Received: 10 October 2024 | Revised: 9 December 2024 | Accepted: 18 February 2025 | Published online: 6 March 2025

Artificial Intelligence and Applications
2025, Vol. 00(00) 1–14
DOI: 10.47852/bonviewAIA52024517
BON VIEW PUBLISHING

RESEARCH ARTICLE

Preterm Birth Prediction Using an Explainable Machine Learning Approach

Ahmad Hassan[1,*], **Saba Nawaz**[2], **Sidra Tahira**[3] **and Arslan Ahmed**[4]

[1] *Department of Computer Science, COMSATS University Islamabad, Pakistan*

[2] *Department of Computer Science, Capital University of Science and Technology, Pakistan*

[3] *Department of Computer Science and Information Technology, Virtual University, Pakistan*

[4] *Department of Medicine and Allied, King Edward Medical University, Pakistan*

Abstract: Pregnancy and childbirth are critical periods, with complications like preterm birth presenting serious risks to both mother and baby. Preterm birth is defined as delivery before 37 weeks of gestation. It poses serious health risks to newborns, including respiratory complications, developmental delays, and long-term disabilities. Early prediction of preterm birth can enable timely interventions to reduce these risks and improve maternal-fetal health outcomes. This paper proposes a machine learning-based framework for preterm birth prediction using a blended approach. This approach combines multiple classifiers into blended model to enhance the prediction accuracy. Feature selection techniques such as variance threshold, Pearson's correlation, and mutual information are applied exclusively to the training set to enhance model performance. The proposed blended model consistently outperforms the standalone base models. It achieves the highest accuracy of 74.61%, precision of 71.83%, recall of 70.94%, $F1$ score of 71.38%, and an area under the curve (AUC) of 84.56% during 5-fold cross-validation. For holdout testing, it also maintains a superior performance with an accuracy of 73.82% and AUC of 83.91%. Additionally, explainable artificial intelligence techniques are applied to interpret the model's predictions. It identifies that abnormal amniotic fluid, prenatal care, and household income have a substantial impact on predicting both preterm and normal birth outcomes. The proposed approach shows strong potential for real-world healthcare applications to offer clinicians valuable insights for early preterm birth intervention.

Keywords: preterm birth, machine learning, healthcare analytics, predictive modeling, explainable artificial intelligence

1. Introduction

Pregnancy typically lasts around 40 weeks [1]. During this time, the baby develops and grows inside the womb. This period known as gestation is crucial for the baby's health and development. A full-term pregnancy allows the baby to reach important milestones that prepare it for life outside the womb [2]. However, when labor begins before 37 weeks, it results in a preterm birth [3]. Preterm birth is a significant concern because it often leads to various health challenges for the newborn. Globally, around 15 million babies are born preterm each year [4]. This represents more than one in ten of all live births. The rates of preterm births are particularly high in low- and middle-income countries [5]. The estimated national preterm birth rates in 2020 on the world map are depicted in Figure 1 [6], that shows the global prevalence of this significant medical issue [6]. Factors like inadequate prenatal care and maternal health issues contribute to the problem.

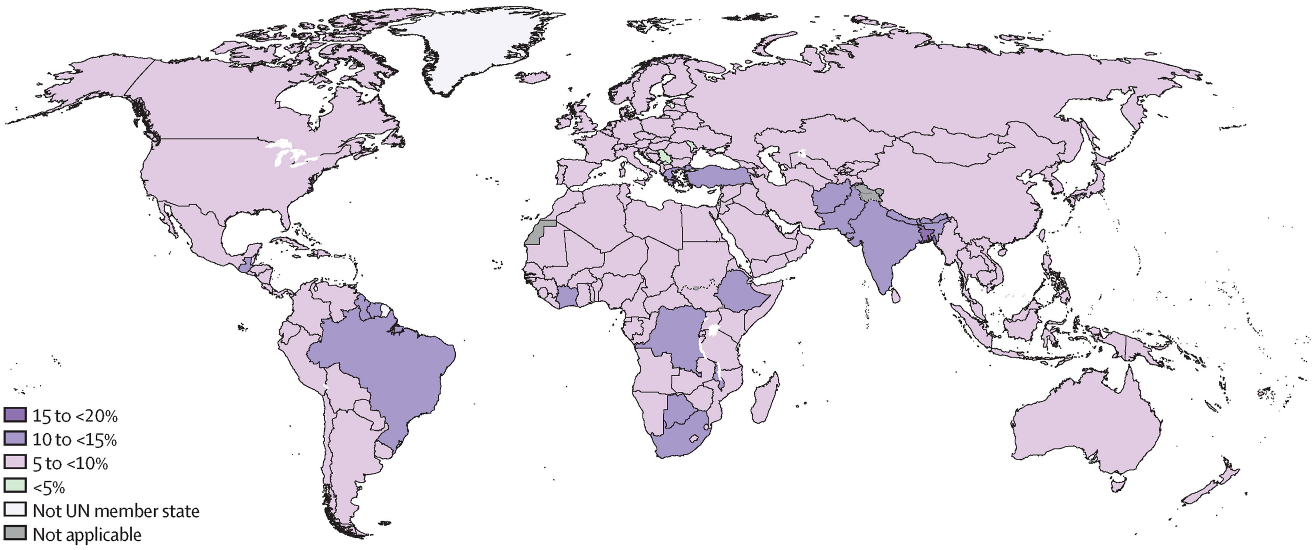
The South Asia and sub-Saharan Africa regions contribute to 49% and 40% of all neonatal deaths caused by preterm, respectively [7]. Complications arising from preterm birth are responsible for nearly one million deaths annually [8]. These complications can also lead to long-term disabilities in survivors. This makes preterm birth a critical issue in global health. Preterm birth significantly impacts the health of both the mother and the baby at various stages of life. For the mother, preterm labor often leads to physical and emotional stress [9]. She may experience anxiety and fear about the health of her baby, especially if the birth occurs very early. Physically, the mother might face complications related to the delivery, including infections or the need for emergency interventions [10]. Additionally, the emotional toll of caring for a premature baby, who may require intensive medical attention, can lead to postpartum depression or anxiety [11].

For the baby, the consequences of being born preterm vary depending on how early the birth occurs. In the fetal stage, preterm birth interrupts crucial developmental processes. Organs like the eyes, kidneys, lungs, brain, and heart may not be fully developed, leading to immediate health challenges after birth [12]. As an infant, a preterm baby is at higher risk for respiratory issues, infections, and feeding difficulties [13]. These babies often require prolonged stays in neonatal intensive care units to monitor and support their development [14]. As they grow, preterm children may face long-term health issues such as chronic lung disease, vision and hearing impairments, and developmental delays [15]. The impact of preterm birth can extend into adulthood, where

---

*__Corresponding author:__ Ahmad Hassan, Department of Computer Science, COMSATS University Islamabad, Pakistan. Email: fa22-rcs-002@cuiwah.edu.pk

Figure 1
The visual overview of global estimated national preterm birth rates



individuals may have an increased risk of chronic conditions like hypertension, diabetes, and cardiovascular disease [16]. Therefore, preterm birth affects health across the lifespan. It is a critical issue for both the immediate and long-term well-being of the mother and child.

Preterm birth imposes significant health and economic burdens. Medical costs for premature infants are often several times higher than those for full-term babies [17]. These costs stem from extended hospital stays, specialized care, and long-term health complications that may require ongoing treatment throughout life. Machine learning (ML) offers a promising solution to address these challenges by enhancing the ability to predict and manage preterm births [18]. By analyzing medical datasets, ML models can identify patterns and risk factors that traditional methods might overlook. This enables early prediction of preterm births and appropriate interventions [19]. Early detection not only improves health outcomes for both the mother and baby but also reduces the associated healthcare costs by preventing complications and optimizing care pathways. By integrating ML into prenatal care, healthcare providers can better allocate resources and deliver personalized care [20]. This ultimately improves the quality of life for millions of families affected by preterm birth.

Moreover, interpretability of prediction results using ML is crucial in healthcare, where decisions affect patient lives [21]. For preterm prediction, clinicians need models they can trust and understand. This is where explainable artificial intelligence (XAI) becomes essential. XAI makes ML models transparent by showing how and why a prediction is made by highlighting key factors like medical history [22]. This clarity helps clinicians ensure that the model's decisions align with medical knowledge that enables them to use predictions confidently. The XAI improves decision-making for preterm birth predictions by bridging the gap between complex models and clinical needs.

1.1. Research gap

Despite significant advancements in understanding preterm birth risk factors and complications, prediction remains a challenge. This is due to the complex interplay of medical, genetic, and socio-environmental factors. Existing studies often lack interpretability. This limits their integration into clinical workflows. Furthermore, many studies focus on traditional statistical methods or single ML models. These may not provide robust or reliable results. Only a limited number of studies apply blending techniques. Even fewer leverage XAI to interpret predictions. To the best of the authors' knowledge, there is only one prior study that utilizes the same dataset for preterm birth prediction. However, it does not explore blending models (BM) or multiple feature selection techniques. This research addresses these gaps by proposing a novel BM framework. It is augmented with XAI and multiple feature selection techniques to improve prediction accuracy and interpretability.

1.2. Research goal

The primary goal of this study is to enhance the prediction and understanding of preterm births using ML. The study develops a BM that combines the strengths of multiple standalone classifiers. It incorporates feature selection techniques to achieve more reliable predictions. Additionally, the use of XAI ensures that the predictions are interpretable. This aids clinicians in making informed decisions.

1.3. Research question

To achieve the aforementioned research goal, this study seeks to answer the following research questions:

1) Can a BM outperform standalone ML models in predicting preterm birth?
2) What are the most informative and effective features for preterm prediction using various feature selection techniques?
3) How can XAI methods enhance the interpretability of preterm birth predictions and make them clinically actionable?

To provide a comprehensive overview, Section 2 continues to expand the discussion on the body of knowledge available by presenting recent works that focus on preterm prediction. Section 3 provides an overview of the main research methodology. It explains the preterm prediction dataset and preprocessing techniques in a clear and precise manner. Furthermore, it provides a comprehensive explanation of the proposed methodology for predicting preterm. Section 4 presents the details of the prediction results and includes a discussion subsection. In conclusion, Section 5 summarizes the findings of this research and presents the conclusion. It also looks at possible areas for subsequent research studies.

2. Literature Review

The literature review section focuses on examining the most recent studies on the use of diverse materials and techniques to predict preterm delivery. It highlights the latest advancements and discoveries in this crucial preterm prediction field. Preterm birth is a leading cause of newborn death. A recent study [23] utilized data from 9,550 pregnant women in China from 2008 to 2018. At 27 weeks of gestation, the random forest (RF) model had the greatest accuracy of 0.816 and area under the curve (AUC) of 0.885, outperforming the other five algorithms that were tested. It was discovered that magnesium levels, fundal height, maternal age, and serum biomarkers were significant predictive indicators. In another study [24], Raman spectroscopy was utilized to investigate the maternal metabolome in the first trimester plasma samples of patients who delivered preterm and those who delivered at term. The study identified fifteen significant metabolites predictive of preterm birth. The study highlights the importance of features, i.e., maternal age and body mass index (BMI) as key contributors to metabolic changes associated with preterm.

By including cytokine levels into models, this study [25] tried to predict preterm birth. Two RF models were created using data from pregnant women between 18 and 23.6 weeks of gestation: first an adjusted model that concentrated on maternal age, interleukin-2, and cervical length, and second one a comprehensive model with 12 variables. With an 87% detection rate and an AUC of 0.875, the second adjusted model increased the prediction performance. Another study [26] explored the use of prenatal oral microbiome to predict preterm birth, involving 59 pregnant women with 30 in the preterm birth group and 29 in the full-term birth group. The study identified 25 differentially abundant taxa, with 22 enriched in full-term births and 3 in preterm births. Based on nine most significant taxa, researchers achieved 76.5% accuracy using RF classifier. This study suggests the potential of using oral microbiome for preterm birth prediction.

Another study [18] utilized electronic health records for the prediction of preterm births. The study employed a cross-sectional design and utilized multifactor logistic regression analysis to assess the risk factors associated with preterm birth. Five different ML models including logistic regression, decision tree (DT), naive Bayes (NB), support vector machine (SVM), and AdaBoost model were constructed. Among them, the AdaBoost demonstrated the highest accuracy of 72.73% for the preterm. This study [27] used Shapley additive explanations (SHAP) to identify key risk factors of preterm to construct ML models. A total 35 risk factors were investigated through the analysis of data from 3,509 pregnant women in the United Arab Emirates. The XGBoost model performed the best, with an AUC of 72.3% for nulliparous women and 73.5% for parous women. Previous preterm birth, cesarean sections, and preeclampsia were important risk factors for parous women, while amniotic infection, maternal age, and BMI were crucial for nulliparous women, according to the analyses.

A recent study [28] presented a different approach for predicting preterm deliveries using electrohysterography (EHG) signals. It is a non-invasive monitoring technique crucial for assessing labor progression. This involved three phases: preprocessing EHG signals with a band-pass filter and wavelet transform to remove noise, extracting features like Shannon energy and median frequency. It predicted preterm outcomes using an enhanced sheep flock-optimized hybrid extreme artificial neural learning network. Yu et al. [29] also developed ML models on the data of 22,603 singleton pregnancies to predict preterm birth. The data were collected from 51 midwifery clinics and hospitals in China. Based on permutation importance, various algorithms including CatBoost, RF, and deep neural networks, were applied, with feature selection. After 26 weeks of gestation, the CatBoost model achieved the best performance with an AUC of 70%. It identified key predictors like antenatal care visits and maternal health metrics.

Another study [30] evaluated the efficacy of deep learning algorithms for predicting preterm birth using data from 30,965 births. Four ML models were compared: logistic regression, RF, SVM, and transformer. With an AUC of 79.20%, the transformer model outperformed the others and an achieved accuracy of 72.61%. This study [31] investigated the effects of preterm delivery on infant temperament development in order to address the worldwide public health challenge of preterm birth. The study used both broad and narrow temperament variables to classify birth status as preterm vs. full-term. A meta-analysis involving 19 samples with 201 preterm and 402 full-term children was carried out by the study. The work used statistical techniques in this field while also adding more understanding of temperament in preterm children by merging data from multiple investigations.

In summary, while significant advancements have been made in preterm birth prediction through the use of ML models, several gaps still exist. Most of the datasets utilized in these studies are not publicly accessible due to privacy concerns, which limits reproducibility and wider validation. Additionally, many studies focus on standalone models and have yet to explore more sophisticated ensemble techniques such as stacking models with meta-learning, which have the potential to improve predictive performance by leveraging multiple algorithms. Addressing these gaps could further enhance the accuracy and generalizability of preterm birth prediction models. Based on the gaps identified in the existing literature, this study hypothesizes that:

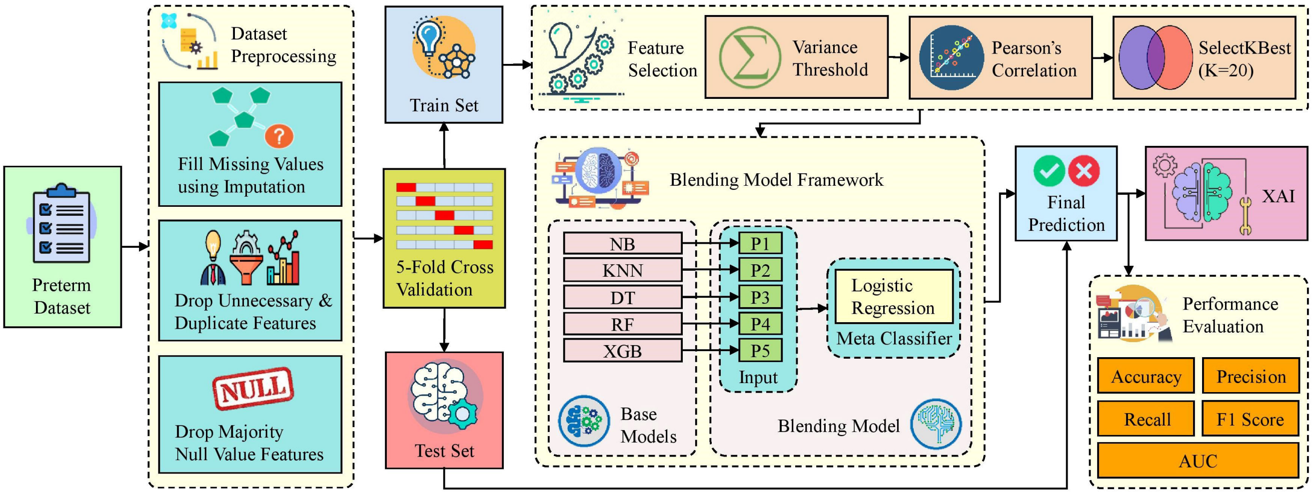
1) Applying multiple feature selection techniques (variance threshold, Pearson's correlation, and mutual information) can enhance the predictive power of ML models by identifying the most relevant factors.
2) Developing a BM with meta-learning can outperform standalone models in terms of prediction accuracy and robustness for preterm birth classification.
3) Integrating XAI methods will provide interpretable predictions and enable clinicians to trust and utilize ML results in real-world settings.

These hypotheses aim to address the lack of ensemble techniques, interpretability challenges, and limited datasets in existing studies. They guide the research methodology and are validated using rigorous experimental evaluation.

3. Proposed Methodology

The methodology of this study focuses on developing a ML framework for preterm birth prediction by using several key steps, as shown in Figure 2. First, the preterm dataset undergoes preprocessing, where missing values are filled through imputation, and unnecessary or duplicate features are removed to ensure data quality. After this, various feature selection techniques are applied to the training set, including variance threshold, Pearson's correlation, and mutual information (SelectKBest), to identify the most important features. The model development phase involves

Figure 2
Overview of the proposed research framework for preterm birth prediction



training and comparing multiple standalone classifiers, such as NB, k-nearest neighbors (KNN), DTs, RF, and xgboost, using 5-fold cross-validation. Consequently, a novel BM is constructed by combining all predictions from the standalone classifiers as its base models with a logistic regression meta-classifier. This approach aims to achieve better accuracy and consistency. The final proposed model is then evaluated using a range of performance metrics, including accuracy, precision, recall, $F1$ score, and AUC. Later, XAI is used to interpret the results. Each part of the proposed methodology is discussed in detail in the coming subsections. This research aims to enhance the prediction of preterm birth using advanced ML techniques. By focusing on improving model accuracy and interpretability, the study offers several key contributions to the field:

1) Apply multiple feature selection methods on training sets to enhance model performance.
2) Perform a comparative analysis of the performance of standalone classification models.
3) Develop a robust BM by combining standalone models to achieve superior results.
4) Evaluate the proposed model's performance using various metrics to ensure comprehensive assessment.
5) Incorporate XAI techniques to interpret and provide detailed explanations of the model's predictions.

Overall, this research study aims to develop an accurate predictive model for preterm birth. This research will help predict preterm birth early on to improve maternal and fetal health during and after the pregnancy. Additionally, explainability is critical for ML models in healthcare. Clinicians need to understand how and why a model makes certain predictions to trust its output. In this study, XAI techniques are used to provide transparency into the model's decision-making process. Specifically, feature importance and SHAP values are employed to make the predictions interpretable. Feature Importance ranks the input features based on their contributions to the model's predictions [32]. It highlights which maternal and environmental factors are most influential in predicting preterm birth. This allows healthcare professionals to focus on key factors that have a higher impact on the risk of preterm birth.

The SHAP values go one step further by explaining the contribution of each feature to the prediction of individual instances. SHAP values assign each feature a value that quantifies its contribution to increasing or decreasing the likelihood of preterm birth [33]. This interpretability is especially important for clinical decision-making, as it helps explain the rationale behind each prediction to patients and clinicians alike. SHAP values capture both linear and non-linear relationships between features, making them a powerful tool for understanding complex interactions in the data. In this study, SHAP values are used to provide a visual explanation of how the model reaches its predictions for individual cases. The use of SHAP and feature importance ensures that the model's decisions are transparent and interpretable, allowing clinicians to trust the predictions and integrate them into medical practice. This transparency is essential in building trust in AI-driven healthcare tools.

3.1. Proposed BM framework

The proposed BM framework is designed to integrate predictions from several base classifiers to generate a final, more accurate prediction. In this framework, five base models such as NB, KNNs, DT, RF, and extreme gradient boosting (XGB) are trained independently on the dataset to predict the binary classification of preterm. Each model's prediction is used as input into a higher-level meta-classifier, which in this case is logistic regression. The framework is designed to address the limitations of standalone models by leveraging the strengths of multiple algorithms.

3.1.1. NB

NB is a probabilistic classifier based on Bayes' theorem. It assumes that the features are independent given the class, which simplifies the calculations [34]. Despite this assumption, NB often performs well in real-world applications. Mathematically, it calculates the probability of each class label yyy given a feature set X = {x1,x2, . . . ,xn} using the following Equation (1):

$$P(y|X) = \frac{P(X|y)\ P(y)}{P(X)} \tag{1}$$

Here, $P(y|X)$ is the posterior probability of class y given the feature set X. $P(X|y)$ is the likelihood of the features given the class, $P(y)$ is the prior probability of the class, and $P(X)$ is the evidence, which can be ignored in classification since it remains constant across different classes. In the proposed blending framework, NB acts as one of the base learners and contributes its probabilistic predictions to the final ensemble.

3.1.2. KNN

KNN is a simple model. It is an instance-based learning algorithm that classifies a data point based on the majority class among its KNNs in the feature space [35]. No explicit training phase occurs in KNN, but during prediction, the distance between data points, which is often the Euclidean distance is calculated to find the nearest neighbors:

$$d(x, y) = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2} \tag{2}$$

where x and y represent two feature vectors, and $d(x, y)$ is the Euclidean distance between them. Once the KNNs are identified, the algorithm performs a majority vote to assign the class label. In the proposed framework, KNN helps capture local data patterns and contributes its predictions to the BM.

3.1.3. DT

DT is a non-parametric supervised learning method that splits the data into subsets based on the most significant features, creating j node, the algorithm chooses the feature that maximizes the Gini index or information gain values [36]. For binary classification, the information gain IG is calculated as:

$$IG(S, A) = H(S) - \sum_{c \in Values(A)} \frac{|S_v|}{|S|} H(S_v) \tag{3}$$

where the $H(S)$ is the entropy of the dataset S, and A is the feature being used to split the data. DT are easy to interpret and can model complex data patterns. In the blending framework, DT models complex interactions in the feature space, and its predictions are passed to the meta-classifier.

3.1.4. RF

RF is an ensemble of DTs where each tree is trained on a random subset of the data and features. The trees make individual predictions, and the final prediction is made by majority voting [37]. Mathematically, the RF prediction $\hat{y}$ is given by:

$$\hat{y} = mode(\hat{y}_1, \hat{y}_2, \ldots, \hat{y}_T) \tag{4}$$

where the $\hat{y}_1, \hat{y}_2, \ldots, \hat{y}_T$ are the predictions of individual trees, and T is the total number of trees. The RF helps to reduce overfitting, making it a powerful tool in the proposed framework by contributing stable and robust predictions to the meta-classifier.

3.1.5. XGB

The XGB is an advanced implementation of the gradient boosting algorithm. It builds an ensemble of trees sequentially. It focuses on correcting errors made by previous trees by minimizing a differentiable loss function, such as log loss in binary classification [38]. The objective function in XGB is:

$$Obj(\theta) = \sum_{i=1}^{n} l(y_i, \hat{y}_i) + \sum_{k=1}^{T} \Omega(f_k) \tag{5}$$

where the $l(y_i, \hat{y}_i)$ is the loss function. The $\Omega(f_k)$ represents the regularization term, and T is the number of trees. XGB is known for its speed and performance in handling structured data. In the blending framework, XGB provides high accuracy and captures complex patterns.

3.1.6. Logistic regression meta-classifier

Later the base models (NB, KNN, DT, RF, XGB) generate their predictions. These predictions are combined and fed into a logistic regression model, which acts as the meta-classifier. Logistic regression is widely used for binary classification problems [39]. It models the probability of the binary outcome using the logistic function:

$$P(y = 1|X) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_n X_n)}} \tag{6}$$

where $X_1, X_2, \ldots, X_n$ are the inputs (predictions from the base models), and $\beta_0, \beta_1, \ldots, \beta_n$ are the parameters learned during training. Logistic regression calculates the final prediction based on the predictions of the base models, effectively blending their outputs to produce a more reliable and accurate result.

3.1.7. BM

Let P_1, P_2, P_3, P_4, P_5 represent the predictions made by the base models NB, KNN, DT, RF, and XGB, respectively. The final prediction $\hat{y}$ from the BM using logistic regression meta-classifier is given by:

$$\hat{y} = \sigma(\beta_0 + \beta_1 P_1 + \beta_2 P_2 + \beta_3 P_3 + \beta_4 P_4 + \beta_5 P_5) \tag{7}$$

where the σ is the logistic function, and $\beta_0, \beta_{1,}\beta_{2,}\beta_{3,}\beta_{4,}\beta_5$ are the coefficients learned by the meta-classifier during training. The framework is designed to harness the strengths of each base model and to improve the model's generalizability. This framework optimizes the model's ability to handle diverse patterns in the data and improve the accuracy and robustness of preterm prediction.

4. Results

The dataset used in this study focuses on maternal factors related to preterm birth prediction [40]. The dataset is based on data collected from women in the State of Acre located in the Western Brazilian Amazon. The Ethics and Research Committees of the Federal University of São Paulo (UNIFESP) and Hospital São Paulo (#1772,931), as well as the Federal University of Acre (UFAC) (#1797,567), approved the study project. The authors collected the dataset through validated instruments and hospital records to ensure comprehensive and accurate data collection. This dataset contains approximately 800 instances, with a balanced distribution between preterm and full-term birth cases. It includes 78 features that capture a range of maternal and pregnancy-related information, such as demographic details, medical history, prenatal care, and various risk factors known to influence preterm birth outcomes. The features in this dataset provide a comprehensive representation of the factors impacting pregnancy outcomes in this particular region. The balanced nature of the dataset ensures that models trained on it do not favor one class over the other, which is crucial for reliable predictions. In-depth details about the dataset, including its collection process, feature definitions, and original objectives, are thoroughly discussed in its original research paper [41].

4.1. Dataset preprocessing

The data preprocessing stage involves handling the missing values, eliminating irrelevant or redundant features, and conducting a thorough analysis of the dataset. Several features are included in the dataset that could only be obtained after the end of the pregnancy, such as birth weight, Apgar scores at 1 and 5 min, delivery type, gestational age, and cause of prematurity. These features are removed to ensure that the model only relies on the information available before delivery. Additionally, the non-informative features, such as record identification numbers, and any duplicate entries are discarded to improve the dataset's quality.

The dataset also contains a significant number of null values, and removing all instances with missing data will reduce the dataset by half. Therefore, further analysis is conducted to handle the missing values more effectively. For instance, the feature "previous cesarian birth" has the most null entries. Upon closer inspection, it is determined that these missing values likely represent the mothers who had not experienced any prior pregnancies. This is confirmed by comparing the "premature child previous" feature with the "previous cesarian birth" feature, which allowed the missing values to be filled with a new category indicating no prior pregnancies. For other features like "BMI" and "household income", the missing values are filled using the mean of the existing values. Features with excessive missing data that could not be reasonably imputed are removed. After completing this preprocessing phase, the dataset is reduced to 696 instances with 61 features.

4.2. Training and testing data

The authors use 5-fold cross-validation to train and test the models in this study. In 5-fold cross-validation, the dataset is split into five equal parts, or folds [42]. Each fold takes a turn being the test set while the remaining four folds serve as the training set. This process repeats five times, ensuring that every instance in the dataset is tested once [43]. The results are then averaged to provide a more robust measure of model performance [44]. This technique helps to reduce the risk of overfitting [45], a common issue in small datasets. Feature selection methods are only applied to the training folds to avoid data leakage. This ensures that information from the test set does not influence the model during training, which would otherwise lead to overly optimistic performance results and potentially biased predictions. By applying feature selection exclusively to the training data, the authors maintain the integrity of the validation process and ensure that the model generalizes well to unseen data.

4.3. Feature selection

Feature selection plays a crucial role in improving model performance [46]. It improves the model's predictive power by identifying the most important features while reducing redundancy and noise in the data. In this study, various feature selection techniques are applied to the training set to ensure that the model only learns from relevant and significant information. This process helps in preventing overfitting and enhances the model's predictive ability. The upcoming subsections will discuss the feature selection methods used in detail.

4.3.1. Variance threshold

The variance threshold method is a simple yet effective feature selection technique that removes features with low variance [47]. This method operates on the principle that features with little variation across data points do not contribute significantly to the model's predictions. Features that are mostly constant or have little fluctuation contain minimal information, as they do not help in distinguishing between different classes. In the context of ML, variance is a measure of how much a feature's values spread from the mean. If this variance is below a set threshold, the feature is considered uninformative and can be removed from the dataset [48]. Mathematically, the variance for a given feature X is calculated as:

$$Var(X) = \frac{1}{n}\sum_{i=1}^{n}(x_i - \mu)^2 \tag{8}$$

In the Equation (8), x_i represents each data point, μ is the mean of the feature, and n is the number of instances. If the variance of a feature falls below a certain threshold (usually close to zero), it indicates that the feature exhibits high homogeneity and does not provide significant insight for the model [49]. In this study, all features with high homogeneity, or those that showed little to no variance are removed from the dataset. It ensures that the model focuses only on features with meaningful differences across instances. This helps streamline the dataset and improves both model training efficiency and prediction accuracy by eliminating irrelevant data.

4.3.2. Pearson's correlation

Pearson's correlation is a widely used statistical method for measuring the linear relationship between two features. It quantifies how strongly two variables are related by calculating a correlation coefficient that ranges from −1 to 1 [50]. A value of 1 indicates a perfect positive correlation, where both variables increase together, while −1 represents a perfect negative correlation, meaning one variable increases as the other decreases. A value of 0 suggests no linear relationship between the two variables. In feature selection, highly correlated features can introduce redundancy, as they essentially carry the same information, which can lead to overfitting and reduce model performance [51]. The Pearson correlation coefficient between two features X and Y is computed as:

$$r_{xy} = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2}\sqrt{\sum(y_i - \bar{y})^2}} \tag{9}$$

In the Equation (9), $\mathrm{x_i}$ and $\mathrm{y_i}$ are the individual data points, and $\bar{\mathrm{x}}$ and $\bar{\mathrm{y}}$ are the means of the respective features. The result $\mathrm{r_{xy}}$ indicates the strength and direction of the linear relationship between the two features. In this study, the authors set a Pearson correlation coefficient threshold of 0.7, meaning that any pair of features with a correlation higher than 0.7 is considered too closely related. The correlated features are removed to reduce redundancy and ensure the model focuses on distinct and relevant information. This threshold helps in simplifying the model without losing important predictive power, allowing the remaining features to provide more insights that are meaningful.

4.3.3. Mutual information (SelectKBest)

The mutual information method is specifically used with the SelectKBest function. It is a feature selection technique that measures the mutual dependence between two variables [52]. In ML, mutual information quantifies how much information the presence of one feature provides about the label or target variable. Unlike methods that measure linear relationships, mutual information can capture both linear and non-linear associations which makes it particularly useful for complex datasets [53]. The higher the mutual information score, the more relevant the feature is in predicting the target. Mathematically, mutual information between two variables X and Y is given by:

$$I(X;Y) = \sum_{x\in X}\sum_{y\in Y} p(x,y)\log\left(\frac{p(x,y)}{p(x)p(y)}\right) \tag{10}$$

where $\mathrm{p(x,y)}$is the joint probability distribution of X and Y. The $\mathrm{p(x)}$ and $\mathrm{p(y)}$ are the marginal probabilities of X and Y, respectively.

In this study, the authors calculate the mutual information of each feature relative to the target label and retained the top 18 features with the highest mutual information scores. This selection process ensures that only the most informative features are kept, improving the efficiency and accuracy of the model.

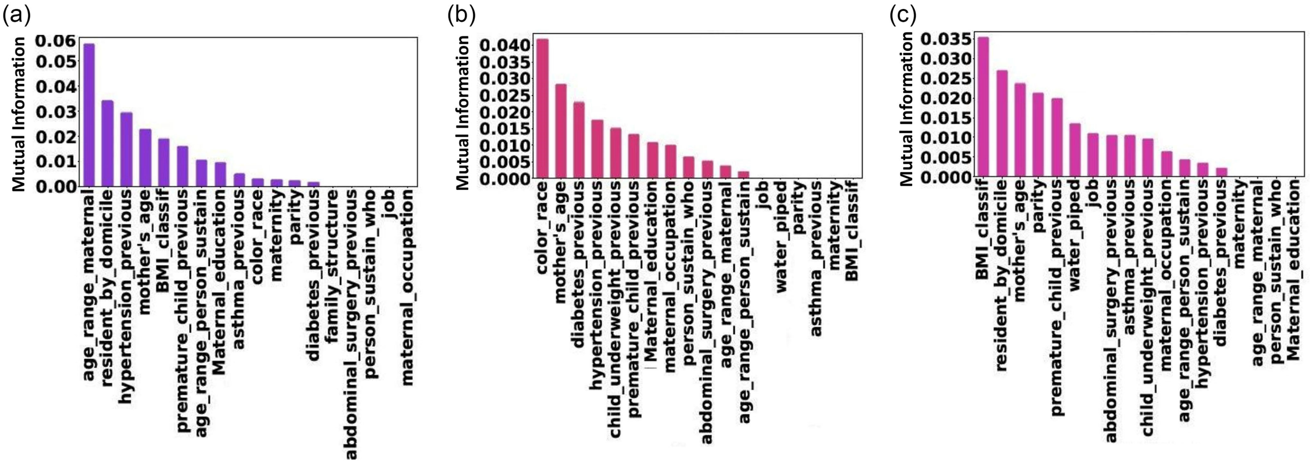
Figure 3 illustrates the 18 features selected after applying the mutual information method, with the x-axis showing the features and the y-axis representing their mutual information scores. Interestingly, even features with a mutual information score of zero were not always removed, as mutual information does not capture all complex relationships within the data. Therefore, additional features were randomly selected to ensure that exactly 18 features were used in training the model across all folds. The final set of 18 features after this selection process are then used to train the models.

4.4. Model performance evaluation

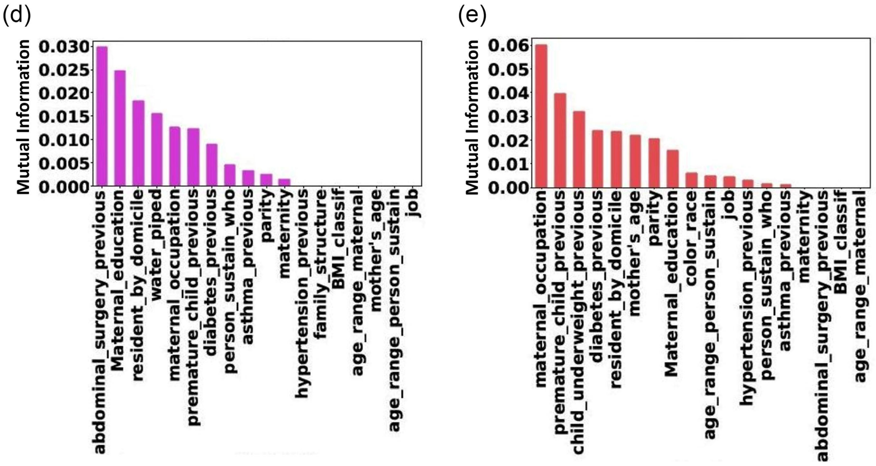
The authors in this study evaluate the proposed model's performance by employing various metrics to measure its classification accuracy and interpretability. Performance metrics provide a quantitative understanding of the model's predictive capability, while XAI methods offer interpretability into how the model makes predictions. To effectively evaluate the classification performance of the model, several metrics are employed. Each of the metric provides unique insights into the model's ability to predict preterm births. The metrics consider the relationships between true positives, true negatives, false positives, and false negatives, which are critical in medical classification tasks. The performance metrics are summarized in Table 1.

The performance of the proposed BM is compared with standalone models using both 5-fold cross-validation and holdout testing. The results indicate the effectiveness of the BM in improving prediction accuracy for preterm birth classification. The models are initially evaluated using 5-fold cross-validation, and results are given in Table 2. Among all standalone models, XGB and KNN perform better than other base classifiers. XGB shows the highest recall of 71.28% and $F1$ score of 70.41%, while KNN achieves higher accuracy of 71.14% but relatively lower precision and recall compared to XGB. The proposed BM consistently outperforms all base models across all metrics, with an accuracy of 74.61%, a precision of 71.83%, a recall of 70.94%, an $F1$ score of 71.38%, and an AUC of 84.56%. This demonstrates that the blended approach provides more reliable and balanced results across different metrics than any standalone model.

Figure 3
Mutual information scores of the top 18 selected features for preterm prediction across different folds



Features for fold 1 Features for fold 2 Features for fold 3



Features for fold 4 Features for fold 5

Table 1
Performance evaluation metrics and their mathematical descriptions for preterm prediction

Metric	Description	Equation
Accuracy	Measures the proportion of correctly classified instances (both preterm and non-preterm) out of the total instances.	$\frac{TP + TN}{TP + TN + FP + FN}$
Precision	The ratio of true positive predictions to the total predicted positive instances. Precision is critical in reducing false positives, which could mislead clinical interventions.	$\frac{TP}{TP + FP}$
Recall	The ratio of true positive predictions to total actual positive instances. High recall means model identifies the majority of true preterm cases to minimize false negatives.	$\frac{TP}{TP + FN}$
F1 Score	The harmonic mean of precision and recall. The $F1$ score balances both false positives and false negatives, making it useful in preterm prediction.	$\frac{2 \times Precision\ mesRecall}{Precision + Recall}$
AUC	Represents the model's ability to distinguish between classes. It provides a broader measure of the model's performance.	TP and FP rate

Table 2
5-fold cross-validation results for preterm prediction

Models	Accuracy	Precision	Recall	$F1$ score	AUC
DT	65.47%	63.61%	43.08%	51.37%	69.74%
KNN	71.14%	56.09%	47.29%	51.32%	70.39%
NB	65.13%	66.09%	45.71%	54.04%	71.01%
RF	68.02%	67.74%	49.86%	57.44%	74.97%
XGB	72.92%	69.57%	71.28%	70.41%	79.38%
BM	74.61%	71.83%	70.94%	71.38%	84.56%

Table 3
Holdout (80/20) testing results of standalone models and the blending model

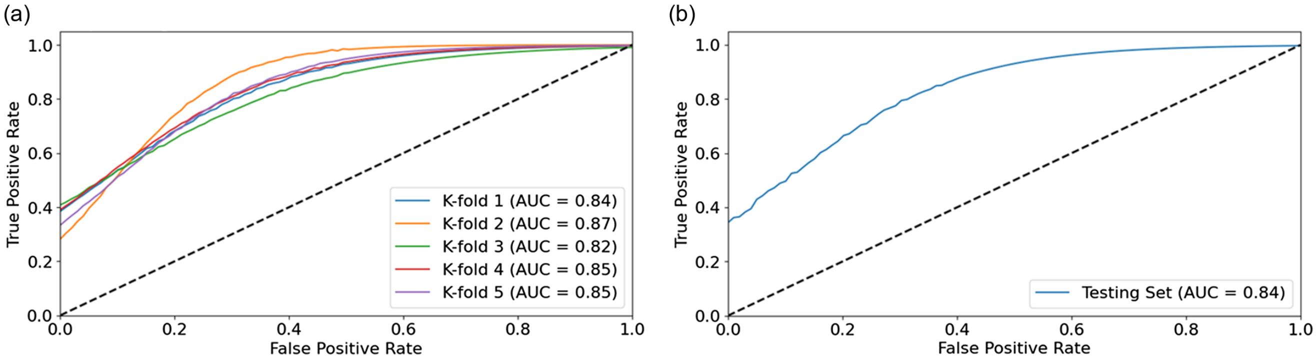
Models	Accuracy	Precision	Recall	$F1$ score	AUC
DT	61.36%	62.17%	41.46%	49.75%	69.01%
KNN	59.74%	49.65%	45.43%	47.45%	69.36%
NB	54.79%	65.12%	44.97%	53.21%	69.97%
RF	65.34%	62.93%	47.38%	54.06%	70.42%
XGB	70.74%	67.43%	73.33%	70.26%	78.35%
BM	73.82%	72.46%	69.74%	71.07%	83.91%

The models are further evaluated using an 80/20 train-test split to assess generalization performance. As shown in Table 3, the BM continues to outperform the base models with the highest accuracy of 73.82%, precision of 72.46%, recall of 69.74%, and $F1$ score of 71.07%. The AUC for the BM is also superior at 83.91%, indicating that the model provides a strong balance between sensitivity and specificity as shown in Figure 4. Among the base models, XGB performs best in terms of recall of 73.33% and $F1$ score 70.26%, but still lags behind the proposed BM in terms of overall performance. The results demonstrate that the proposed BM provides more robust and consistent performance across both cross-validation and holdout testing. These results highlight its potential for improving preterm birth prediction.

In terms of comparative results, this study stands out to apply predictive modeling alongside a previous study [54] on the same dataset for preterm birth classification. The referenced study employs an 80/20 holdout training and testing split and utilizes information gained for feature selection before implementing a RF model. The proposed BM undergoes the same evaluation using the same 80/20 holdout technique but applies three distinct feature selection methods to enhance its predictive capabilities. The comparative performance of the models is shown in Table 4. Although the RF model from the prior study achieves a higher accuracy, the BM outperforms it in AUC, which indicates stronger balance and reliability in differentiating between preterm and non-preterm cases. This higher AUC underscores the proposed BM's potential to provide more consistent results across varying preterm risk factors.

Figures 5 and 6 display SHAP summary plots for both normal and preterm predictions, respectively. These visualizations rank features based on their importance and the impact they have on the prediction outcome. In both cases, the feature "abnormal

Figure 4
5-fold cross-validation and holdout (80/20) testing AUC plots



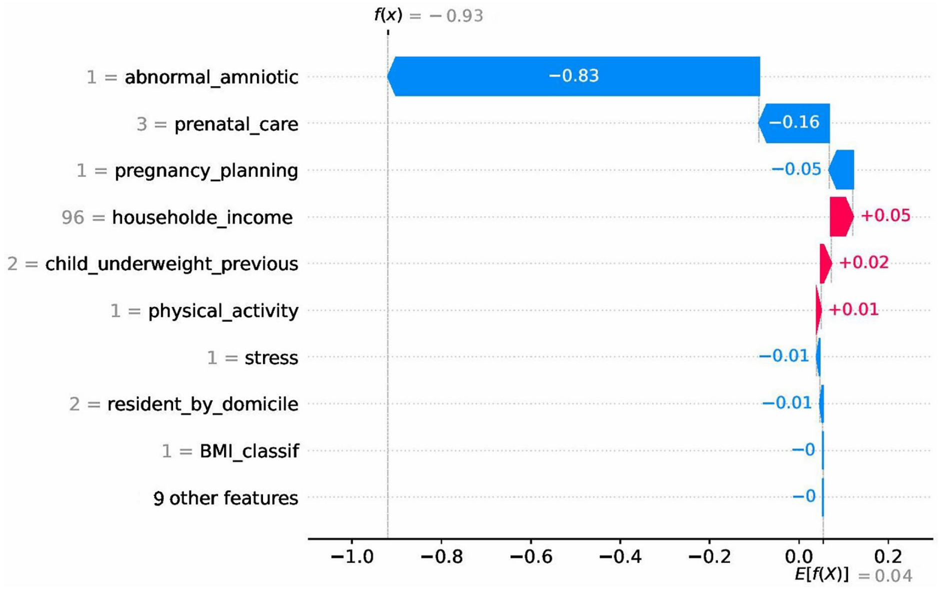
5-fold cross-validation AUC plot

Holdout (80/20) testing AUC plot

Table 4
Comparative results analysis of studies done on the same preterm prediction dataset

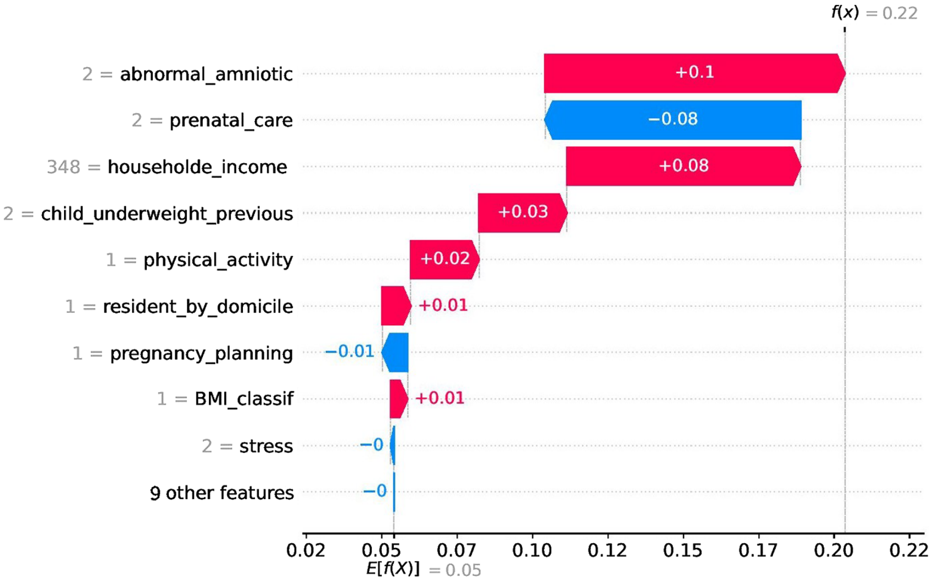
Models	Accuracy	Precision	Recall	$F1$ score	AUC
Random Forest [54]	79.25%	72.00%	68.00%	70.00%	80.00%
Proposed Blending Model	73.82%	72.46%	69.74%	71.07%	83.91%

Figure 5
SHAP summary plot showing key features contributing to the normal birth prediction



amniotic" emerges as the most influential. It has a very high SHAP value indicating its significant contribution towards predicting preterm birth. Figure 5 highlights features contributing to the prediction of a normal birth (non-preterm). The "abnormal amniotic" feature is the most impactful feature, but in this case, it has a strongly negative SHAP value of −0.83, suggesting that its absence is crucial for a normal birth prediction. The "prenatal care" feature also has a significant negative value of −0.16, reinforcing the idea that good prenatal care lowers the likelihood of preterm birth. Other features like "pregnancy planning", "household income", and "child underweight previous" have smaller but notable positive contributions to the normal birth prediction. This indicates that economic stability, child health history, and pregnancy planning are also important factors for a healthy pregnancy outcome. These SHAP-based insights validate the model's reliance on medically significant features. They also

Figure 6
SHAP summary plot showing the most important features contributing to the preterm birth prediction



demonstrate its ability to align with clinical knowledge. This reinforces the rationale for feature selection and enhances the model's interpretability.

The "abnormal amniotic" feature has the most significant positive impact, indicated by a SHAP value of +0.10, meaning that abnormal amniotic fluid strongly contributes to the model predicting a preterm birth. Similarly, "household income" feature also has a substantial positive SHAP value of +0.08, showing that economic factors play a role in preterm birth predictions. On the other hand, "prenatal care" feature has a negative SHAP value of −0.08, suggesting that better prenatal care reduces the likelihood of predicting preterm birth. Other features like "child underweight previous" and "physical activity" contribute more modestly but still positively affect the prediction outcome. These insights align with medical research showing that lack of prenatal care and abnormal amniotic conditions are critical risk factors for preterm birth. By leveraging SHAP values, the model provides a transparent explanation of how key features influence preterm predictions. This bridges the gap between algorithmic decisions and clinical interpretability. It highlights the importance of explainability techniques in validating and refining feature selection for real-world applications.

The bar plot shown in Figure 7 ranks the features based on their overall contribution to the model's predictions. The "abnormal

Figure 7
Overall feature importance bar plot for preterm prediction

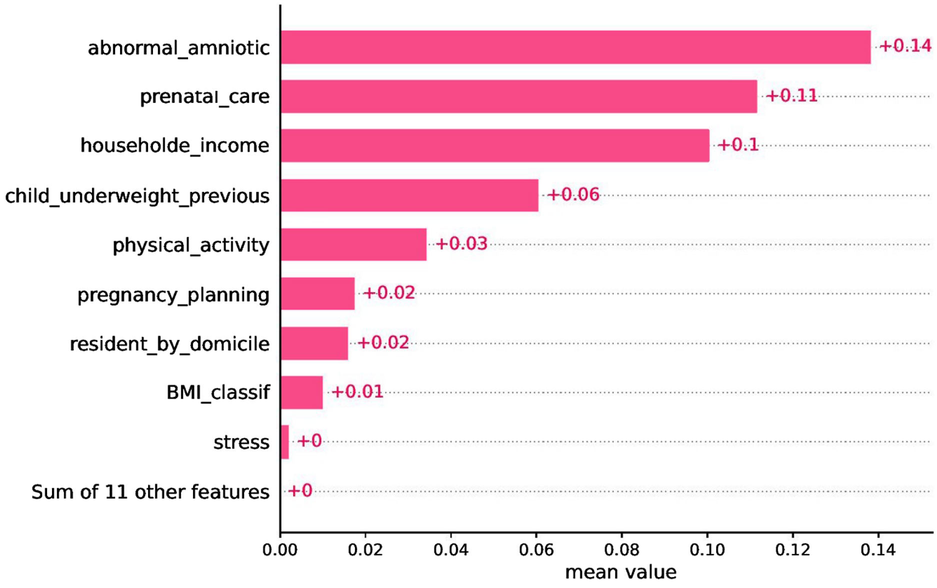
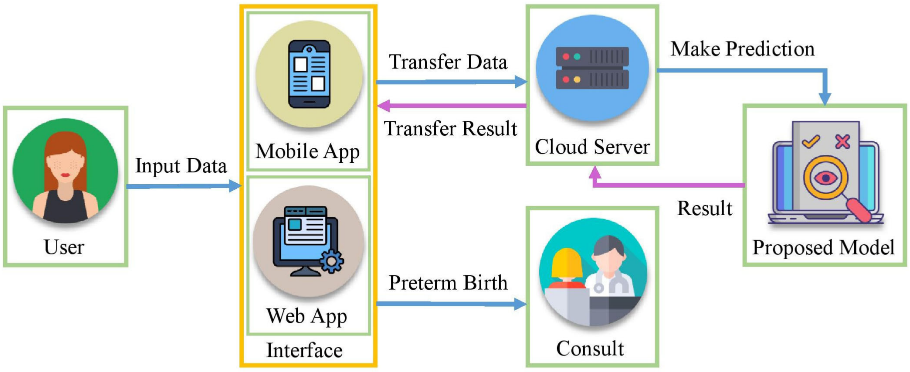


Figure 8
Suggested real-world application of the proposed model



amniotic" feature has the highest mean SHAP value of +0.14, making it the most critical feature for predicting preterm birth. Following closely, features are "prenatal care" and "household income" with values of +0.11, and +0.10, respectively. Both of the features are vital socio-economic and medical factors influencing the outcome. Other features like "child underweight previous" and "physical activity" contribute more modestly, but they still play a role in the model's accuracy. The SHAP visualizations show that a combination of medical and socio-economic features significantly impacts the model's ability to predict preterm births. Abnormal amniotic fluid and prenatal care are consistently the most influential features across both positive and negative predictions. These results emphasize the importance of early intervention in medical care and socio-economic support to mitigate the risk of preterm births. The integration of explainability techniques like SHAP and LIME ensures the feature selection process is both data-driven and clinically meaningful. This enhances the model's reliability and supports its adoption in healthcare decision-making.

4.5. Discussion

The results of this study provide valuable insights into predicting preterm birth using ML models, with a focus on understanding feature importance through XAI techniques. Key findings from the SHAP plots reveal that medical and socio-economic factors, like abnormal amniotic fluid, prenatal care, and household income, have a substantial impact on predicting both preterm and normal birth outcomes. For instance, abnormal amniotic fluid is consistently identified as the most critical predictor that positively influencing preterm birth risk while contributing negatively to the prediction of a normal birth outcome. The analysis also highlights that prenatal care plays a dual role; when absent or insufficient, it increases the likelihood of preterm birth, whereas its presence contributes to a healthier pregnancy outcome. Interestingly, household income emerges as another influential factor, with lower income being associated with higher risks of preterm births. Other features like physical activity and child's underweight history also contribute with comparatively smaller impacts.

The proposed preterm prediction model can be integrated into a practical system involving a mobile app and a web-based interface as shown in Figure 8. Users, such as pregnant women or healthcare providers, can input relevant medical and socio-economic data via these platforms. These data are then transferred to a cloud server, where the prediction model processes the input and makes real-time predictions regarding the likelihood of a preterm birth. Upon generating a prediction, the results are sent back to the app or interface, where they can be reviewed by the user. Studies have highlighted the importance of integrating predictive models into mobile health platforms to provide timely interventions and improve health outcomes, particularly in resource-limited settings [55, 56]. In the event of a preterm birth prediction, the system also facilitates consultation with healthcare professionals, providing a direct link between the model's output and practical medical advice. This two-way communication ensures that predictions lead to actionable outcomes. This real-world application of the proposed preterm prediction model can enhance the effectiveness of early interventions and allow for better monitoring of high-risk pregnancies.

5. Conclusion

Preterm birth is defined as delivery before 37 weeks of gestation. It is a serious public health concern that can lead to severe health complications for newborns. These include respiratory distress, developmental delays, and long-term neurological deficits. For mothers, preterm births can lead to medical complications and emotional distress. Due to these risks, predicting preterm births early is critical. Early prediction can help for timely medical interventions that can mitigate harm and improve the chances of survival for the infant. In this research, authors have developed a ML model using a blend of various classification techniques to predict preterm birth. The proposed model consistently outperformed standalone models in both 5-fold cross-validation and holdout testing. Specifically, the proposed model reached 74.61% accuracy with an AUC of 84.56% in cross-validation and 73.82% accuracy with an AUC of 83.91% in holdout testing. The XAI results show that significant factor influencing preterm risk is abnormal amniotic fluid and prenatal care. In the normal birth case, the positive influence of abnormal amniotic conditions and household income is highlighted. However, in preterm cases, abnormal amniotic fluid, prenatal care quality, and pregnancy planning emerged as the leading indicators.

Despite the promising results, this research has certain limitations. The study is based on a single dataset. This may limit the model's applicability to other populations with different socio-demographic or medical profiles. The dataset size is also relatively small compared to those used in some large-scale studies. Additionally, the proposed model has not yet been tested in real-world clinical settings. Implementation challenges such as data variability and user interaction could influence its performance. The limitations of this study provide a foundation for the future work. The future work can include expanding the dataset to include diverse populations from different regions which would enhance the model's generalizability. Additionally, incorporating more advanced features such as genetic or environmental factors could further refine predictive performance. Testing the model in clinical environments will also be crucial for assessing its practical utility and effectiveness in real-world scenarios.

Acknowledgement

The authors express their gratitude to all the anonymous reviewers for their insightful comments and suggestions that contributed to the enhancement of this research paper.

Ethical Statement

This study does not contain any studies with human or animal subjects performed by any of the authors.

Conflicts of Interest

The authors declare that they have no conflicts of interest to this work.

Data Availability Statement

The data that support the findings of this study are openly available in Midwifery at https://doi.org/10.1016/j.midw.2020.102670, reference number [40].

Author Contribution Statement

Ahmad Hassan: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Resources, Writing – original draft, Writing – review & editing, Visualization, Supervision, Project administration. **Saba Nawaz:** Validation, Formal analysis, Investigation, Resources, Writing – original draft, Writing – review & editing, Project administration. **Sidra Tahira:** Validation, Investigation, Data curation, Writing – review & editing, Visualization, Project administration. **Arslan Ahmed:** Investigation, Data curation, Writing – review & editing, Supervision, Project administration.

References

[1] Szczuko, M., Palma, J., Kikut, J., Komorniak, N., & Ziętek, M. (2020). Changes of lipoxin levels during pregnancy and the monthly-cycle, condition the normal course of pregnancy or pathology. *Inflammation Research*, *69*(9), 869–881. https://doi.org/10.1007/s00011-020-01358-6

[2] Hua, J., Barnett, A. L., Lin, Y., Guan, H., Sun, Y., Williams, G. J., . . . , & Du, W. (2022). Association of gestational age at birth with subsequent neurodevelopment in early childhood: A national retrospective cohort study in China. *Frontiers in Pediatrics*, *10*, 860192. https://doi.org/10.3389/fped.2022.860192

[3] Dong, D., Ru, X., Huang, X., Sang, T., Li, S., Wang, Y., & Feng, Q. (2022). A prospective cohort study on lactation status and breastfeeding challenges in mothers giving birth to preterm infants. *International Breastfeeding Journal*, *17*(1), 6. https://doi.org/10.1186/s13006-021-00447-4

[4] Ji, X., Wu, C., Chen, M., Wu, L., Li, T., Miao, Z., . . . , & Ding, H. (2022). Analysis of risk factors related to extremely and very preterm birth: A retrospective study. *BMC Pregnancy and Childbirth*, *22*(1), 818. https://doi.org/10.1186/s12884-022-05119-7

[5] Montemor, M. S., Demarque, G. F., Rodrigues, A. S., Francisco, R. P. V., & de Carvalho, M. H. B. (2022). Association between preterm births and socioeconomic development: Analysis of national data. *BMC Public Health*, *22*(1), 2014. https://doi.org/10.1186/s12889-022-14376-2

[6] Ohuma, E. O., Moller, A. B., Bradley, E., Chakwera, S., Hussain-Alkhateeb, L., Lewin, A., . . . , & Moran, A. C. (2023). National, regional, and global estimates of preterm birth in 2020, with trends from 2010: A systematic analysis. *The Lancet*, *402*(10409), 1261–1271. https://doi.org/10.1016/S0140-6736(23)00878-4

[7] Liu, T., Xu, Y., Gong, Y., Zheng, J., & Chen, Z. (2024). The global burden of disease attributable to preterm birth and low birth weight in 204 countries and territories from 1990 to 2019: An analysis of the Global Burden of Disease Study. *Journal of Global Health*, *14*, 04109. https://doi.org/10.7189/jogh.14.04109

[8] Ayele, T. B., & Moyehodie, Y. A. (2023). Prevalence of preterm birth and associated factors among mothers who gave birth in public hospitals of east Gojjam zone, Ethiopia. *BMC Pregnancy and Childbirth*, *23*(1), 204. https://doi.org/10.1186/s12884-023-05517-5

[9] Ukhawounam, U., Limruangrong, P., Pungbangkadee, R., & Vongsirimas, N. (2023). Effects of education and guided imagery program on stress level and coping behaviors among pregnant women at risk of preterm birth. *International Journal of Women's Health*, *15*, 1581–1591. https://doi.org/10.2147/IJWH.S418693

[10] Zivaljevic, J., Jovandaric, M. Z., Babic, S., & Raus, M. (2024). Complications of preterm birth—The importance of care for the outcome: A narrative review. *Medicina*, *60*(6), 1014. https://doi.org/10.3390/medicina60061014

[11] Girchenko, P., Robinson, R., Rantalainen, V. J., Lahti-Pulkkinen, M., Heinonen-Tuomaala, K., Lemola, S., . . . , & Räikkönen, K. (2022). Maternal postpartum depressive symptoms partially mediate the association between preterm birth and mental and behavioral disorders in children. *Scientific Reports*, *12*(1), 947. https://doi.org/10.1038/s41598-022-04990-w

[12] Akalay, S., Rayyan, M., Fidlers, T., van den Heuvel, L., Levtchenko, E., & Arcolino, F. O. (2024). Impact of preterm birth on kidney health and development. *Frontiers in Medicine*, *11*, 1363097. https://doi.org/10.3389/fmed.2024.1363097

[13] Nava, C., di Gallo, A., Biuso, A., Daniele, I., Lista, G., Comberiati, P., . . . , & D'Auria, E. (2023). Early-life nutrition in preterm infants and risk of respiratory infections and wheezing: A scoping review. *Nutrients*, *15*(13), 3031. https://doi.org/10.3390/nu15133031

[14] Singh, H., Cho, S. J., Gupta, S., Kaur, R., Sunidhi, S., Saluja, S., . . . , & Sun, Y. (2021). Designing a bed-side system for predicting length of stay in a neonatal intensive care unit. *Scientific Reports*, *11*(1), 3342. https://doi.org/10.1038/s41598-021-82957-z

[15] Rysavy, M. A., Colaizy, T. T., Bann, C. M., DeMauro, S. B., Duncan, A. F., Brumbaugh, J. E., . . . , & Bell, E. F. (2021). The

relationship of neurodevelopmental impairment to concurrent early childhood outcomes of extremely preterm infants. *Journal of Perinatology*, *41*(9), 2270–2278. https://doi.org/10.1038/s41372-021-00999-7

[16] Sixtus, R. P., Dyson, R. M., & Gray, C. L. (2024). Impact of prematurity on lifelong cardiovascular health: Structural and functional considerations. *npj Cardiovascular Health*, *1*(1), 2. https://doi.org/10.1038/s44325-024-00002-0

[17] Kovács, G., Abonyi-Tóth, Z., Fadgyas-Freyler, P., & Kaló, Z. (2023). Incremental cost of premature birth: A public health care payer perspective from Hungary. *BMC Health Services Research*, *23*(1), 686. https://doi.org/10.1186/s12913-023-09697-w

[18] Zhang, Y., Du, S., Hu, T., Xu, S., Lu, H., Xu, C., . . . , & Zhu, X. (2023). Establishment of a model for predicting preterm birth based on the machine learning algorithm. *BMC Pregnancy and Childbirth*, *23*(1), 779. https://doi.org/10.1186/s12884-023-06058-7

[19] Zhou, X., Cai, F., Li, S., Li, G., Zhang, C., Xie, J., & Yang, Y. (2024). Machine learning techniques for prediction in pregnancy complicated by autoimmune rheumatic diseases: Applications and challenges. *International Immunopharmacology*, *134*, 112238. https://doi.org/10.1016/j.intimp.2024.112238

[20] Karalis, V. D. (2024). The integration of artificial intelligence into clinical practice. *Applied Biosciences*, *3*(1), 14–44. https://doi.org/10.3390/applbiosci3010002

[21] Nasarian, E., Alizadehsani, R., Acharya, U. R., & Tsui, K. L. (2024). Designing interpretable ML system to enhance trust in healthcare: A systematic review to proposed responsible clinician-AI-collaboration framework. *Information Fusion*, *108*, 102412. https://doi.org/10.1016/j.inffus.2024.102412

[22] Metta, C., Beretta, A., Pellungrini, R., Rinzivillo, S., & Giannotti, F. (2024). Towards transparent healthcare: Advancing local explanation methods in explainable artificial intelligence. *Bioengineering*, *11*(4), 369. https://doi.org/10.3390/bioengineering11040369

[23] Sun, Q., Zou, X., Yan, Y., Zhang, H., Wang, S., Gao, Y., . . . , & Ma, X. (2022). Machine learning-based prediction model of preterm birth using electronic health record. *Journal of Healthcare Engineering*, *2022*(1), 9635526. https://doi.org/10.1155/2022/9635526

[24] Synan, L., Ghazvini, S., Uthaman, S., Cutshaw, G., Lee, C. Y., Waite, J., . . . , & Bardhan, R. (2023). First trimester prediction of preterm birth in patient plasma with machine-learning-guided Raman spectroscopy and metabolomics. *ACS Applied Materials & Interfaces*, *15*(32), 38185–38200. https://doi.org/10.1021/acsami.3c04260

[25] Borboa-Olivares, H., Rodríguez-Sibaja, M. J., Espejel-Nuñez, A., Flores-Pliego, A., Mendoza-Ortega, J., Camacho-Arroyo, I., . . . , & Estrada-Gutierrez, G. (2023). A novel predictive machine learning model integrating cytokines in cervical-vaginal mucus increases the prediction rate for preterm birth. *International Journal of Molecular Sciences*, *24*(18), 13851. https://doi.org/10.3390/ijms241813851

[26] Hong, Y. M., Lee, J., Cho, D. H., Jeon, J. H., Kang, J., Kim, M. G., . . . , & Kim, J. K. (2023). Predicting preterm birth using machine learning techniques in oral microbiome. *Scientific Reports*, *13*(1), 21105. https://doi.org/10.1038/s41598-023-48466-x

[27] Khan, W., Zaki, N., Ghenimi, N., Ahmad, A., Bian, J., Masud, M. M., . . . , & Ahmed, L. A. (2023). Predicting preterm birth using explainable machine learning in a prospective cohort of nulliparous and multiparous pregnant women. *PLoS ONE*, *18*(12), e0293925. https://doi.org/10.1371/journal.pone.0293925

[28] Kavitha, S. N., & Asha, V. (2024). Predicting risk factors associated with preterm delivery using a machine learning model. *Multimedia Tools and Applications*, *83*(30), 74255–74280. https://doi.org/10.1007/s11042-024-18332-7

[29] Yu, Q. Y., Lin, Y., Zhou, Y. R., Yang, X. J., & Hemelaar, J. (2024). Predicting risk of preterm birth in singleton pregnancies using machine learning algorithms. *Frontiers in Big Data*, *7*, 1291196. https://doi.org/10.3389/fdata.2024.1291196

[30] Zhang, F., Tong, L., Shi, C., Zuo, R., Wang, L., & Wang, Y. (2024). Deep learning in predicting preterm birth: A comparative study of machine learning algorithms. *Maternal-Fetal Medicine*, *6*(3), 141–146. https://doi.org/10.1097/FM9.0000000000000236

[31] Seamon, E., Mattera, J. A., Keim, S. A., Leerkes, E. M., Rennels, J. L., Kayl, A. J., . . . , & Gartstein, M. A. (2024). Leveraging machine learning to study how temperament scores predict pre-term birth status. *Global Pediatrics*, *9*, 100220. https://doi.org/10.1016/j.gpeds.2024.100220

[32] Wang, H., Liang, Q., Hancock, J. T., & Khoshgoftaar, T. M. (2024). Feature selection strategies: A comparative analysis of SHAP-value and importance-based methods. *Journal of Big Data*, *11*(1), 44. https://doi.org/10.1186/s40537-024-00905-w

[33] Sebastián, C., & González-Guillén, C. E. (2024). A feature selection method based on Shapley values robust for concept shift in regression. *Neural Computing and Applications*, *36*(23), 14575–14597. https://doi.org/10.1007/s00521-024-09745-4

[34] Abdulameer, A. G., Hammood, A. S., Abdulwahed, F. M., & Ayyash, A. A. (2024). Naïve Bayes algorithm for timely fault diagnosis in helical gear transmissions using vibration signal analysis. *International Journal on Interactive Design and Manufacturing*, *18*(6), 1–12. https://doi.org/10.1007/s12008-024-02037-0

[35] Halder, R. K., Uddin, M. N., Uddin, M. A., Aryal, S., & Khraisat, A. (2024). Enhancing k-nearest neighbor algorithm: A comprehensive review and performance analysis of modifications. *Journal of Big Data*, *11*(1), 113. https://doi.org/10.1186/s40537-024-00973-y

[36] Mienye, I. D., & Jere, N. (2024). A survey of decision trees: Concepts, algorithms, and applications. *IEEE Access*, *12*, 86716–86727. https://doi.org/10.1109/ACCESS.2024.3416838

[37] Ali Linkon, A., Rahman Noman, I., Rashedul Islam, M., Chakra Bortty, J., Kumar Bishnu, K., Islam, A., . . . , & Abdullah, M. (2024). Evaluation of feature transformation and machine learning models on early detection of diabetes mellitus. *IEEE Access*, *12*, 165425–165440. https://doi.org/10.1109/ACCESS.2024.3488743

[38] Bhandari, V., Londhe, N. D., & Kshirsagar, G. B. (2024). Comparative evaluation of facial paradigms for Devanagari script-based P300 speller: Preliminary findings. *Journal of Data Science and Intelligent Systems*. Advance online publication. https://doi.org/10.47852/bonviewJDSIS42021463

[39] Bala, I., Kelly, T. L., Lim, R., Gillam, M. H., & Mitchell, L. (2024). An effective approach for multiclass classification of adverse events using machine learning. *Journal of Computational and Cognitive Engineering*, *3*(3), 226–239. https://doi.org/10.47852/bonviewJCCE32021924

[40] Santos, C., Costa, K., Dourado, J., Lima, S., Dotto, L., & Schirmer, J. (2020). Data for: Maternal factors associated with prematurity in public maternity hospital the Brazilian Western Amazon. *Mendeley Data*. https://doi.org/10.17632/ypj6f5bjc2.1

[41] Santos, C. L., Costa, K. M. M., Dourado, J. E. C., Lima, S. B. G., Dotto, L. M. G., & Schirmer, J. (2020). Maternal factors associated with prematurity in public maternity hospitals at the Brazilian Western Amazon. *Midwifery*, *85*, 102670. https://doi.org/10.1016/j.midw.2020.102670

[42] Wang, G., Li, C., Tang, F., Wang, Y., Wu, S., Zhi, H., . . . , & Zhang, J. (2023). A fully-automatic semi-supervised deep learning model for difficult airway assessment. *Heliyon*, *9*(5), e15629. https://doi.org/10.1016/j.heliyon.2023.e15629

[43] Hassan, A., Gulzar Ahmad, S., Ullah Munir, E., Ali Khan, I., & Ramzan, N. (2024). Predictive modelling and identification of key risk factors for stroke using machine learning. *Scientific Reports*, *14*(1), 11498. https://doi.org/10.1038/s41598-024-61665-4

[44] Hassan, A., & Yousaf, N. (2022). Bankruptcy prediction using diverse machine learning algorithms. In *2022 International Conference on Frontiers of Information Technology*, 106–111. https://doi.org/10.1109/FIT57066.2022.00029

[45] Gousiadou, C., Doganis, P., & Sarimveis, H. (2023). Development of artificial neural network models to predict the PAMPA effective permeability of new, orally administered drugs active against the coronavirus SARS-CoV-2. *Network Modeling Analysis in Health Informatics and Bioinformatics*, *12*(1), 16. https://doi.org/10.1007/s13721-023-00410-9

[46] Noroozi, Z., Orooji, A., & Erfannia, L. (2023). Analyzing the impact of feature selection methods on machine learning algorithms for heart disease prediction. *Scientific Reports*, *13*(1), 22588. https://doi.org/10.1038/s41598-023-49962-w

[47] Eljialy, A. E. M., Uddin, M. Y., & Ahmad, S. (2024). Novel framework for an intrusion detection system using multiple feature selection methods based on deep learning. *Tsinghua Science and Technology*, *29*(4), 948–958. https://doi.org/10.26599/TST.2023.9010032

[48] Pudjihartono, N., Fadason, T., Kempa-Liehr, A. W., & O'Sullivan, J. M. (2022). A review of feature selection methods for machine learning-based disease risk prediction. *Frontiers in Bioinformatics*, *2*, 927312. https://doi.org/10.3389/fbinf.2022.927312

[49] Scholbeck, C. A., Casalicchio, G., Molnar, C., Bischl, B., & Heumann, C. (2024). Marginal effects for non-linear prediction functions. *Data Mining and Knowledge Discovery*, *38*(5), 2997–3042. https://doi.org/10.1007/s10618-023-00993-x

[50] Khatoon, A., Daud, A., & Amjad, T. (2024). Categorization and correlational analysis of quality factors influencing citation. *Artificial Intelligence Review*, *57*(3), 70. https://doi.org/10.1007/s10462-023-10657-3

[51] Kuzudisli, C., Bakir-Gungor, B., Bulut, N., Qaqish, B., & Yousef, M. (2023). Review of feature selection approaches based on grouping of features. *PeerJ*, *11*, e15666. https://doi.org/10.7717/peerj.15666

[52] Huang, L., Zhou, X., Shi, L., & Gong, L. (2024). Time series feature selection method based on mutual information. *Applied Sciences*, *14*(5), 1960. https://doi.org/10.3390/app14051960

[53] Young, A. L., van den Boom, W., Schroeder, R. A., Krishnamoorthy, V., Raghunathan, K., Wu, H. T., & Dunson, D. B. (2023). Mutual information: Measuring nonlinear dependence in longitudinal epidemiological data. *PLoS ONE*, *18*(4), e0284904. https://doi.org/10.1371/journal.pone.0284904

[54] Meem, K. R. H., Islam, S., Adnan, A. O. S., & Momen, S. (2022). Premature birth prediction using machine learning techniques. In *Artificial Intelligence Trends in Systems: Proceedings of 11th Computer Science On-line Conference*, *2*, 270–284. https://doi.org/10.1007/978-3-031-09076-9_26

[55] Dospinescu, O., & Perca, M. (2011). Technological integration for increasing the contextual level of information. *Scientific Annals of Alexandru Ioan Cuza University of Iasi: Economic Series*, *58*(1), 571–581.

[56] Lou, Z., Wang, L., Jiang, K., Wei, Z., & Shen, G. (2020). Reviews of wearable healthcare systems: Materials, devices and system integration. *Materials Science and Engineering: R: Reports*, *140*, 100523. https://doi.org/10.1016/j.mser.2019.100523

How to Cite: Hassan, A., Nawaz, S., Tahira, S., & Ahmed, A. (2025). Preterm Birth Prediction Using an Explainable Machine Learning Approach. *Artificial Intelligence and Applications*. https://doi.org/10.47852/bonviewAIA52024517