Journal of Engineering and Technology for Industrial Applications

ITEGAM-JETIA

Manaus, v.12 n.58, p.808-828. March/April, 2026.
DOI: https://doi.org/10.5935/jetia.v12i58.3078

ISSN ONLINE: 2447-0228

RESEARCH ARTICLE OPEN ACCESS

A MACHINE LEARNING AND HYBRID FEATURE SELECTION FRAMEWORK FOR PREDICTING PRETERM BIRTH IN NULLIPAROUS WOMEN USING ELECTRONIC MEDICAL RECORDS

T. Ummal Sariba Begum[1], R. Renuga Devi*[2]

[1, 2]Department of Computer Science and Applications, Faculty of Science and Humanities, SRM Institute of Science and Technology, Ramapuram Campus, Chennai, India

[1]http://orcid.org/0009-0000-2190-4226, [2]http://orcid.org/ 0000-0002-6197-236X

Email: *renugadr@srmist.edu.in, ut0110@srmist.edu.in

ARTICLE INFO

Article History
Received: December 08, 2025
Reviewed: January 10, 2026
Accepted: March 10, 2026
Published: April 30, 2026

Keywords:
Nulliparous women,
Pre term birth monitoring,
Deep learning,
Dragon Fly, Entropy,
Feature selection,
nuMoM2b.

ABSTRACT

Preterm birth (PTB) is a leading cause of neonatal mortality worldwide, underscoring the need for accurate early prediction. This study presents a machine learning framework combined with hybrid feature selection to predict PTB risk in nulliparous women using data from electronic medical records collected during prenatal consultations. Clinical, demographic, and physiological data were obtained from the publicly available nuMoM2b dataset, covering three gestational intervals: 6–13 weeks, 16–21 weeks, and 22–29 weeks. Data preprocessing and standardization were performed using Differential Evolution (DE) to enhance quality and improve model performance. A hybrid feature selection approach, integrating the Dragonfly Optimizer with entropy-based relevance ranking, was employed to identify informative and non-redundant predictors, reducing dimensionality and noise while maintaining clinical interpretability. A Multilayer Perceptron (MLP) classifier trained on the selected features differentiated term and preterm deliveries. The framework achieved 88.41% accuracy, AUC = 0.80, precision = 85.84%, recall = 87.24%, and F1-score = 88.15%. Incorporating ultrasound features such as cervical length and Plasticity Index further improved predictive performance. Notably, at the third prenatal consultation, the model reached 85.62% sensitivity for predicting very preterm infants. These findings highlight the importance of ultrasound measurements and demonstrate that integrating machine learning with evolutionary optimization and entropy-based feature selection can significantly enhance early PTB risk detection. The approach enables timely interventions for high-risk pregnancies, potentially improving maternal and neonatal outcomes. This study underscores the value of computational methods in clinical decision support and emphasizes how machine learning can transform prenatal care for nulliparous women.

I. INTRODUCTION

High-risk pregnant women can get prompt preventative measures to lower their risk of preterm birth by being identified early [1]. Although invasive screening and imaging tests have the potential to be successful screening techniques, they are still in the experimental stage due to their high cost, potential risks, and limited accessibility [2]. Non-invasive screening techniques using machine learning (ML) algorithms based on comprehensive pregnancy surveillance data alongside multilevel information connection to delivery records, without the need for any additional physiologic or imaging tests, have the potential to support clinical decision making to predict adverse pregnancy outcomes and guide pregnancy management [3]. The frequent medical care that survivors of preterm birth need, which has a detrimental effect on their quality of life and mental health, further strains the family economy and health care system [4]. Therefore, early detection of preterm birth is essential to enable medical professionals to perform prompt therapies to enhance infant outcomes. Because of the intricacy of the system, no set standards have been established for the prediction of preterm birth [5].

At the moment, evaluation of biochemical markers, cervical measurement, and risk factors serve as the primary foundation for clinical evaluation of PTB [6]. However, not all of these diagnostic methods are therapeutically acceptable for every pregnant woman because of their lack of safety or cost-effectiveness, even though they may be able to identify particular preterm birth characteristics [7]. For instance, examinations like cervical measurement and biochemical marker assessment will put the expectant mother under more physical and mental stress, and the expense of the tests will put more financial strain on the family [8]. The main driving force for our work is the importance of making an accurate prediction of PTB. These forecasts are essential to manage the logistical aspects of newborn care as well as for starting prompt interventions, particularly in areas with a shortage of advanced care facilities.

Steroids to encourage lung maturity and magnesium sulfate injections to safeguard the developing brain are essential treatments when access to skilled neonatal care is restricted. Accurately anticipating preterm birth also makes it easier to transfer patients either prenatally or postnatally to facilities that can offer critical care in situations where the required facilities are not available immediately, improving outcomes for both mothers and newborns [9]. Finding preterm birth risk factors is the main goal of the majority of current research on the prediction of PTB. Numerous risk factors have been demonstrated to increase the likelihood of preterm birth, namely body mass index (BMI), maternal family history, cervical length, vaginal hemorrhage, and depression [10], [11].

In recent years, a lot of research has focused on the use of Electronic Medical Records (EMRs) to predict premature birth. For instance [12] employed a stabilized sparse logistic regression approach to attain an AUC of 0.79 while concentrating on appropriate data preparation, addressing problems including data leaking and class imbalance. By merging racial-ethnic groupings [13] were able to predict PTB in nulliparous women utilizing data from almost 2 million patients, achieving an AUC of 0.67 [14] used a C5.0 Decision Tree to develop a model that achieves a sensitivity of 90.9% and a specificity of 71.8%, surpassing the prediction quality of costly and invasive fibronectin marker testing. By [15] utilized a variety of ML models to predict the delivery of extreme PTB (before 28 weeks), and they were able to get an AUC of 0.827 using an ensemble of LSTM models.

I.1 IDENTIFIED RESEARCH GAPS:

Although there is growing interest in using ML to predict preterm birth (PTB), existing models struggle to handle feature redundancy, integrate temporal clinical data, and achieve high sensitivity in the early stages of pregnancy. Although DL has been applied to EMR in a number of recent studies, many of these studies either ignore the temporal evolution of risk factors throughout prenatal Consultations or depend on static feature sets. For instance, Traditional time-series LSTM techniques by Zhang et al., (2022) [16], enhances prediction performance but frequently depend on insufficient gestational data, which restricts their use in late-trimester categorization. Similar issues with interpretability and dimensionality found by [17], they use ensemble methods to Electronic Health Record data, which affected their practical clinical applicability. On full multi-Consultation EMR data [18] showed that ensemble approaches can be adopt moderate accuracy, but they had trouble with feature interpretability and redundancy.

Crucially, the majority of current models do not dynamically include ultrasound-derived factors such as cervical length and plasticity index over gestational intervals. For extremely preterm and severe preterm forecasts, this omission results in decreased accuracy, especially in the later trimesters when intervention is essential. Furthermore, not much research has available in terms of hybrid evolutionary techniques (such DE or Dragonfly Optimization) in combination with entropy-based methods for novel and comprehensible feature selection in medical datasets. Additionally, there is discussion about an efficient feature fusion method for multi-Consultation data, which is crucial consideration for longitudinal PTB prediction. Thus, a significant research gap still exists in creating an end-to-end system for accurate and interpretable PTB risk prediction that uses publically available EMR datasets such as nuMoM2b and integrates DL classification, hybrid feature selection, and adaptive feature preprocessing.

I.2 OBJECTIVES

The objective of this research is to use longitudinal EMRs from publicly accessible nuMoM2b datasets to create a novel ML framework for early PTB prediction. The study specifically tried to address the limitations in existing research gaps by (i) integrating temporal clinical data from multiple prenatal Consultations, (ii) integrating ultrasound-derived parameters such as cervical length and pulsatility index to improve predictive sensitivity , and (iii) implementing a novel hybrid feature selection approach that combines entropy-based relevance ranking with evolutionary optimization techniques (e.g., Differential Evolution and Dragonfly Optimization) to manage feature redundancy and guarantee clinical interpretability. By using this method, the study hopes to facilitate clinically useful risk assessments that enhance the outcomes for mothers and newborns.

I.3 MAJOR CONTRIBUTION OF THIS PAPER:

In order to overcome a number of shortcomings in discussed methods, this study offers an innovative ML framework for the early PTB. The following are this study main contributions:

- Integration of Multi-Consultation Prenatal-EMR Data: This study integrates longitudinal data from three different prenatal intervals (6–13, 16–21, and 22–29 weeks), which captures the changing nature of risk variables during pregnancy.
- Methodology for Hybrid Feature Selection: The Dragonfly Optimization Algorithm and entropy-based relevance ranking are used to provide a novel hybrid feature selection approach. This approach dynamically provides a balance among redundancy and feature relevance which is lack from conventional filter or embedded selection methods.
- Ultrasound Feature Fusion: The model incorporates clinical data that are generally left out of many PTB prediction models, such as cervical length and pulsatility index, but are generated from ultrasound. The model sensitivity to identify the extremely preterm and very preterm babies, particularly in later trimesters, is increased by the inclusion of these data.
- Data Preprocessing using Evolutionary Optimization: To improve data quality and reduce the imbalance and noise in prenatal records, the system uses Differential Evolution (DE) for preprocessing and standardizing clinical variables from our previous study.

- Interpretable DL Model: To overcome the common shortcomings of DL models, a MLP classifier is trained using the optimized feature set to achieve excellent accuracy which guarantees the model transparency and clinical application.
- Validation on Public Dataset: The methodology is validated on the nuMoM2b dataset to ensure the reproducibility and encouraging the open research. The experimental results show better performance in terms of accuracy, AUC, sensitivity, and specificity when compared to baseline models.

II. THEORETICAL REFERENCE

II.1 LITERATURE REVIEW

The relevant literature on PTB prediction using ML and DL techniques is included in this section. By [19] utilized readily available variables from prenatal visits and machine learning to forecast premature birth in nulliparous women. Five-fold cross-validation was used to assess Elastic-net regularized logistic-regression models on data from 8,830 women at three prenatal visits who were part of the nuMoM2b dataset. Along with noteworthy gains in sensitivity and specificity, AUC values rose from 0.6161 in the first visit to 0.7087 in the third. The predictive capacity of the models was significantly improved by the additional measurements. Interestingly, at the third prenatal visit, the model's sensitivity for predicting very-preterm and extreme-preterm infants was 0.8254 and 0.9295, respectively. By [20] aims to create a Risk Prediction Conceptual Model (RPCM), a ML model, for PTB prediction. This work proposes a feature selection method based on the concept of entropy. The goal of the unique strategy is to forecast the classifier's accuracy at the greatest level by identifying the best maternal features that causes PTB from the obstetrical dataset. We then gather obstetrical data from rural communities' Community Health Centers (Kamdara, Jharkhand).

To categorize the considered data into term birth and PTB, the recommended method is utilized to the data gathered to determine the great maternal features of pregnant women. At the end, SVM classifier secured improved performance of 90.9% accuracy. By [21] predicts neonatal mortality and neonatal morbidities such as necrotizing enterocolitis, retinopathy of prematurity, and bronchopulmonary dysplasia in very low birth weight infants using machine learning. Time series data and clinical characteristics gathered at Helsinki University Hospital's Children's Hospital's neonatal ICUS as our predictors. We look at nine distinct classifiers. In order to increase classifier sensitivity, our systematic analysis also includes various data pretreatment techniques. When it comes to predict the classes, the analyzed classifier AUROC is 0.922, 0.899, and 0.846, respectively. According to [22] suggests a ML-based PTB prediction model. All of the patients in this study were chosen from a Chinese hospital, and the data evaluated ranged from 2008 to 2018. 4775 of the 9550 pregnant women who were enrolled in the research had PTB. 4775 individuals in all were chosen at random to serve as controls. The accuracy (0.816) and AUC (0.885) of the RF model based on 27 weeks of gestation were the highest when compared to other approaches.

By [23] create, train, and evaluate ML models for singleton pregnancies at various gestational intervals that predict preterm birth (less than 37 weeks gestation). The models were developed using complete data using 22,603 singleton pregnancies from prospective based on populations cohort research conducted in 51 hospitals through midwifery clinics in Wenzhou City, China, during 2014 and 2016. To build prediction models, the top 10%, 25%, and 50% of predicting features were selected. Employing the training data set with five-fold cross-validation, prediction models were developed for internal validation. AUC results were utilized to assess the model performance. After 26 weeks of gestation, the CatBoost-based prediction model outperformed the others with an accuracy of 0.81 and an AUC value of 0.70 (0.67, 0.73). According to [24] intends to use a reduced PSG-based system that solely uses airflow data to find sleep apnea in pregnant nulliparous women. For sleep-disordered breathing (SDB) studies, we used the nuMOM2b dataset, which included 3012 female participants. It is based on computer-aided diagnostic tool to detect sleep apnea in expectant mothers automatically. Wavelet coefficients are used to extract the L1- norm features.

To automatically identify apnea, the chosen features are subsequently passed into a variety of ML classifiers. Using the RUS-Boosted tree classifier, the created model achieved the best accuracy of 83.9% with the F1 score of 0.91. In turn [25] utilized ML techniques to find the main predictors of PTB among clinical and dental factors. We collected cohort data from 60 women who had cesarean sections for singleton infants (30 PTB, 30 FTB). PTB and spontaneous PTB (SPTB) were the dependent variables. For the ML analysis, 15 independent variables which includes10 clinical and 5 dental factors were chosen. The key predictors of PTB and SPTB were found using the Random Forest (RF) variable significance. The relationships among the predictors and PTB/SPTB were examined using Shapley additive explanation (SHAP) values. Pregnancy BMI, modified gingival index (MGI), preeclampsia and two more factors were the top five that were identified as major factors to predict PTB based on RF variable relevance. Mother age, Premature rupture of the membranes and pre-pregnancy BMI were among the key predictors of PTB/SPTB that showed positive relationships with SHAP levels.

The necessity of integrated medical and dental care during pregnancy is highlighted by the positive correlations shown among these predictors and PTB. By [26] suggests a two-phase approach in which Phase I will construct a graphical user interface (GUI) based on ML for the early PTB prediction and will choose the key factors that cause PTB based on the opinions of experts and a review of the literature. Using paper based data collected from a Jharkhand hospital, 15 sociodemographic factors are extracted from digitized birth dataset (DD-I). To determine their connection with PTB, parameters are statistically examined. With a mean accuracy of 88.02% (with SD.: 0.001), the RF learner is then examined as the superior for early prediction, outperforming traditional manual methods.

These studies demonstrate the increasing interest in predicting preterm births using EMRs. There is still a sizable gap in the literature, though, as many of the discussed models either don't assess the model performance at various pregnancy stages. By concentrating on readily available characteristics that can be gathered during routine prenatal visits and evaluating the model's prediction power at various pregnancy stages, our study aims to close this gap.

III. MATERIALS AND METHODS

Write in detail the research project, including background and limitations. The selection of materials and methods, procedures and equipment must be justified so that the work can be reproduced. Modifications or new methods must be described in detail. You must clearly define the universe and specify how the sample was selected and why it is representative.

Data processing represents the practical development of a theoretical basis, deriving the model equations to program the calculation algorithm, according to the need. In materials, they include the technical specifications and the quantities, the origin and, if necessary, the method for its elaboration . The proposed architecture in Figure 1 presents a robust, ML pipeline designed for accurate PTB using EMR. The process comprises of five phases including data preprocessing, data structuring (Temporal), feature selection, classification and model assessment. Data Preprocessing: Initially the raw clinical data is preprocessed to remove the records with missing or irrelevant features, no indication of birth type as PTB or full term, or unrelated risk factors. Differential Evolution (DE) based preprocess on medical PTB data from our previous research is used to optimize this process for data cleaning. The cleaned dataset is divided into training and testing subsets of 70:30 stratified sampling. To class imbalance is addressed by SMOTE (Synthetic Minority Oversampling Technique) method.

Temporal Data structure: The training data is further divided longitudinal subsets based on prenatal Consultations intervals as baseline, Consult1 (6 to 13 weeks), Consult2 (16 to 21 weeks) and Consult3 (22 to 29 weeks). This segmentation helps model for risk factors evolution among the gestation. Feature selection using H-EDFO: This hybrid approach combines entropy-based feature relevance score with the DFO. It ranks the features based on their clinical relevance and removes the redundant or less informative features by improving the interpretability and performance. Classification using MLP: The optimized feature subset is fed as input to MLP that classify the outcomes as PTB and non-PTB. Model evaluation and Interpretation: The final phase states include discrimination and calibration analysis to evaluate the model performance and reliability using the metrics accuracy, sensitivity and AUC across various gestational stages.

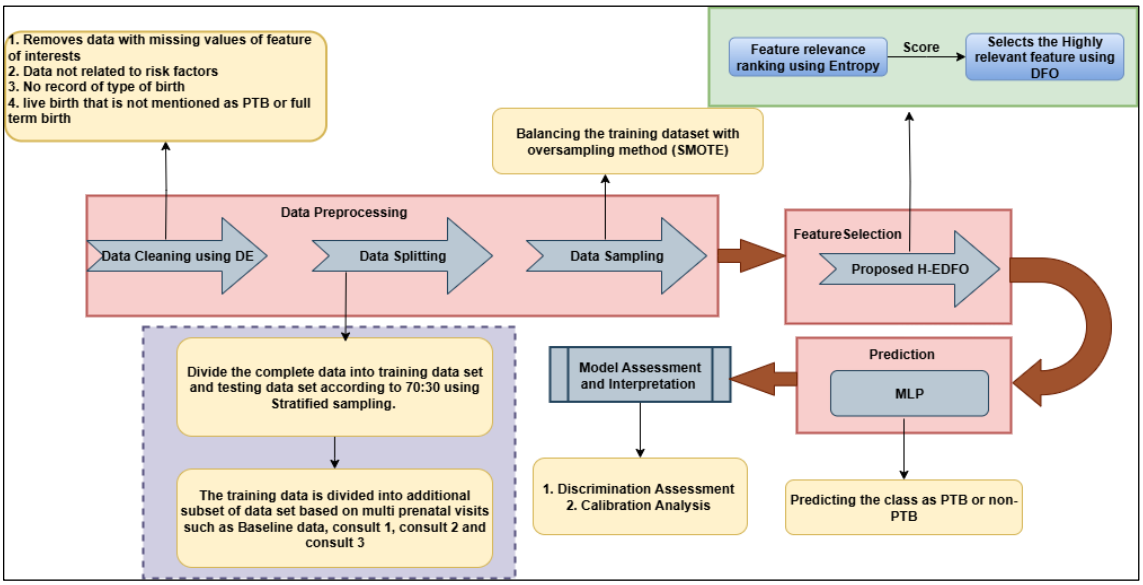


Figure 1: Proposed architecture of PTB prediction using longitudinal EMR data from nuMom2b dataset.
Source: Authors, (2026).

III.1 DATA SET DESCRIPTION AND SELECTION

One useful tool in our research is the Nulliparous Pregnancy Outcomes Study: New Mothers-to-Be (nuMoM2b) dataset made publicly available by the NICHD DASH repository (2015), served as the dataset for this investigation. Its usage in this study didn't require any further ethics clearances or data use agreements [27] which is an extensive collection 10038 nulliparous women with singleton pregnancies [27] who were sourced from eight clinical facilities connected to research institutes are included in the dataset respectively, Western Reserve University (Site 1), Columbia University (Site 2),Indiana University (Site 3), Magee-Women's Hospital (Site 4), Northwestern University (Site 5), University of California Irvine (Site 6), University of Pennsylvania (Site 7), and the University of Utah (Site 8) [19].

The nuMoM2b study's main goals were to: (1) identify maternal traits that influence or predict an adverse pregnancy outcome, such as genetics, epigenetics, and the body's response to environmental influences and pregnancy (2) identify specific elements of placental growth and function that lead to a poor pregnancy outcome; and (3) outline the genetic, growth, and developmental characteristics of the fetus that are connected to a poor pregnancy outcome.

The requirement of study inclusion and exclusions are listed in Table 1:

Table 1: Considered inclusion and exclusion data requirements of prenatal Consultation.

Inclusion	Exclusion
Pregnant women who were nulliparous (defined as having no previous pregnancies longer than 20 weeks), (2) identify specific elements of placental growth and function that lead to a poor pregnancy outcome; and (3) outline the genetic, growth, and developmental characteristics of the fetus that are connected to a poor pregnancy outcome.	1) Maternal age <13 years, 2) history of at least three spontaneous abortions, (3) identified fetal aneuploidy, (4) a surrogate pregnancy, (6) multifetal pregnancy reduction, (7) involvement in an intervention study that could impact maternal or fetal morbidities/mortality (8) previous enrollment in the nuMoM2b study, (9) planned pregnancy termination, and (10) inability to give informed consent

Source: Authors, (2026).

Based on the dataset description, collection of data is based on interviews, clinical measurements, ultrasounds, self-administrated questionaries and medical records review at the scheduled Consultations as stated in Table 2. To guarantee data consistency and integrity across trial sites, particular data collection protocols were adhered to at every Consultation. Standardized training on data collection procedures was provided to all study staff, and frequent monitoring Consultations were carried out to make sure study protocols were being followed. Some inaccuracies might have remained, and some data might still be outside of expected ranges or otherwise illogical in spite of these data quality control procedures. Before performing analysis and deciding how to handle inaccurate data, investigators utilizing the data must assess fields of interest.

Table 2: Gestational time periods of prenatal Consultations of the dataset.

Consultation	Gestation weeks	Predictors
Consult 1	6 to 13 weeks	35
Consult 2	16 to 21 weeks	51
Consult 3	22 to 29 weeks	73
Consult 4	Delivery time	-

Source: Authors, (2026).

Of the 10,038 individuals, 9,289 agreed to share their anonymised data for research purposes, and 9,127 volunteered to share their baby's data. Numerous variables were covered by the information gathered, such as physiological measurements, food habits, psychosocial factors, demographics, and pregnancy outcomes. The [28] publication contains a detailed explanation of the procedures used in the Nulliparous Pregnancy Outcomes Study. To ensure model stability and dependability, logistic regression-based prediction models usually recommend a minimum of 10 events per predictor variable (EPV)[29]. 7,59 preterm births were among the 8,830 individuals in our final analytical sample. With each Consultation, the number of predictor factors are listed in Table 2. All models demonstrated suffic29ient statistical power by exceeding or meeting the minimum suggested criterion of 10 EPV. Furthermore, the minimum of 2,000 participants advised through reviews for reliable and broadly applicable ML applications in healthcare is significantly exceeded by our sample size of 8,830 participants [30].

Dataset Variables:

A methodical, multi-step procedure was used to choose the candidate predictor variables. We found established predictors of preterm birth through a thorough assessment of the literature [31-33], including clinical variables, specialized imaging data, and biological markers. Some known predictors have the potential to have a high predictive value, but they were not included in database because of the resource-intensive factor. Clinically accessible predictors that were easily obtainable throughout standard prenatal care were the main focus of our final variable selection process. A number of criteria were used to evaluate the variables, including availability in the routine prenatal Consultations, cost-effectiveness, and proven clinical relevance from published research. As shown in Table 3, the selected factors included demographic data, non-invasive evaluations, medical history, family medical history, and standard diagnostic procedures. The selection of these variables was informed by cost-effectiveness and feasibility concerns, as they can be gathered quickly, cheaply, and without requiring invasive or resource-intensive procedures. As a result, ordinary clinical practice can easily implement the prediction models [34].

Table 3: nuMom2b dataset identified characteristics with variables.

Characteristics	Variables
Basic Preconception	maternal age, height, BM), adoption of assisted reproductive technologies, PTB family history, pregnancy complications, hypertension, and cervical excisional procedures/surgery history (Loop Electrosurgical Excision Procedure (LEEP)/conization),
Lifestyle	Mental stress, alcohol consumption, tobacco usage, and maternal vitamin-D deficiency
Sociodemographic	ethnicity, marital status, educational attainment, and socioeconomic status
Obstetric/Pregnancy	Vaginal hemorrhage, serial transvaginal cervical length, gestational diabetes and hypertension, and cervical insufficiency
Medical History	history of obstetrics, diabetes, and hypertension prior to pregnancy

Source: [19].

III.2 DATA PRE-PROCESSING

III.2.1 Data Cleaning Using DE for Analyzing Missing Data

A comprehensive analysis of missing data patterns was carried out, which included a review of the dataset documentation and a close look at the data gathering methods. Even while many variables had incomplete data patterns lacking a clear documentation explanation, several variables displayed systematic missingness that was related to specific therapeutic regimens.

The research by [35] utilizes the DE based preprocessing to handle the missing values of medical datasets. In this study used the same model to handle missing and irrelevant data of nuMoM2b dataset. DE is the population based approach that searches globally to find the optimal imputation values. The steps for DE based imputation is shown as follows:

Step 1: Preprocessing

- The dataset is divided into continuous and categorical values
- Normalize the continuous features values in the range [0, 1]
- Categorical values are encoded using one-hot encoding

Step 2: Fitness function: The candidate imputation is evaluated using clinical constraints such as cervical length in the range 15 to 50mm.

Step 3: DE optimization

- Population is initialized for the candidate imputation
- DE operations such as mutation, crossover and selection are applied as:

Step 4: Post processing: Reverse Normalization and selects the categorical variables with the variables having highest probability of DE search. The value of mutation factor is declared as 0.6 and number of population N is 50 for 100 iterations with crossover value of 0.8.

Nonetheless, certain variables were missing for systematic reasons that could be explained. Missing entries were taken to mean that there was no tobacco use or bleeding beyond spotting for variables like "Usage of Tobacco in Consult-1" and "Bleeding in Vaginal" at all three Consultations, respectively, and were imputed appropriately. Only in the event that the results of the uterine artery Doppler at Consult-2 were abnormal and uterine artery Doppler ultrasonography at Consult-3 is required during data collection. This procedure is in line with a study by Hofstaetter et al., that found that the uterine artery Pulsatility Index (PI) did not change significantly with gestational age in uncomplicated pregnancies [36].

Therefore, we randomly selected from the normal PI (Pulsatility Index) range appropriate among 22 and 29 weeks, as reported in the study by [37], to impute these missing values for cases with missing data of "Left uterine artery -PI at Consult-3" and "Right uterine artery - PI at Consult-3" where Consult-2 measurements were normal. This strategy was required to preserve the dataset's integrity while taking into account the clinical judgment that guides these assessments. Before variable extraction for every Consultation, imputation was done once for every training dataset. Based on this analysis, the records inclusion and exclusion are stated in the Figure 2.

There were 749 of the 10,038 women who were first enrolled in the nuMoM2b research were disqualified because their risk factor information was lacking. There were 438 of the 9,289 women who still had data available were disqualified (432 had no birth type record, 4 had live births without records, and 2 had stillbirths without records). There were 21 more participants were disqualified because their data was missing more than 80% of the time. These cases mostly involved very early preterm deliveries or early pregnancy loss, for which data from follow-up Consultations could not be obtained. A final analysis sample of 8,830 women with adequate data through delivery of 8,071 with full-term births and 759 with preterm deliveries was obtained through this systematic filtering approach.

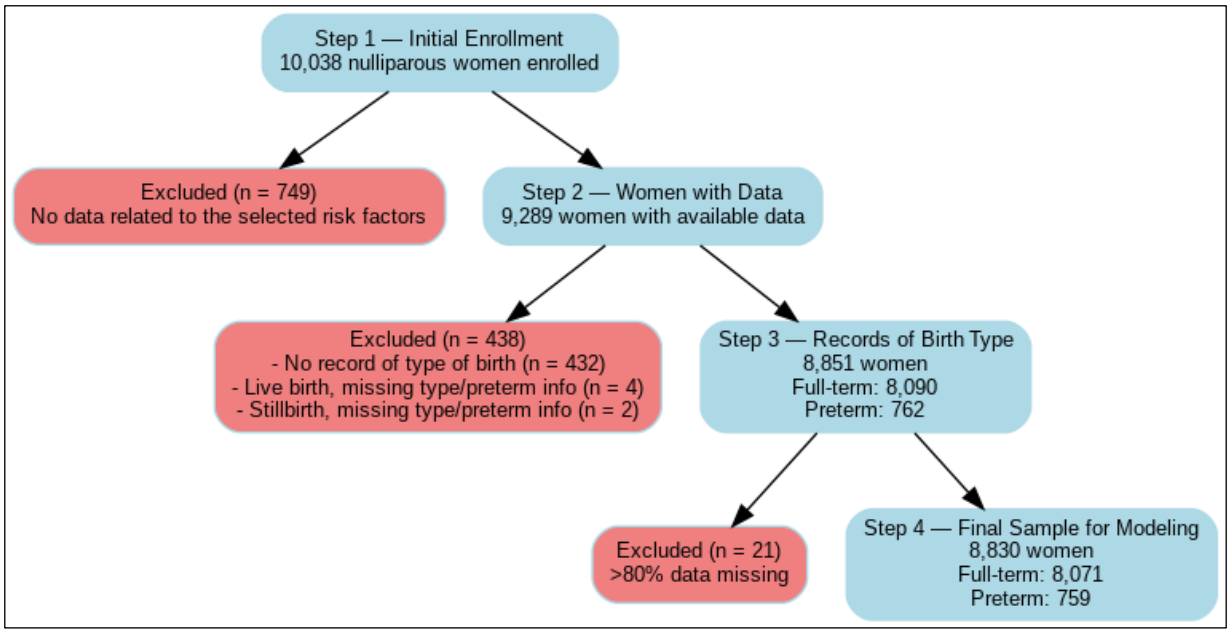


Figure 2: Records inclusion and exclusion based on preprocessing.
Source: Authors, (2026).

III.2.2 Temporal Data Splitting

A total of 35 variables were accessible during the initial Consultation, 51 during the second Consultation, and 73 during the third Consultation. Basic data, such as ethnicity, level of poverty and obesity, pre-gestational diabetes, marital status, and educational attainment, were gathered during the initial Consultation (dataset B).

Furthermore, we have fundamental measurement factors (such as surveys and questionnaires) that are measured at all three Consultations, including BMI, smoking, vaginal bleeding, level of stress, gestational diabetes and hypertension diagnosis (dataset C1, C2, and C3). Following that, ultrasound measurements were added to the data at the second and third Consultations (datasets U2 and U3). The overview of dataset is described in Table 4.

Table 4: The overview of dataset.

Datasets	Variables
B	poverty, color, marital status, education, obesity level, and prenatal diabetes diagnosis
C1	Basic assessments during consultation 1 include BMI, level of stress, smoking, gestational diabetes, vaginal bleeding and hypertension.
C2	fundamental measurements Weight, body mass index, smoking status, vaginal hemorrhage, gestational diabetes, and hypertension during the second consultation
C3	Weight, BMI, smoking history, vaginal-bleeding, gestational diabetes, and hypertension are the fundamental parameters taken at consultation3.
U2	Pulsatility Index and cervical length ultrasound measures at consultation2
U3	Pulsatility Index and cervical length ultrasound measures at consultation3

Source: Authors, (2026).

III 2.3 Data Sampling

Given that preterm births and other unfavorable pregnancy outcomes affect a smaller percentage of the population than term deliveries, the nuMoM2b) dataset is naturally unbalanced. Machine learning classifiers may become biased toward the majority class as a result of this imbalance, making it harder for them to identify high-risk situations. The Synthetic Minority Oversampling Technique (SMOTE) [38] is used as a data-level balancing technique prior to model training in order to overcome this. By interpolating between a data point and its closest neighbors in feature space, SMOTE [38] creates synthetic samples of the minority class, lowering the possibility of overfitting that can happen with straightforward minority sample duplication.

This preserves specificity while increasing the classifier's sensitivity to uncommon occurrences, like extremely premature birth. The method improves the representation of minority patterns in the training set while maintaining the original data distribution, and it has been demonstrated to be successful in medical prediction tasks, especially when paired with feature selection and ensemble classifiers. After preprocessing the nuMoM2b dataset, SMOTE was applied in this study to produce a balanced input set for predictive modeling, preventing majority-class dominance from overshadowing early markers of preterm birth.

III.3 FEATURE SELECTION USING PROPOSED H-EDFO

In ML, choosing a subset of exceptional factors during the creation of the prediction model is referred to as feature selection, which is also referred as feature- subset selection. The existence of duplicate and unnecessary features in any dataset, but particularly in medical datasets, can lower prediction accuracy and negatively affect the model's performance. To cut down on training time and improve the classifier's predictive accuracy, the primary objective of any feature selection technique is to choose the optimal feature subset by eliminating superfluous and irrelevant characteristics from the datasets. The variables selections are based on systematic process.

Based on the literatures, the identified PTB predictors includes clinical variables, biological markers, and specific imaging findings as shown in Table 3. Clinically useful predictions that could be obtained and were included in the database using the suggested H-EDFO approach were given priority in the final variable selection. The proposed conceptual framework for H-EDFO is show in Figure 3. To improve predictive modeling in medical datasets, the suggested approach uses a feature selection and fusion strategy based on multi-subset entropy and Dragonfly Optimization (DFO). The preprocessed nuMoM2b dataset is first divided into several subsets that correspond to various stages of consultation and clinical classifications.

Entropy-based feature ranking is applied to each subset in order to measure the discriminative ability of specific features. In order to optimize the selection process and determine the most pertinent properties from each subset, the ranking features are further put through DFO-based feature subset selection, which simulates the static and dynamic swarming behaviors of dragonflies (Meraihi, 2020) [39]. A union procedure is then utilized to fuse the chosen features from each subset, creating an ideal feature set that removes redundancy and maintains complementary information. With increased accuracy and computing efficiency, this fused set can be applied to downstream tasks like illness risk classification or prediction.

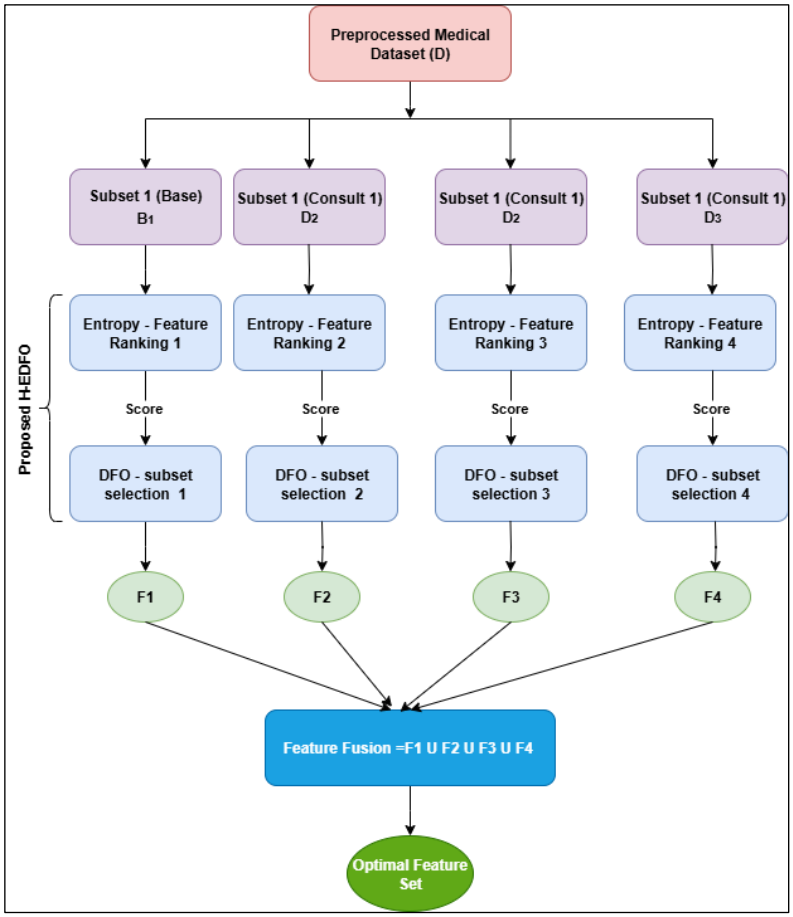


Figure 3: Proposed FS conceptual framework.
Source: Authors, (2026).

Entropy notion

Strongly correlated attributes cannot be included in a feature subset, per the study in [40]. Additionally, the more independent the attributes are from one another, the more information they will have, which will ultimately produce better results than data that is not visible. The feature selection strategy works better for medical datasets, which are the subject of current study since they are more sensitive. To extract most relevant features from the PTB dataset, an entropy-based feature selection strategy is described by this section. These characteristics are used to categorize all birth cases into PTB and term birth. The entropy-based feature ranking is shown in Algorithm 1.

Algorithm 1: Entropy based Feature Ranking

Step 1: Assume the medical dataset is D having N attributes such that $\mathcal{F}_i$ $where$ $i = 1,2, \ldots N$. Let A_0 is the group of features of original dataset.

Initially, $\mathcal{F}_0 = \{F_1, F_2, F_3, \ldots F_4\}$. Since D is divided into four subsets such as B, C1, C2 and C3 and after the feature selection we will get four feature subsets such as fs1, fs2, fs3 and fs4 respectively. F is the fused subset derived from the subsets and Initialize $fs_k = fs_0$ for $k = 1,2,3,4$.

Assume P is the problem for classification defined from set of N attributes called F_i $where$ $i = 1,2,3, \ldots, N$ and consider F denotes the features resultant from actual dataset.

Step 2: Initialize $F = fs_0 = \{\mathcal{F}_1, \mathcal{F}_2, \mathcal{F}_3, \ldots, \mathcal{F}_N\}$

For every subset $C_i \in D$ where i=1,2,3,4 do

For every attribute $\mathcal{F}_i \in fs_o$ do

Compute Gain (S, $\mathcal{F}_i$) using Eqn (1)

$$Gain(S, \mathcal{F}_i) = Entropy(S) - \sum_{v_j \in \mathcal{F}_i} \left({\left|S_{v_j}\right|}/{|S|} \right) \quad (1)$$

$$Entropy(S) = \sum P_m log_2 P_m \quad (2)$$

Where v_j denotes of attribute $\mathcal{F}$ values and S is amount of samples of P and non-zero probability s$\in S$ is denoted as P_m of class 'c' among 'm'.

End-for

calculate

$$R = (\text{MAX } \_gain(S, \mathcal{F}_i) - \min \_gain(S, \mathcal{F}_i)/n) \quad (3)$$

Where i=1,…n and R is the threshold of the chosen variables.

For every attribute Compute$\mathcal{F}_i \in f_0$ do

If $Gain(\mathrm{S}, \mathcal{F}_i) < R$ then

Update $fs_k = fs_k = \{\mathcal{F}_i\}$ removes $\mathcal{F}_i$ from fs_k

End-if

End-For

End-For

Pass theses selected attributes scores to DFO for optimized feature selection. The sample feature name with entropy score is listed in Table 5. Based on the entropy score values, the features are ranked and ordered its importance. These scores are given as input to the DFO as fitness function for optimal feature selection.

Table 5: Feature rank and entropy score of each feature.

Dataset	Feature Name	Entropy Score (0–1)	Rank
B (Baseline)	Maternal Age	1.000	1
	BMI at Enrollment	0.942	2
	Race/Ethnicity	0.912	3
	Gravidity	0.871	4
	Smoking Status	0.836	5
	Education Level	0.801	6
	Previous Preterm Birth	0.764	7
	Household Income	0.725	8
	Marital Status	0.691	9
	Health Insurance Type	0.654	10
C1 (First Trimester Clinical)	Systolic BP	1.000	1
	Diastolic BP	0.956	2
	Fundal Height	0.928	3
	Fetal Heart Rate	0.899	4
	Hemoglobin Level	0.866	5
	Cervical Length	0.825	6
	Uterine Artery PI	0.791	7
	Weight Gain (trimester)	0.756	8
	Urine Protein	0.724	9
	Blood Glucose Level	0.688	10
C2 (Second Trimester Clinical)	Cervical Length (mid)	1.000	1
	Amniotic Fluid Index	0.952	2
	Placental Position	0.917	3
	Fetal Biparietal Diameter	0.881	4
	Abdominal Circumference	0.842	5
	Femur Length	0.805	6
	Uterine Artery Notching	0.768	7
	Weight Gain (mid)	0.734	8
	Hematocrit Level	0.703	9
	Glucose Challenge Test	0.667	10
C3 (Third Trimester Clinical)	Estimated Fetal Weight	1.000	1
	Amniotic Fluid Index (late)	0.948	2
	Cervical Dilation	0.913	3
	Fetal Head Circumference	0.876	4
	Abdominal Circumference (late)	0.844	5
	Femur Length (late)	0.809	6
	Blood Pressure (late)	0.772	7
	Proteinuria (late)	0.738	8
	NST Result	0.705	9
	Weight Gain (late)	0.668	10
U1 (Ultrasound 1)	Fetal Heart Rate (US)	1.000	1
	Crown–Rump Length	0.946	2
U2 (Ultrasound 2)	Femur Length (US)	1.000	1
	Abdominal Circumference (US)	0.935	2

Source: Authors, (2026).

III.3.1 DFO

In order to migrate and hunt, dragon fly optimizers (DO) [41] imitate the behavior of dragonflies. It could be dynamic or static. The minimum number of dragonflies (DF) in the static swarm is relocated to hunt other swarms nearby. The greater number of dragonflies in the dynamic swarm travel far together in a similar direction. Five weights such as food, opponent, cohesiveness, separation, and alignment along with inertia weights determine how DF moves. The two main features of DF in SI are exploration and exploitation. Low alignment with greater weight cohesion characterizes the exploitation search space, while high alignment with decreased weight cohesion characterizes the exploration search space. These weights are adjusted during the optimization procedure. The characteristics of DFO are shown in Figure 4 and its expressions are listed in Table 6.

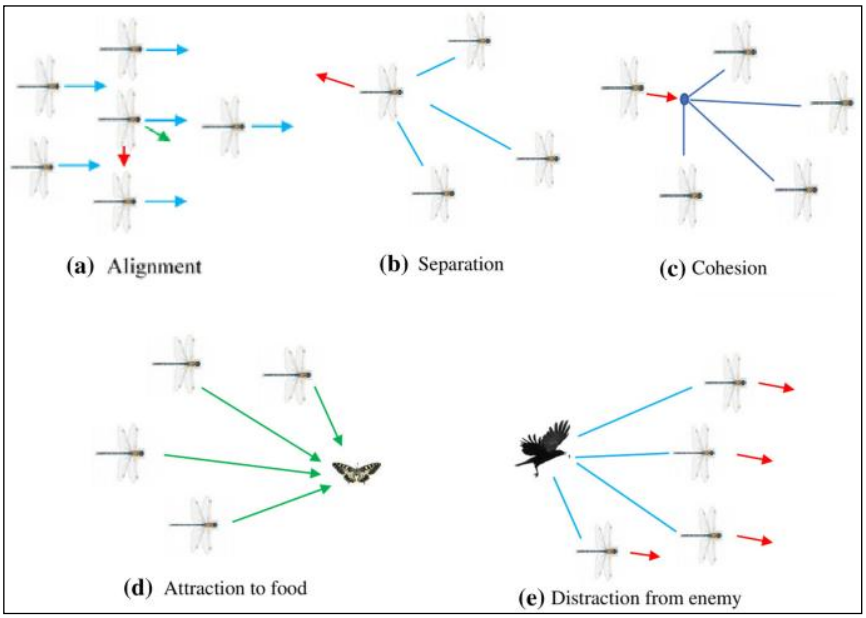


Figure 4: DFO characteristics.
Source: [42].

Table 6: Description of DFO characteristics.

Characteristics	Description	Expression
Separation	As indicated by Eqn (4), separation is the individual's avoidance of their neighbor.	$S_i = -\sum_{j=1}^{N} X - X_j$ (4) Where, X – current position, X_j is the jth neighbor position, N is the amount of neighbors and S is the separation motion of individual i.
Alignment	Eqn (5) is used to calculate the matching velocity between an individual and its neighbor.	$A_i = \frac{\sum_{j=1}^{N} V_j}{N}$ (5) A is the alignment motion, V is the velocity of neighbor j.
Cohesion	Eqn (6) is the measurement of the person with respect to the neighborhood center.	$C_i = \frac{\sum_{j=1}^{N} X_j}{N} - X$ (6) Where, C is the ith individual cohesion
Food	This is the motion of DF in the direction of its food attraction, as indicated by Eqn (7).	$F_i = X^+ - X$ (7) Where, F is the ith individual food attraction, X^+ is the food source
Enemy	This serves as DF's diversion from its adversaries and is indicated in Eqn (8)	$E_i = X^- + X$ (8) Where, E is the individual enemy distraction motion, X^- is the enemy position.

Source: Authors, (2026).

The step and position vector are used to update the individual DF position. Eqn (9) computes the PSO [42] velocity vector, which is comparable to the step factor Δx.

$$\Delta X_{t+1} = (S_i + A_i + C_i + F_i + E_i) + \omega \Delta X_t \quad (9)$$

Where t is the number of iterations and ω is the inertia weight. This inertia weight is added to the regular DO to create the best possible outcome. Eqn (10) improves the step vector's inertia weight.

$$\omega = \omega_{max} - \frac{\omega_{max} - \omega_{min}}{t_{max}} . t \quad (10)$$

Where t_max is the maximum iteration, t indicates the number of iterations, and ω_max and ω_min declare the dragonfly's beginning and ending values. The weights and the number of iterations are inversely correlated. The weight will drop as the number of iterations increases, making the global search capability stronger. Eqn (11) is used to update the location vector.

$$X_{t+1} = X_t + \Delta X_{t+1} \quad (11)$$

III.3.2 Hybrid E-DFO-Based Optimal Feature Selection and Feature Fusion

The entropy score of each features are treated as the fitness function of DFO for optimal selection using Algorithm 2.

Algorithm 2: Hybrid E-DFO based feature selection and fusion

Step 1: Input preprocessing datasets and Entropy score of each features.

Step 2: DFO encoding

Each DF position vector is a binary string such that 1 means feature selected and 0 means feature eliminated. For ex: [1,0,0,1,0] means feature 1 and 4 are selected.

Step 3: Initial population

- Instead of random population, the top entropy features are given as population
  For Example, dataset B – the top 5 ranked features are having the probability P~0.8 and low ranked features are having the probability P~0.2.

Step 4: Fitness function

The multi objective fitness of DFO is combined as:

$$F = \alpha Accuracy + \beta Entropy_{Score} - \gamma Feature\_count\_penalty \quad (12)$$

Where, Accuracy is the model accuracy of the selected featuers, Entropy_Score is the selected features entropy mean, Feature_count_penalty is penalizing much features selection and α, β and γ are the weights controlling parameters.

Step 5: DF searching behavior:

Separation, Cohesion and Alignment are adjusted to select bitstrings, food attraction moves towards increased fitness solutions and distraction from enemies avoid poor solutions.

Step 6: Feature fusion: The find optimal features from all the datasets are fused to find the optimal feature set as in Eqn (13) where F1, F2, F3 and F4 are the feature sets from the considered four data subsets.

$$F = F1 \cup F2 \cup F3 \cup F4 \quad (13)$$

The extracted features and optimal feature subsets are discussed in result section.

III.3.3 Classification Using MLP

To forecast the models with several learning procedures and minimal limitations, this work used a Python tool to create a feed forward back propagation algorithm [43], one of the Multi Layer Perceptron (MLP) functions. A typical MLP consists of one input, one or more hidden layers, and one output layer. Most studies handle the transition using a single hidden layer as shown Figure 5. The pixel and transit model samples from the determination classes were used in the training procedure.

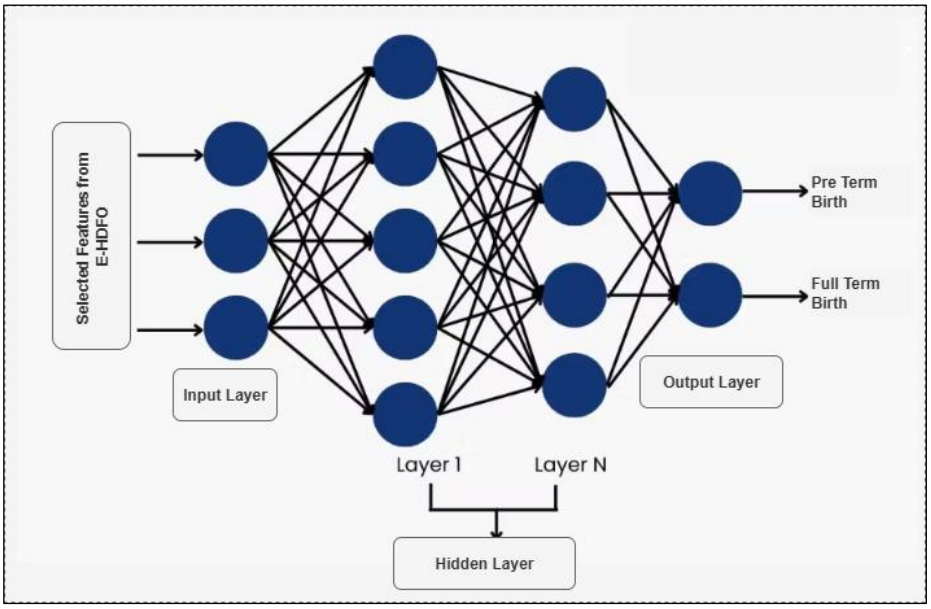


Figure 5: MLP structure for PTB prediction.
Source: Authors, (2026).

The training model, which looks at the beginning and ending learning rates, is where the MLP in the categorization begins. The hidden layer node is used to run the submodel, and all parameters—aside from the number of hidden layers—are set to their default settings. Based on several trials, the value is doubled if the overall accuracy is raised; if not, the final value is taken into account. The process of predicting PTB using MLP is shown in Algorithm 3.

Algorithm 3: MLP based PTB prediction

Step 1: Let X be the input with m features and Y is the output belongs to two classes such as 0 means non PTB and 1 means PTB

Step 2: Compute Entropy based feature ranking using Algorithm 1

Step 3: The entropy score is treated as the fitness function for H-EDFO and select optimal feature subset and fusion using Algorithm 2

Step 4: MLP based PTB classification

Let $X \in R^d$ be vector of fused features and $d = |F\ (fused\ feature\ set|$
An MLP among with L layers transforms X as,

$$h^{(1)} = \sigma(w^1 x + b^1) \tag{14}$$

$$h^{(l)} = \sigma(w^l h^{l-1} + b^1)\ where\ l = 1,2,3, \ldots L-1 \tag{15}$$

$$y' = \phi(w^L h^{L-1} + b^L) \tag{16}$$

Where, σ is the ReLu activation function and ϕ is the sigmoid activation function for output and y' is the predicted output as 0 or 1. The loss function is computed using Binary cross entropy loss as

$$Loss = -\frac{1}{N}\sum_{i=1}^{N}[Y_i \log(Y'_i) + (1 - Y_i)\log(1 - Y'_i)] \tag{17}$$

Steo 5: Prediction

A threshold T of 0.5 is used on while converting y' final decision

$$class = \begin{cases} 1 & Y' \geq T \quad (PTB) \\ 0 & otherwise \quad (non - PTB) \end{cases} \tag{18}$$

The hyper parameters of proposed PTB prediction system are shown in Table 7.

Table 7: Hyper parameters of H-EDFO-MLP based PTB prediction.

Approaches	Hyperparameter	Values	Description
Entropy-based Feature Scoring	Entropy	-	Shannon entropy used for each feature.
	Score Normalization	Min–Max scaling [0, 1]	Scales entropy values for uniform weighting in DFO.
Dragonfly Optimization (DFO)	Population size	30	Number of dragonflies (candidate solutions).
	Iterations	100	Maximum optimization cycles.
	Inertia weight (w)	0.9 → 0.4 (linearly decreasing)	Balances exploration vs. exploitation.
	Alignment coefficient (A)	1.0	Controls alignment behavior in swarm.
	Cohesion coefficient (C)	1.0	Controls attraction to swarm center.
	Separation coefficient (S)	1.5	Avoids overcrowding in search space.
	Food factor (F)	2.0	Attracts toward optimal solutions.
	Enemy factor (E)	-2.0	Repels from poor solutions.
	Selection threshold	0.5	Keeps features with probability > threshold.
MLP Classifier	Input layer size	Number of fused features	Matches output of DFO stage.
	Hidden layers	2	Depth for learning feature interactions.
	Neurons per layer	[64, 32]	First layer: 64, Second layer: 32 neurons.
	Activation function (hidden)	ReLU	Non-linear transformation.
	Activation function (output)	Sigmoid	Sigmoid for binary PTB
	Optimizer	Adam	Adaptive learning rate optimization.
	Learning rate	0.001	Controls step size in weight update.
	Batch size	32	Samples per gradient update.
	Epochs	100	Full passes over dataset.
	Loss function	Binary cross-entropy	Matches binary classification of PTB.
	Early stopping patience	10 epochs	Stops training if no improvement.
	Dropout rate	0.3	Regularization to prevent overfitting.

Source: Authors, (2026).

IV. RESULTS AND DISCUSSIONS

The WEKA toolbox (http://www.cs.waikato.ac.nz/ml/weka) or Python 3.12 and the Scikit-Learn package are used to conduct the experiment. The birth dataset's observations are thoroughly examined in order to predict birth cases. This is actually a binary class dataset, where all births that take place between weeks 28 and 37 are referred to as PTB classes with the label "1" while all births that take place beyond weeks 37 are referred to as Full Term Birth (FTB) classes with the label "0." As per our preprocessing and data splitting approaches, there were 8830 women with 8071 as full-term birth and 759 PTB are selected for analyzing the performance of the proposed model. Using AUC ratings obtained from 5-fold cross-validation comprising information from several prenatal consults, we assessed the predictive ability of the models. A thorough examination of pooled performance indicators such as accuracy, sensitivity, specificity, and AUC scores, came after this evaluation. The performance metrics are listed in Table 8.

Table 8: Performance metrics for classifiers.

Metrics	Expressions
Accuracy	$\frac{TP+TN}{TP+TN+FP+FN}$
Precision	$\frac{TP}{TP+FP}$
Sensitivity / Recall	$\frac{TP}{TP+FN}$
F1 score	$2*\frac{Precision*Recall}{Precision+Recall}$
AUC	$\int_0^1 TPR\left(FPR^{-1}(x)\right)dx$ $TPR=\frac{TP}{TP+FN}$, $FPR=\frac{FP}{FP+TN}$

Source: Authors, (2026).

IV.1 IMPACT ON PREPROCESSING MODEL

Table 9 shows the performance of the original dataset outcome with balanced and preprocessed dataset using DE model. The PTB prediction model's performance on the original dataset and the dataset preprocessed using Differential Evolution (DE) is contrasted in the preprocessing table you supplied. The DE-based preprocessing notably improves outcomes across all measures, such as accuracy, precision, recall, F1-score, and AUC. Accuracy increased from 0.772 to 0.834 and F1-score from 0.739 to 0.881, demonstrating improved predictive capabilities and a more balanced trade-off among precision and recall. It is crucial to compute the p-value using statistical tests like the paired t-test and McNemar's test to make sure that the observed gains are not the result of chance which is shown in Table 10.

The accuracy improvement is statistically significant ($p < 0.001$), according to the paired t-test, which indicates that it is very unlikely to have happened by accident. The distribution of misclassifications, on the other hand, does not significantly change according to McNemar's test ($p > 0.05$), indicating that the improvement is mostly attributable to increased performance and confidence on situations that are previously foreseeable rather than a change in the pattern of right and wrong predictions. Table 10 illustrates this statistical validation, which supports the assertion that DE preprocessing significantly improves the PTB prediction model.

Table 9: Impact on preprocessing data with original data using MLP classifier.

Metrics	Original dataset	Preprocessed dataset using DE	Improved %
Accuracy	0.772	0.834	6.2
Precision	0.701	0.818	16.7
Recall	0.827	0.892	7.9
F1 score	0.739	0.881	19.2
AUC	0.7	0.8	14.3

Source: Authors, (2026).

Table 10: Statistical validation (P test) for Preprocessing result improvement.

Test	Statistic	p-value	Interpretation
Paired t-test (accuracy per fold)	19.487	4.09×10^{-5}	DE significantly improves accuracy over the original dataset ($p < 0.001$).
McNemar's test (prediction changes)	29.0	0.188	No significant change statistically in the proportion of misclassifications ($p > 0.05$).

Source: Authors, (2026).

IV.2 IMPACT ON PROPOSED H-EDFO FEATURE SELECTION

There were 35 available variables at the first appointment, 51 available factors at the second consultation, and 73 available factors at the third consultation. The selected features for each dataset of multi consultation is shown in Table 11. Once the features are selected from each data subset, it is fused and the fused optimal feature set is shown in Figure 6.

Table 11: Selected feature subsets of each multi-consultation dataset using proposed H-EDFO.

Feature Name	Entropy score	DFO Feature selection		Feature Name	Entropy score	DFO Feature selection
B dataset				**C2**		
Maternal Age	0.92	1		Lab Hemoglobin	0.93	1
BMI	0.87	1		WBC Count	0.89	0
Smoking Status	0.84	0		Platelet Count	0.85	1
Blood Pressure	0.81	1		CRP Level	0.82	0
Parity	0.79	0		Glucose Tolerance	0.8	1
Ethnicity	0.75	0		Proteinuria	0.77	0
Education Level	0.71	1		Serum Ferritin	0.73	1
Income Bracket	0.68	0		Vitamin D Level	0.7	0
Alcohol Use	0.65	1		Thyroid Function	0.68	1
Prenatal Visits	0.63	1		Bilirubin	0.65	0
C1				**C3 dataset**		
Ultrasound GA	0.91	1		Gestational Diabetes	0.94	1
Fetal Heart Rate	0.88	1		Hypertensive Disorder	0.9	1
Placenta Position	0.83	0		Anemia Diagnosis	0.86	0
AFI	0.8	1		Infection History	0.83	0
Cervical Length	0.78	1		Preterm Labor Signs	0.81	1
Fetal Weight Estimate	0.74	0		Chorioamnionitis	0.78	0
Fetal Movement Count	0.72	0		Placental Abruption	0.75	1
Multiple Gestation	0.69	1		Oligohydramnios	0.72	0
Cord Position	0.66	1		Polyhydramnios	0.69	0
Presentation	0.62	0		Preeclampsia	0.66	1
U1				**U2**		
Genetic Marker A	0.95	1		Environmental Exposure A	0.91	1
Genetic Marker B	0.88	1		Environmental Exposure B	0.85	1

Source: Authors, (2026).

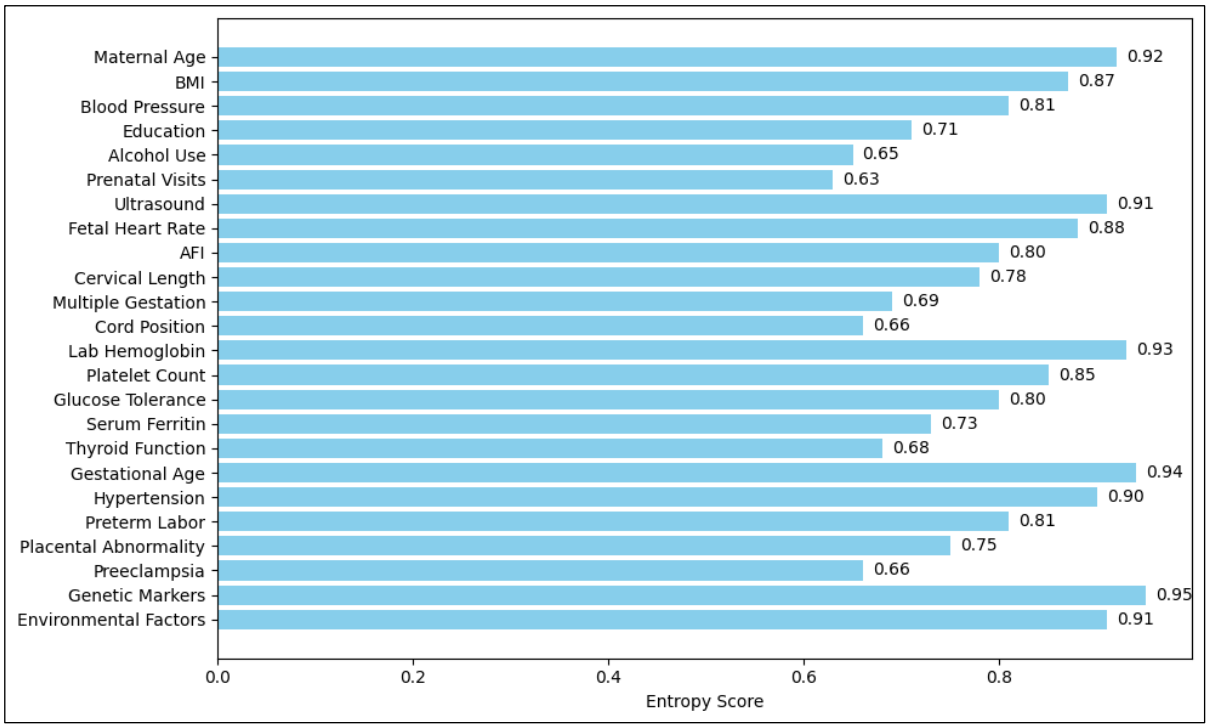


Figure 6: Fused feature set.
Source: Authors, (2026).

IV.3 IMPACT ON CLASSIFICATION MODEL WITH PROPOSED FEATURE SELECTION AS OVERALL PERFORMANCE

To assess the performance of the proposed model, the considered consultation datasets with Ultrasound data are combined together and form three datasets such as D1, D2 and D3 where D1 is the combination of B+C1, D2 is the grouping of B+C1+C2+U1 and D3 is the grouping of B+C1+C2+C3+U1+U2. We looked at the prediction performance for datasets 1, 2, and 3 in the section titled "Model performance improved when data from later prenatal consults were added" before moving on to the Sensitivity analysis on subgroups of PTB section to look at the prediction performance for the severity of preterm birth. In the section titled "Analysis of key variable contribution across three prenatal consult datasets," we also looked at the most crucial variables in the prediction. Based on the dataset's pooled AUC curves, which are likewise depicted in Figure 7, the model effectiveness summary is displayed in Table 12.

Table 12: Validated Gestational period-based datasets – AUC scores.

Datasets	Training	Validation	Testing
D1 =B+C1	0.7511±0.0121	0.7213±0.0225	0.7186±0.0217
D2=B+C1+C2+U1	0.7719±0.0138	0.7426±0.0204	0.7418±0.0136
D3=B+C1+C2+C3+U1+U2	0.8228±0.0092	0.8187±0.0125	0.7992±0.0227

Source: Authors, (2026).

Based on Table 12, the training data performance has been increased from Dataset 1 to Dataset 3 with the value of consult 1 as 0.7511 (Std. Dev as 0.0121) to Consult 3 as 0.8228 (Std. Dev as 0.0092). Similarly, the validation dataset performance also increased from D1 at consult 1 as 0.7213 (Std. Dev as 0.0225) to consult 3 as 0.8187 (Std. Dev as 0.0125). This pattern also reflected in testing data where the AUC scores improved from 0.7186 (Std. Dev as 0.0217) at consult 1 to 0.7992 (Std. Dev as 0.0227) at Consult 3. According to this pattern, predictive power significantly improves with the availability of more thorough gestational data, especially that gathered during the third consultation.

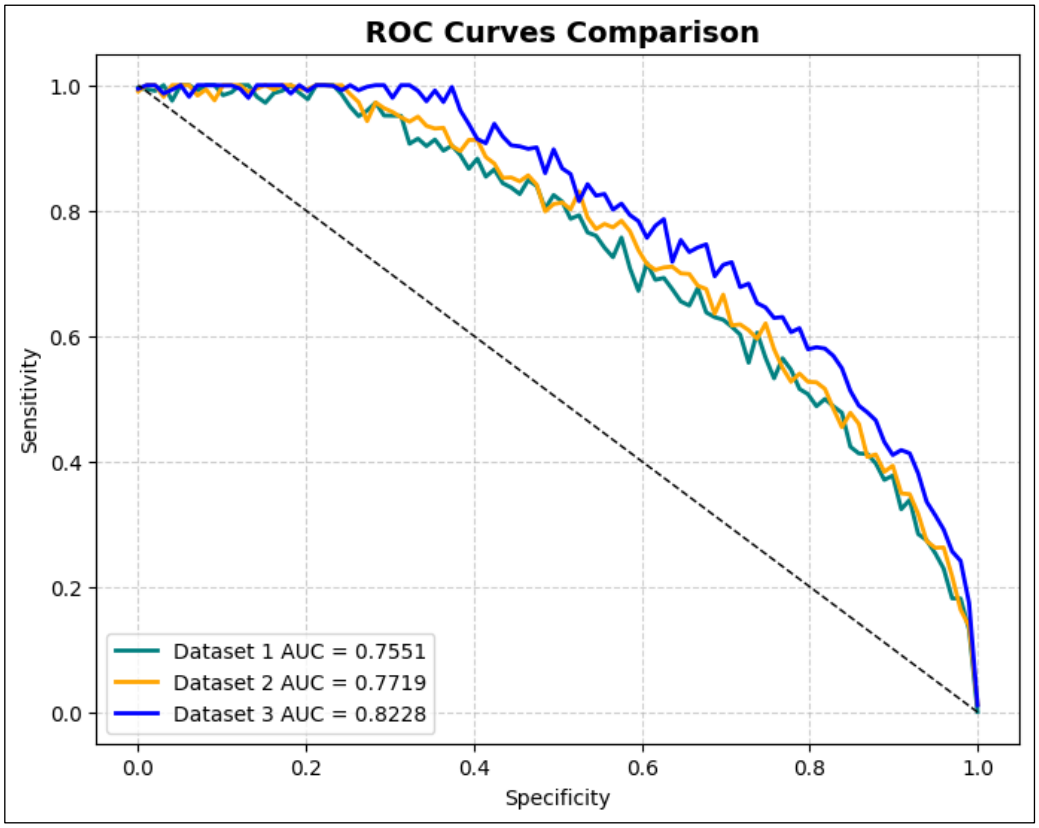


Figure 7: Pooled AUC curves of 6 test fold (B, C1, C2, C3, U1 and U2) merge into 3 datasets (D1, D2 and D3).
Source: Authors, (2026).

The pooled ROC curve in Fig 7, which is created by combining the five test-sets through cross-validation and calculating their combined ROC curve for each dataset, is shown in Fig 2. The AUC of the datasets having data from the first prenatal visit, the first and second prenatal consultations, and the dataset including data from all three prenatal consultations increased noticeably, it offers a visual representation of these improvements. The outcome demonstrated that adding data from consults 2 and 3 significantly enhanced the prediction ability, as seen by a discernible rise in AUC scores.

There appears to be a clear relationship among the amount of gestational data and the predicted accuracy of prenatal assessments, as seen by the improvement in model performance with the addition of later prenatal consult data.A closer look at several performance measures provided crucial information about the models' prediction ability. Table 13 provides more specific information about how predictive power changes when additional prenatal data becomes available over subsequent consultation.

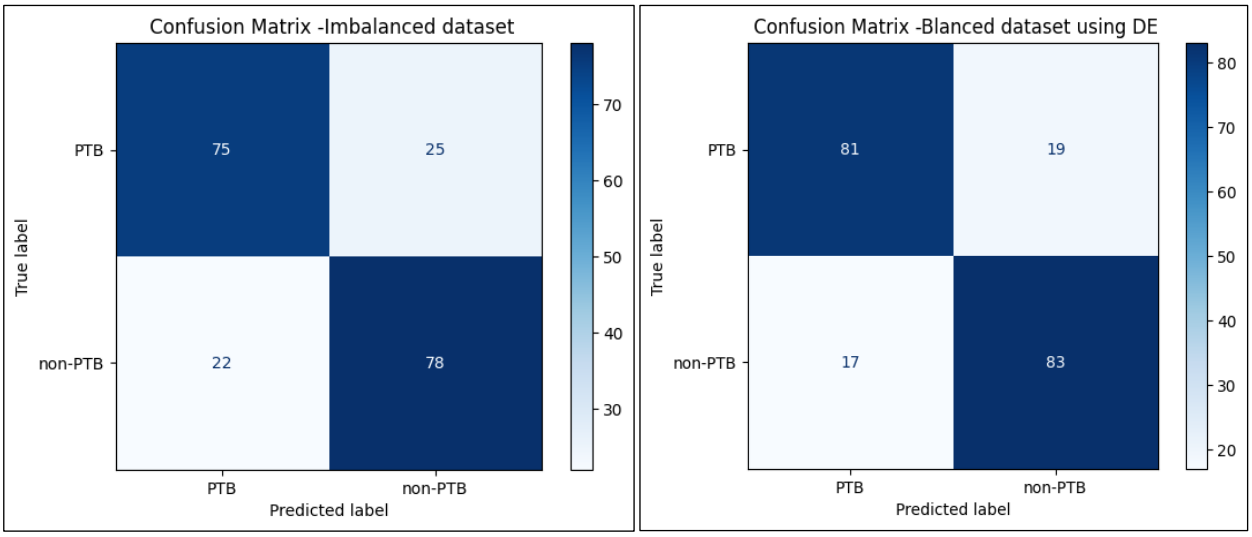
The accuracy at consultation 1 is 72.63 % (Confidence interval of 70% to 73%) which is increased to consultation 2 as 72.83% (Confidence interval of 71.8% to 73.9%) and further improved in the consultation 3 including the selected feature variables as 88.41% with the confidence interval of 82.6% to 89.6%. Similarly, the precision, recall and F1 score values are also increased from consultation 1 to consultation 2 and getting optimum outcome with consultation 3. That proves the improved performance of proposed model efficiency on predicting the PTB with selected fused features with the consultation 3.

Table 13: Pooled performance metrics on the considered 3 datasets with confidence intervals of 95%.

Datasets	AUC score	Accuracy	Precision	Recall	F1 Score
D1	0.7511	0.7263 [0.7018,0. 7328]	0.7283 [0.7128, 0.7372]	0.7467 [0.7362,0.7563]	0.7232 [0.7128, 0.7392]
D2	0.7719	0.7283 [0.7182,0.7394]	0.7364 [0.7283, 0.7462]	0.7653 [0.7526, 0.7782]	0.7563 [0.7462, 0.7628]
D3	0.8228	0.8841 [0.8263, 0.8962]	0.8584 [0.8028, 0.8673]	0.8724 [0.8629, 0.9038]	0.8815 [0.8726, 0.8972]

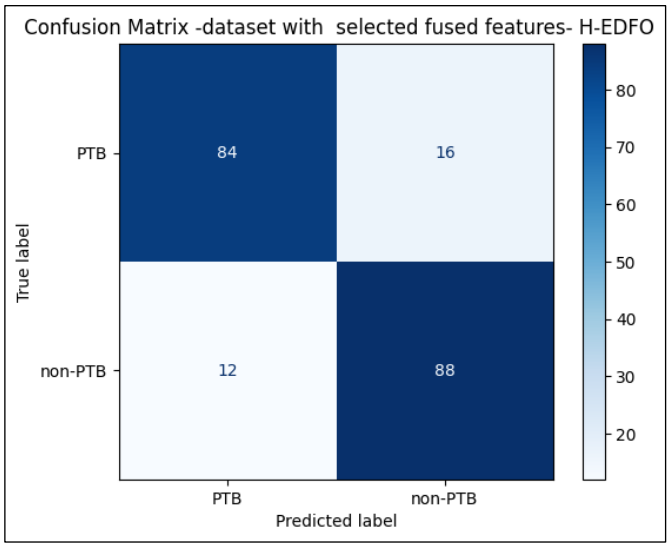
Source: Authors, (2026).

As a whole, the importance of using efficient preprocessing, feature selection based classification is illustrated in Figure 8.



(a)

(b)



(c)

Figure 8: Confusion Matrix for the analysis (a) Imbalanced dataset (b) Balanced dataset (c) Proposed H-EDFO based selected fused features based Prediction.
Source: Authors, (2026).

Three confusion matrices that show how well PTB (preterm birth) prediction models perform under various feature selection and dataset preparation techniques are shown in the figure. The confusion matrix for the unbalanced dataset is displayed in Fig (a). The model misclassified 22 non-PTB instances as PTB and 25 PTB cases as non-PTB, while it properly recognized 75% PTB cases and 78% non-PTB cases.

Due to the model's propensity to favor the dominant class, the comparatively higher number of misclassifications highlights the detrimental effects of class imbalance. The matrix of confusion for the balanced dataset produced by Differential Evolution (DE) is shown in Fig (b). Following balancing, there were fewer misclassifications (19% PTB instances were incorrectly categorized as non-PTB, and 17% non-PTB cases were incorrectly classified as PTB), and the accurate classifications increased to 81% PTB and 83% non-PTB. This illustrates how dataset balance improves the model's capacity to identify minority class occurrences with greater accuracy.

The dataset with specific fused features acquired using the suggested Hybrid Enhanced Feature Optimization (H-EDFO) method is represented by the final confusion matrix in Fig (c) With 84% PTB and 88 non-PTB cases accurately categorized and the lowest misclassification rates (16% PTB cases wrongly projected as non-PTB and 12% non-PTB cases misclassified as PTB), this setup produced the best results. The enhancement demonstrates how well the suggested H-EDFO extracts the most pertinent and discriminative elements, increasing sensitivity and specificity.

IV.4 ANALYSIS ON PTB SUBGROUPS

PTB is defined by the WHO as delivery that occurs before 37 full weeks of pregnancy. Preterm birth is further divided into subcategories according on gestational age [44]. Understanding the variation in results and adjusting interventions appropriately depend on these classifications. These subcategories are described in Table 14.

Table 14: PTB sub groups with gestational age at the time of delivery.

Sub-Category	Gestational age	No. of patients
Extreme -PTB	<28 weeks	69
Very-PTB	28 to 32 weeks	63
Moderate-PTB	32 to 34 weeks	113
Late -PTB	34 to 27 weeks	514

Source: Authors, (2026).

A thorough analysis of our predictive models' pooled sensitivity with a 95% CI for each PTB subgroup across three datasets is given in Table 15. Extreme PTB (delivery before 28 weeks) showed the most remarkable performance, with sensitivity gradually increasing across data sets from 71% to 75% on dataset 1 to dataset 3 respectively. Given the seriousness of extremely preterm deliveries, Data Set 3's unusually high sensitivity is especially significant. The closeness of consult 3 to the delivery date for extreme-PTBs explains this very high sensitivity in Dataset 3. As a result, the data gathered at this visit are more suggestive of the imminence of preterm birth. Given that extremely PTBs necessitate the most urgent and extensive treatment, this great sensitivity is especially important.

With sensitivity beginning at 71.82% (95% CI: 0.7011- 0.7395%) in Dataset 1, peaking at 84.64% (95% CI: 73.82- 85.23%) in Dataset 2, and maintaining durable performance at 85.62% (95% CI: 72.81, 86.23%) in Dataset 3, very-PTB(28–32 weeks) demonstrated equally strong detection rates. The model's sensitivity was relatively modest but remained significant for moderately preterm babies (32–34 weeks). Dataset 1's 74.41% saw a significant improvement to 75.82%, while Dataset 2 had a minor decline to 76.83%. With sensitivity ranging from 74.32% in Dataset 1 to a peak of 75.63% in Dataset 2, before falling to 78.62 in Dataset 3, late-PTB (34–37 weeks) showed the most difficult to forecast. The model demonstrated moderate sensitivity levels when taking into account general-PTB prediction: 74.67% in Dataset 1, increasing to 76.53 % in Dataset 2, and retaining 87.24% in Dataset 3.

Table 15: Pooled Sensitivity analysis of PTB subgroups.

Subgroups	Dataset-1	Dataset-2	Dataset-3
Extreme PTB	0.7161 [0.7113,0. 7227]	0.7382 [0.7221, 0.7471]	0.7523 [0.7423-0.7625]
Very PTB	0.7182 [0.7011,0.7395]	0.8464 [0.7382, 0.8523]	0.8562 [0.7281, 0.8623]
Moderate PTB	0.7441 [0.7333, 0.7728]	0.7582 [0.7422, 0.7671]	0.7683 [0.7521, 0.7724]
Late PTB	0.7432 [0.7327, 0.7592]	0.7563 [0.7434, 0.7627]	0.7862 [0.7234, 0.7937]
General PTB	0.7467 [0.7362,0.7563]	0.7653 [0.7526, 0.7782]	0.8724 [0.8629, 0.9038]

Source: Authors, (2026).

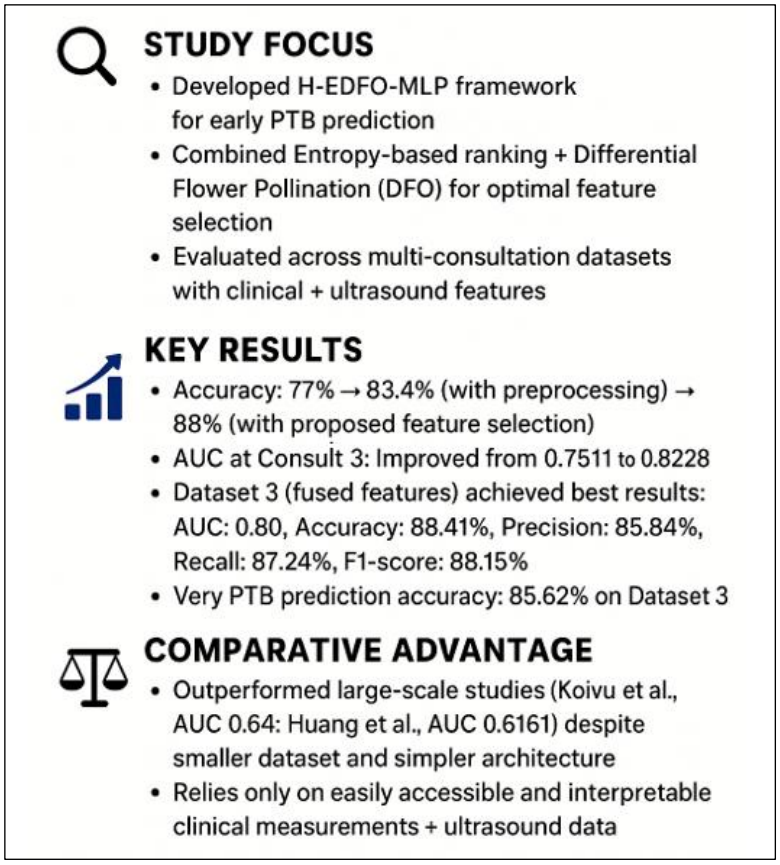


Figure 9: Research Highlights.
Source: Authors, (2026).

IV.5 DISCUSSIONS

This study demonstrates the predictive performance of the proposed feature selection model PTB at multi-consulting which are more effective in consult 3 at the gestational age of 22 to 29 weeks, with Improved AUC of 0.7511 to 0.8228 at consult 3. These results are comparable to large scale analysis of existing literatures such as [45] employed ML algorithms such as gradient bossing and neural network with 16 million observations and obtained 0.64 as AUC score for PTB prediction. Huang et al., [19] employed ElasticNet based PTB prediction on multi visit data with the inclusion of all the variables and obtained the AUC score of 0.6161. The suggested model performed well using only easily accessible clinical factors as well as ultrasound data at Consult 3, despite the fact that both studies used much larger datasets and more sophisticated algorithms. This suggests that the timing as well as type of selected clinical data using the suggested H-EDFO is important than sample size.

The result highlights are pictorially illustrated in Figure 9 for better understanding. The analysis of PTB prediction based on efficient preprocessing, feature selection and classification using proposed models efficiently improves the performance on consult 3 during the 22 to 29 gestational weeks with the selected features. The DE approach improves the model performance on preprocessing data with balancing factors. The model with preprocessed data can improve 6% of accuracy without preprocessing data along with all variables. Further, the feature selection model is proposed to select the most influencing features of all considered visits using entropy and DFO methods. The entropy score is computed for all the base and consultation datasets variables and based on the score; the feature ranks are given. The entropy score is treated as the fitness function to find optimal feature subsets using DFO which selects the relevant features.

These selected feature subsets from all the considered consultation periods are fused as a single feature set for further processing. These results are analyzed and compared as a confusion matrix and accuracy without preprocessing is 77%, with preprocessing is 83.4% and with feature selection model, it is improved as 88% of accuracy. Hence, the proposed model efficiently improves the prediction of PTB with improved accuracy and in terms of other metrics too. Furthermore, the model efficiency is implemented with the combination of these consultations and framed three datasets by combining Base checkup, C1, C2, C3 along with Ultrasound 1 and Ultrasound 2 data. These three datasets are evaluated using the proposed framework in terms of AUC, Accuracy, precision, recall and F1 score. Noticeable results are obtained on dataset 3 by obtaining 0.8, 88.41%, 85.84%, 87.24% and 88.15% respectively than Dataset 1 and 2.

Which means, the dataset 3 is the combination of all the considered consultation data's fused feature sets. The selected features using proposed H-EDFO efficiently improves the performance while predicting the PTB. We also looked at the model's performance on various preterm birth subgroups, and the fact that the sensitivity varies among PTB categories highlights the significance of customizing model creation and validation for particular clinical situations. The other variant of PTB also predicted using the model and very PTB prediction secured the improved accuracy of 85.62% of accuracy on Dataset 3 than other datasets. It is proved that our model efficiently predicts the PTB on Dataset 3 with fused feature sets from all the consultation. To improve the predicted accuracy of our model, we used MLP to incorporate demographic, lifestyle, obstetric, and medical history data. This method outperformed more intricate models by preserving interpretability and usability, which are essential for clinical decision-making.

IV.6 IMPLICATIONS

IV.6.1 Theoretical Implication

The theoretical knowledge of how entropy-based ranking and DFO optimization might work together to provide optimal, clinically relevant feature subsets is advanced by the suggested H-EDFO framework. The results provide credence to the idea that longitudinal data integration is crucial for early and precise PTB risk assessment by demonstrating that the combination of variables from several consultations improves classification accuracy. The study further supports the idea of effective predictive modeling for medical diagnostics by demonstrating the possibility of combining clinical measurements with ultrasound data to get greater accuracy without the need for high-dimensional, resource-intensive datasets. The results of the confusion matrix, which demonstrate an improvement in accuracy from 77% (without preprocessing) to 83.4% (with preprocessing) and then to 88% (with feature selection), empirically support the theoretical idea that focused feature selection and methodical data preparation are essential to the success of predictive modeling.

In order to find clinically significant features across several consultation points, the suggested H-EDFO framework combines DFO optimization with entropy-based ranking, advancing methodological comprehension. The theoretical claim that the kind and timing of clinical data are just as significant as dataset size or model complexity is supported by the noteworthy AUC improvement from 0.7511 to 0.8228 at Consult 3. By showing that smaller but more carefully crafted feature sets can perform better than bigger, less curated datasets, this supports and expands on earlier findings in the literature and helps create effective and understandable AI models for medical diagnosis.

IV.6.2 Practical Implications

From a clinical standpoint, the suggested approach provides healthcare providers with a workable and affordable decision-support tool, especially in environments with limited resources. The performance improvements, especially on Dataset 3 (AUC: 0.8, Accuracy: 88.41%, Precision: 85.84%, Recall: 87.24%, F1 Score: 88.15%), show the model's potential as a useful clinical decision-support tool from the perspective of real-world applications. The approach can be readily implemented in healthcare settings with low resources or in remote areas because it just uses routinely collected clinical parameters and ultrasound data, negating the need for sophisticated computational equipment. Clinicians are able to implement timely treatments, including increased monitoring, dietary counseling, or expert referrals, thanks to the increased accuracy throughout the 22–29 week gestational period. These interventions have the potential to considerably lower PTB-related problems. Additionally, the model's high interpretability encourages adoption and confidence among medical professionals, which makes it appropriate for incorporation into electronic health record systems.

IV.6.3 Limitations

Compared to large-scale PTB prediction studies, the study's dataset was comparatively smaller, which would have limited its generalizability to larger populations. Because the data came from a particular hospital setting, it might not accurately reflect how PTB risk variables differ in various geographical or ethnic contexts. Even while the model performed well with a few features, several biological and environmental risk factors like genetic markers and comprehensive nutrition data were left out. Robustness must be confirmed by external validation in prospective clinical scenarios and on separate datasets.

V. CONCLUSIONS

This study integrated multilayer perceptron classification, Differential Flower Pollination optimization, and entropy-based ranking to develop and test the H-EDFO-MLP framework for preterm birth (PTB) prediction. Targeted feature selection, multi-consultation data fusion, and meticulous preprocessing all contributed to the framework's notable performance gains. Notably, the AUC rose from 0.7511 to 0.8228 at Consult 3, and accuracy increased from 77% without preprocessing to 88% with the suggested feature selection. Our approach, which just used readily available clinical and ultrasound characteristics, showed better prediction capacity than large-scale studies employing more intricate models.

These findings demonstrate that timing and data type are more important than computational complexity or dataset size. Our findings have ramifications that go beyond predictive modeling. First, utilizing this prediction model during the 22–29-week consult may assist doctors in identifying individuals who need close monitoring and early interventions, as indicated by the model's high sensitivity for very-PTB (85.62% at Consult 3). This timing coincides with important clinical decision points, like the prenatal corticosteroid injection and the planning of the mother's transfer to facilities equipped to provide newborn care.

Without the need for sophisticated computing resources or a large infrastructure, the model can be simply incorporated into current maternal care workflows by using readily available clinical and ultrasound characteristics gathered during normal prenatal consultations. Early interventions like closer monitoring, preventive treatments, or referral to higher-level care facilities are made possible by the increased prediction accuracy, particularly during the gestational window of 22–29 weeks. These actions have the potential to greatly lower negative outcomes. Additionally, the interpretability of the model increases physician acceptance and confidence in AI-driven suggestions.

Additionally, the method offers a framework that is scalable, meaning that it may be applied to other pregnancy risk evaluations or expanded to other populations with little retraining. To increase generalizability, the study will be expanded in future to encompass a variety of healthcare systems and demographics. combining biochemical, lifestyle, environmental, genetic, and ECG data to predict apnea levels and capture a wider risk profile of PTB. Putting in place online learning tools to enable the model to update continually when fresh clinical practice data becomes available.

VI. AUTHOR'S CONTRIBUTION

Conceptualization: T. Ummal Sariba Begum, R. Renuga Devi.
Methodology: T. Ummal Sariba Begum, R. Renuga Devi.
Investigation: T. Ummal Sariba Begum, R. Renuga Devi.
Discussion of results: T. Ummal Sariba Begum, R. Renuga Devi.
Writing – Original Draft: T. Ummal Sariba Begum, R. Renuga Devi.
Writing – Review and Editing: T. Ummal Sariba Begum, R. Renuga Devi.
Resources: T. Ummal Sariba Begum, R. Renuga Devi.
Supervision: T. Ummal Sariba Begum, R. Renuga Devi.
Approval of the final text: T. Ummal Sariba Begum, R. Renuga Devi.

VII. ACKNOWLEDGMENTS

The authors thank the National Institute of Child Health and Human Development (NICHD) and the nuMoM2b research team for providing access to the dataset used in this study. The authors also acknowledge SRM Institute of Science and Technology, Ramapuram Campus, Chennai, India, for providing academic support and resources that facilitated this research

VIII. REFERENCES

[1] C.A. Espinosa, W. Khan, R. Khanam, S. Das, J. Khalid, J. Pervin, et al., Multiomic signals associated with maternal epidemiological factors contributing to preterm birth in low- and middle-income countries, Sci Adv, 9, eade7692 (2023).

[2] R.O. Bahado-Singh, J. Sonek, D. McKenna, D. Cool, B. Aydas, O. Turkoglu, et al., Artificial intelligence and amniotic fluid multiomics: prediction of perinatal outcome in asymptomatic women with short cervix, Ultrasound Obstet Gynecol, 54, 110–118 (2019).

[3] Z. Sharifi-Heris, J. Laitala, A. Airola, A.M. Rahmani, M. Bender, Machine learning approach for preterm birth prediction using health records: systematic review, JMIR Med Inform, 10, e33875 (2022).

[4] N. Baumann, P. Bartmann, D. Wolke, Health-related quality of life into adulthood after very preterm birth, Pediatrics, 137(4), e20153748 (2016).

[5] FIGO Working Group on Good Clinical Practice in Maternal–Fetal Medicine, Good clinical practice advice: prediction of preterm labor and preterm premature rupture of membranes, Int J Gynaecol Obstet, 144(3), 340–346 (2019).

[6] C. Ye, T. Fu, S. Hao, Y. Zhang, O. Wang, B. Jin, et al., Prediction of incident hypertension within the next year: prospective study using statewide electronic health records and machine learning, J Med Internet Res, 20(1), e22 (2018).

[7] Z.A. Oskovi Kaplan, A.S. Ozgu-Erdinc, Prediction of preterm birth: maternal characteristics, ultrasound markers, and biomarkers—an updated overview, J Pregnancy, 2018, 8367571 (2018).

[8] S. Crespi, X.H. Hu, B. Johnson, G. Griesinger, W.L. Ledger, O. Heisel, E.M. Kolibianakis, Patient burden and healthcare resource utilization of regimens for ovarian stimulation, Reprod Biomed Online, 44(3), 573–581 (2022).

[9] M.S. Kramer, K. Demissie, H. Yang, R.W. Platt, R. Sauvé, R. Liston, et al., Contribution of mild and moderate preterm birth to infant mortality, JAMA, 284(7), 843–849 (2000).

[10] A. Koire, D.M. Chu, K. Aagaard, Family history is a predictor of current preterm birth, Am J Obstet Gynecol MFM, 3(1), 100277 (2021).

[11] M. Lv, C. Chen, L. Qiu, N. Jin, M. Wang, B. Zhao, et al., A nomogram to predict extremely preterm birth in women with singleton pregnancies undergoing cervical cerclage, Heliyon, 8(10), e10731 (2022).

[12] T. Tran, W. Luo, D. Phung, J. Morris, K. Rickard, S. Venkatesh, Preterm birth prediction: deriving stable and interpretable rules from high-dimensional data, Stat, 1050, 28 (2016).

[13] A. Weber, G.L. Darmstadt, S. Gruber, M.E. Foeller, S.L. Carmichael, D.K. Stevenson, et al., Application of machine learning to predict early spontaneous preterm birth among nulliparous non-Hispanic black and white women, Ann Epidemiol, 28(11), 783–789 (2018).

[14] A. Esty, M. Frize, J. Gilchrist, E. Bariciak, Applying data preprocessing methods to predict premature birth, in Proc. IEEE Eng Med Biol Soc Conf, 6096–6099 (2018).

[15] C. Gao, S. Osmundson, D.R.V. Edwards, G.P. Jackson, B.A. Malin, Y. Chen, Deep learning predicts extreme preterm birth from electronic health records, J Biomed Inform, 100, 103334 (2019).

[16] Y. Zhang, S. Lu, Y. Wu, W. Hu, Z. Yuan, The prediction of preterm birth using time-series technology-based machine learning: retrospective cohort study, JMIR Med Inform, 10(6), e33835 (2022).

[17] S. Kim, M.K. Higgins, P. Brennan, G.M. Slavich, L. Xiong, V. Hertzberg, et al., Interpretable race/ethnicity-specific machine learning models of chronic stressors to predict preterm birth: a national PRAMS analysis, Stud Health Technol Inform, 315, 267–272 (2024).

[18] Q.-Y. Yu, Y. Lin, Y.-R. Zhou, X.-J. Yang, J. Hemelaar, Predicting risk of preterm birth in singleton pregnancies using machine learning algorithms, Front Big Data, 7, 1291196 (2024).

[19] C. Huang, X. Long, M. van der Ven, et al., Predicting preterm birth using electronic medical records from multiple prenatal visits, BMC Pregnancy Childbirth, 24, 843 (2024).

[20] R. Raja, I. Mukherjee, B.K. Sarkar, A machine learning-based prediction model for preterm birth in rural India, J Healthc Eng, 2021, 6665573 (2021).

[21] J. Jaskari, et al., Machine learning methods for neonatal mortality and morbidity classification, IEEE Access, 8, 123347–123358 (2020).

[22] Q. Sun, X. Zou, Y. Yan, H. Zhang, S. Wang, Y. Gao, et al., Machine learning-based prediction model of preterm birth using electronic health records, J Healthc Eng, 2022, 9635526 (2022).

[23] Q.Y. Yu, Y. Lin, Y.R. Zhou, X.J. Yang, J. Hemelaar, Predicting risk of preterm birth in singleton pregnancies using machine learning algorithms, Front Big Data, 7, 1291196 (2024).

[24] M. Sharma, S. Bapodara, J. Tiwari, U.R. Acharya, Automated sleep apnea detection in pregnant women using wavelet-based features, Inform Med Unlocked, 32, 101026 (2022).

[25] J.S. Park, K.S. Lee, J.S. Heo, et al., Clinical and dental predictors of preterm birth using machine learning methods: the MOHEPI study, Sci Rep, 14, 24664 (2024).

[26] T. Priya, B.K. Sarkar, S.K. Sahana, Use of machine learning techniques for early prediction of birth cases, Procedia Comput Sci, 258, 125–135 (2025).

[27] Nulliparous Pregnancy Outcomes Study: Monitoring Mothers-to-Be (nuMoM2b) dataset, Eunice Kennedy Shriver National Institute of Child Health and Human Development, NICHD DASH (2015).

[28] D.M. Haas, C.B. Parker, D.A. Wing, S. Parry, W.A. Grobman, B.M. Mercer, et al., Methods of the Nulliparous Pregnancy Outcomes Study: monitoring mothers-to-be (nuMoM2b), Am J Obstet Gynecol, 212(4), 539.e1 (2015).

[29] R.D. Riley, J. Ensor, K.I. Snell, F.E. Harrell, G.P. Martin, J.B. Reitsma, et al., Calculating the sample size required for developing a clinical prediction model, BMJ, 368, m441 (2020).

[30] Z. Sharifi-Heris, J. Laitala, A. Airola, A.M. Rahmani, M. Bender, et al., Machine learning approach for preterm birth prediction using health records: systematic review, JMIR Med Inform, 10(4), e33875 (2022).

[31] H. Rawashdeh, S. Awawdeh, F. Shannag, E. Henawi, H. Faris, N. Obeid, et al., Intelligent system based on data mining techniques for prediction of preterm birth in women with cervical cerclage, Comput Biol Chem, 85, 107233 (2020).

[32] G.P. Zhang, Neural networks for classification: a survey, IEEE Trans Syst Man Cybern C, 30(4), 451–462 (2000).

[33] T.T.C. Wong, X. Yong, J.S.Z. Tung, B.J.Y. Lee, J.M.X. Chan, R. Du, et al., Prediction of labour onset in women with symptoms of preterm labour using cervical length, BMC Pregnancy Childbirth, 21(1), 359 (2021).

[34] B. Wouter, M.C. Schut, A. Abu-Hanna, J.G. van Baal, J.J. van Netten, S.A. Bus, Development of a prediction model for foot ulcer recurrence in people with diabetes using easy-to-obtain clinical variables, BMJ Open Diabetes Res Care, 9(1), e002257 (2021).

[35] T. Ummal Sariba Begum, R. Renuga Devi, Neonatal apnea level classification using effective preprocessing of preterm birth health record, in Proc. 6th Int. Conf. Data Intell Cogn Inform (ICDICI), IEEE, 2025.

[36] M.M. Rahman, D.N. Davis, Machine learning-based missing value imputation method for clinical datasets, in IAENG Trans Eng Technol, Springer, Cham, 245–257 (2013).

[37] P. Cavoretto, N. Salmeri, M. Candiani, A. Farina, Reference ranges of uterine artery pulsatility index from first to third trimester based on serial Doppler measurements: longitudinal cohort study, Ultrasound Obstet Gynecol, 61(4), 474–480 (2023).

[38] A. Fernández, S. García, F. Herrera, N.V. Chawla, SMOTE for learning from imbalanced data: progress and challenges, marking the 15-year anniversary, J Artif Intell Res, 61, 863–905 (2018).

[39] Y. Meraihi, A. Ramdane-Cherif, D. Acheli, et al., Dragonfly algorithm: a comprehensive review and applications, Neural Comput Appl, 32, 16625–16646 (2020).

[40] M.A. Hall, Correlation-based feature selection for discrete and numeric class machine learning, in Proc. 17th Int. Conf. Mach Learn, Stanford, USA, 359–366 (2000).

[41] M.C. Rahman, T.A. Rashid, Dragonfly algorithm and its applications in applied science survey, Comput Intell Neurosci, 2019, 9293617 (2019).

[42] T. Mehmod, H.B.M. Rais, Ant colony optimization and feature selection for intrusion detection, in Advances in Machine Learning and Signal Processing, vol. 387, Springer, New York, 305–312 (2016).

[43] P.N. Dadhich, S. Hanaoka, Markov method integration with multilayer perceptron classifier for simulation of urban growth of Jaipur City, in Proc. 6th WSEAS Int. Conf. Remote Sens, Iwate, Japan, 118–123 (2010).

[44] World Health Organization, Preterm birth, WHO Fact Sheet (2023).

[45] A. Koivu, M. Sairanen, Predicting risk of stillbirth and preterm pregnancies with machine learning, Health Inf Sci Syst, 8, 1–12 (2020).