawaz SA, Nizamani MM, Mehmood A, Bhatti MA, Shoukat MU. Feature-based multi-criteria recommendation system using a weighted approach

with ranking correlation. Intelligent Data Analysis. 2021 Jan 1;25(4):1013-29.

[12] Tang J, Alelyani S, Liu H. Feature selection for classification: A review. Data classification: Algorithms and applications. 2014:37.

[13] Piri J, Mohapatra P. Imbalanced Cardiotocography Data Classification Using Re-Sampling Techniques. In Proceedings Of International Conference On Machine Intelligence And Data Science Applications: Midas 2020 2021 (Pp. 681-692). Springer Singapore.

[14] Campos, D. & Bernardes, J. (2000). Cardiotocography [Dataset]. UCI Machine Learning Repository. https://doi.org/10.24432/C51S4N.

[15] Mirjalili S, Saremi S, Mirjalili SM, Coelho LD. Multi-objective grey wolf optimizer: a novel algorithm for multi-criterion optimization. Expert systems with applications. 2016 Apr 1;47:106-19.

[16] Chandrashekar G, Sahin F. A survey on feature selection methods. Computers & electrical engineering. 2014 Jan 1;40(1):16-28.

[17] Saq AH, Zainal A, Al-Rimy BA, Alyami A, Abosaq HA. Intrusion Detection in IoT using Gaussian Fuzzy Mutual Information-based Feature Selection.Engineering, Technology & Applied Science Research. 2024 Dec 1;14(6):17564-71.