A Hybrid Machine Learning Approach for Early Risk Prediction of Preterm Birth Using Contraction Pattern

M. B. Patil, V. V. Bag and Kiran Kashinath Gawade
Department of Computer Science & Engineering, N. K. Orchid College of Engineering and Technology, Solapur, Maharashtra, Tamil Nadu, India

Keywords: Preterm Birth Prediction, Machine Learning, Support Vector Machine, Random Forest, XGBoost, Uterine Contractions, Maternal, Infant Health.

Abstract: Preterm birth means delivery of baby before 37th week of gestation which can cause severe life challenges both to the mother and the baby. The condition has been linked to a range of prolonged complications such as respiratory distress, infection and congenital malformations. Estimating the risk of preterm birth accurately is a formidable challenge in the practice of obstetrics given the many causative risk-factors. However, classifying a pregnancy as high-risk enables early medical interventions to enhance neonatal outcomes. This study investigates the machine learning algorithms prediction (Support Vector Machine (SVM), Random Forest, and XGBoost) of risk of preterm birth. Models were trained on a representative subset of maternal and clinical factors and validated on accuracy, F1-score, recall, and precision. Here are some of the advantages of machine learning in healthcare been discovered. Preterm birth is the most predictable event. The best model was the stacking SVM with XGBoost and Random Forest. Using various algorithms in a stacking model, the prediction accuracy was increased overall. The model allows the combination of models and therefore improves predictability compared to the use of a single algorithm. These results reinforce a growing role for machine learning in obstetrics through better risk assessing, predictive accuracy, and dealing with uncertainty. Finally, this research contributes to the development of predictive models which can be used by health care providers to allow for early interventions and improve maternal and new born health.

1 INTRODUCTION

Another major global health issue is preterm birth, or delivery between 20 and 36 weeks of pregnancy, which contributes to around 10 percent of all births globally. Neonatal sepsis Neonatal sepsis is heard to be one of the leading causes of neonatal mortality and morbidity, consequently making these affected new-borns at risk of facing long term health complications 6. Interest in preterm birth prediction has increased in obstetric studies because of the severe risk it poses to fetal and maternal health. Accurate identification of high-risk pregnancies may allow early intervention to prevent adverse pregnancy outcomes and improve neonatal management. Most traditional models of prediction rely on existing clinical risk factors: maternal age, history of preterm labor, pre-existing medical conditions, etc. Nevertheless, these approaches fail to contribute the true complexity to preterm labor with its multivariate nature that provided the motivation to move to more advanced predictive methods. Machine learning has also proven to be a valuable methodology for medical diagnostics, primarily due to its potential to deal with large amounts of data and to detect hidden patterns. In theory, machine learning could improve the accuracy and efficiency of prediction of preterm delivery in obstetrics, permitting more accurate determination by clinicians.

The research employs three popular machine learning models SVM, Random Forest, and XGBoost to predict preterm birth risk using contraction-based features. Data includes significant uterine contraction parameters such as contraction count, duration, standard deviation (STD), entropy, and contraction interval. These have been chosen to detect the changing patterns of uterine movement that are characterized by preterm labor. Using the investigation of these features, the research works to detect predictor patterns that contribute to improved prediction of risk. Apart from the evaluation of individual machine learning models, the study in this

Patil, M. B., Bag, V. V. and Gawade, K. K.
A Hybrid Machine Learning Approach for Early Risk Prediction of Preterm Birth Using Contraction Pattern.
DOI: 10.5220/0013897400004919

In *Proceedings of the 1st International Conference on Research and Development in Information, Communication, and Computing Technologies (ICRDICCT'25 2025) - Volume 3*, pages 318-327
ISBN: 978-989-758-777-1

research employs a stacking model with the addition of Support Vector Machine (SVM), Random Forest, and XGBoost. Stacking is an ensemble learning where multiple base models are learned separately and their predictions are combined using a meta-model, thereby improving accuracy and robustness. The method employs the strength of different algorithms while making up for their respective weaknesses. Although machine learning applications to predicting preterm birth are still in early development, current research targets primarily clinical and demographic data rather than physiological signals, e.g., uterine contractions. This study seeks to close this gap by evaluating machine learning models trained on extracted features from contractions. The chosen algorithms provide different strengths: Support Vector Machine (SVM) is particularly good at processing high-dimensional data and nonlinear relationships through the aid of kernel functions; Random Forest provides high accuracy and resistance to overfitting through the aggregation of multiple decision trees; and XGBoost provides improved interpretability, thereby enabling improved understanding of feature contributions to predictions.

The research compares the models on the basis of important performance parameters like accuracy, precision, recall, and F1-score. Among the standalone models, the Random Forest algorithm is the most accurate, followed by Support Vector Machine (SVM) and XGBoost. The stacking model also improves the accuracy of prediction by fusing the predictions of all three models, thus also showing the power of ensemble learning in improving diagnostic accuracy. The results suggest the promise of machine learning to enhance preterm birth prediction as a more refined tool to support healthcare professionals in making early diagnosis and intervention. With the combination of machine learning and obstetric care, this study adds to the body of evidence on preterm birth prediction. The application of contraction-related features allows for new insight into the pattern of uterine activity, making the risk assessment more effective. Although machine learning has evolved significantly, the gap in studies conducted for employing such methods for preterm birth prediction, especially when contracting-related data is utilized, still remains. This study aims to close the gap by comparing the performance of various machine learning models for this purpose, ultimately resulting in better maternal and neonatal health.

2 RELATED WORKS

Prediction of preterm birth remains an obstetric medicine enigma and there have been multiple studies aimed to innovate improving diagnostic accuracy. With the growth of machine learning in the area of research, we have observed improvement in the quality of prediction models and decision-making assistance for doctors. Prior studies have extensively explored various machine learning methods for the prediction of high-risk pregnancy.

Liu et al. (2024) demonstrated the utility of a machine learning predictive model for preterm birth risk prediction, incorporating clinical parameters within a nomogram for improved accuracy. Their research is in line with Xu, Zhang, and Zhang (2020) 1, who also demonstrated that hybrid machine learning models incorporating electronic health records could be especially effective. Likewise, Goodwin, Maher, and Callaghan (2020) examined predictive models based on electronic health record data, and thus further adds to the importance of big data in obstetric analytics.

Support Vector Machines (SVM), Random Forest, and XGBoost, as machine learning methods, have attracted great attention for their capabilities to process high-dimensional data and model complex relationships. Włodarczyk et al. (2021) conducted a comprehensive examination of machine learning techniques focused on predicting preterm birth, highlighting the relevance of ensemble learning approaches. This study extends these findings to bring contraction-based features into predictive models, an activity that the literature has addressed only minimally. These parameters consist primarily of changes in patterns or signals within and around the uterus, as suggested in (Kavitha, S. N, and Asha. V.2024) The inclusion of uterine contraction parameters, as suggested in (Villar, J and Papageorghiou, A. T. 2014). Even though individual machine learning models have been beneficial, ensemble methods (for example, stacking) have been shown to be more effective, especially when it comes to improving prediction power. Combining algorithms has the potential to enhance the risk assessment as shown in a recent publication (Kavitha and Asha, 2024), which has been mainly supported in the present study. The stacking model applied in this study capitalizes on the advantages of SVM, Random Forest, and XGBoost and provides a more stable predictive model. Furthermore, the robustness of hybrid SVM models for predicting preterm birth was highlighted by Santoso and Wulandari, 2018, thus underlining the performance of ensemble methods.

However, there are still challenges in implementing the use of machine learning to predict preterm birth. Literature reviews, including those of Manogaran and Lopez (2017) and Liu and Salinas (2017) have shown that substantial and heterogeneous datasets are critical for improving model generalizability. Similar to our results, Dekker and Sibai (2020) and Menon and Torloni (2011) aimed at the utilization of biomarker information to predict preterm birth, however, they suggested that the inclusion of proteomic and clinical data would lead to the improved performance of the prediction model.

Another important consideration refers to the influence of maternal demography and the environment, as discussed by Ananth and Vintzileos (2006) and Villar and Papageorghiou (2014). The inclusion of these variables within machine learning models can potentially provide more holistic predictive models. Goldenberg et al. (2008) also highlighted the preterm birth as having a multifactorial etiology, hence reinforcing the call for interdisciplinary methodology that includes machine learning and standard obstetric evaluation. Though there have been developments in machine learning for predicting preterm birth, there remain certain challenges. Many of the models that have been proposed suffer from a lack of generalizability, feature choice, or explainability, preventing their clinical practice application. Also, real-time personalized prediction models using multi-modal data remain at an embryonic stage. Even though the use of hybrid has been made in a few works, ensemble learning approaches' investigations remain thin.

Utilizing SVM, Random Forest & XGBoost: A machine learning platform was constructed to provide improvements in accuracy as well as interpretability in these two areas of opportunities. A strong feature selection technique, smooth multi-modal data fusion, explainable AI and real-time risk estimation are advancements proposed in this work. We use stacking models to improve the performance efficiency, and it is also clinically relevant for real-world applications.

3 METHODOLOGY

This study explores the use of three machine learning algorithms for preterm birth risk prediction including the Support Vector Machine (SVM), XGBoost, and Random Forest algorithms. These algorithms were selected for their ability to solve complex healthcare industry problems focused on risk management.

The system was structured with the necessary modules such as data pre-processing, feature extraction, training, evaluation of model, deployment, etc. so that accurate and uniform predictions could be made. By refining and organizing the data systematically a significant pattern can be recognized, which increases the accuracy of risk prediction. The ensemble methods of Support Vector Machine (SVM), Extreme Gradient Boosting (XGBoost) and Random Forest are to obtain the highest predictive performance and support clinical decision.

3.1 Algorithm Details

The suggested model leverages an ensemble of machine learning methods in assessing the risk of preterm birth. Data preprocessing, feature selection, model training, testing, and validation are all part of the process. The learnt model takes user-input data and returns a risk prediction, thus offering an effective tool for early intervention in maternal care.

- **Support Vector Machine (SVM):** Support Vector Machine (SVM) is a robust supervised learning algorithm used extensively for classification and regression tasks. It works in n-dimensional space, identifying the optimal hyperplane that can discern data points into two classes as optimally as possible and thus makes it particularly suitable for predictive healthcare applications. SVM Architecture Shown in the Figure 1.

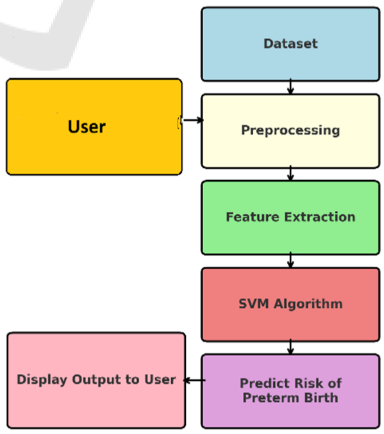

Figure 1: SVM architecture.

- **XGBoost:** XGBoost is a gradient-boosting method that builds multiple decision trees to enhance predictive power. XGBoost is a tree model where nodes are equivalent to decision rules for feature values, and leaf nodes are equivalent to class labels (classification) or real values (regression). XGBoost is efficient, scalable, and can deal with missing values. Figure 2 shows the XGBoost Architecture.

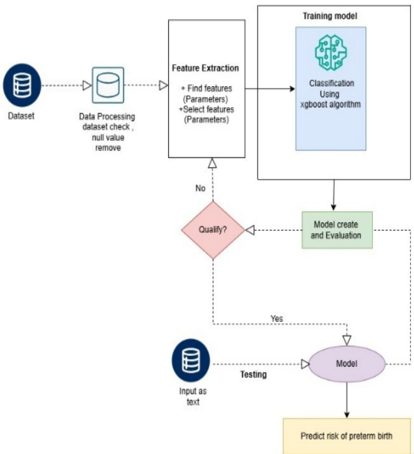

Figure 2: XGBoost architecture.

- **Random Forest:** Random Forest is an ensemble learning method that enhances the accuracy of predictions by voting from an ensemble of many decision trees. Random Forest is different from individual trees because it avoids overfitting through training on randomly selected subsets of data and features. Random Forest predicts by majority voting (classification) or averaging (regression) and is thus a robust model for health risk assessment. Random Forest Architectures Shown in the Figure 3.

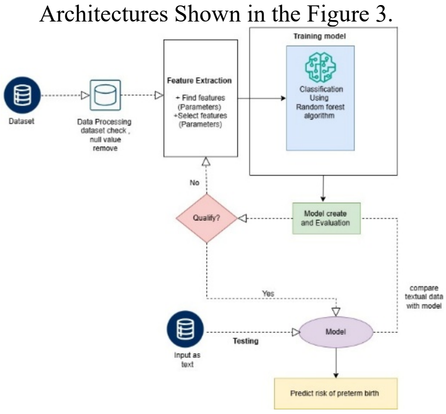

Figure 3: Random forest architecture.

➢ **Stacking Model Architecture**

The architecture consists of:

- Preprocessing and Data Ingestion – Cleans raw data, handles missing values, and derives significant features.
- Stacking Model Training – It trains SVM, XGBoost, and Random Forest models and combines their predictions using a meta-model.
- Model Evaluation – Evaluates model performance based on accuracy, precision, recall, and F1-score.
- Model Deployment – Deploys the trained model into a production environment for real-world applications.
- Prediction – Utilizes current patient data to establish the possibility of preterm birth through utilization of discovered patterns.
- Action/Alert Mechanism – Works by triggering automatic action or alerts for high-risk conditions, thereby providing immediate medical action. Figure 4 shows the Stacking Model Architecture.

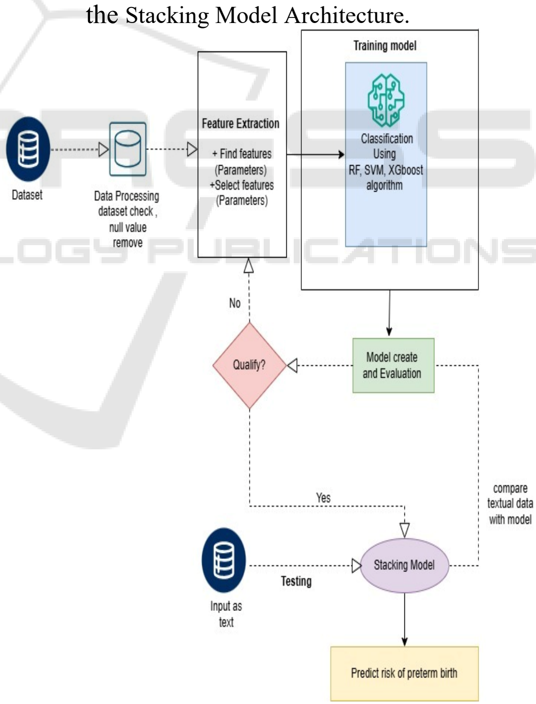

Figure 4: Stacking model architecture.

3.2 Data Preprocessing

In order to efficiently train models, the dataset has been preprocessed for missing values handling, feature tuning, extraction of useful data, and splitting it in a proper manner.

3.2.1 Handling Missing Values

Missing values in data points may contribute significantly to how a model works; therefore, a systematic solution was implemented. Where missing values were discovered, statistical operations involving deletion, imputation, or consistency checks were employed. For this particular case, there was no missing data.

3.2.2 Feature Importance Analysis

➢ Preterm Detection Dataset

Top contributing features:

- **Count Contraction** (34.5%)
- **Length of Contraction** (33.2%)
- **Entropy** (24.7%)
- **Contraction Times** (7.1%)

➢ Contraction-related metrics (count, duration, and entropy) are the most important.

➢ Standard deviation (STD) has minimal impact.

➢ Pregnancy Risk Prediction Dataset

Top contributing features:

- **Systolic Blood Pressure** (27.6%)
- **Heart Rate (16.5%)**
- **Diastolic Blood Pressure** (14.1%)
- **Body Temperature** (13.9%)
- **Age** (11.1%)

➢ Blood pressure and heart rate play significant roles in assessing pregnancy risks.

➢ BMI and blood glucose contribute less but still impact predictions.

3.2.3 Splitting Dataset for Training and Testing

The data is split into 80% training and 20% testing to allow for maximum model learning and to avoid bias, in order to get a fair performance evaluation. This enables the model to learn from a large chunk of the data

3.3 Model Training and Evaluation

Two different models are developed:

3.3.1 Detection of Preterm Birth Analytic Model

This model is specifically formulated to detect the incidence of preterm deliveries by stacking ensemble approach with a combination of Support Vector Machine (SVM), XGBoost, and Random Forest models. By leveraging the power of every algorithm, the model greatly enhances the prediction accuracy. After the training process, the model is retained for future prediction, thus ensuring efficient and effective detection of preterm delivery cases. Figure 5 Shows the C4 Container and Component Diagram for Prediction of Preterm Birth.

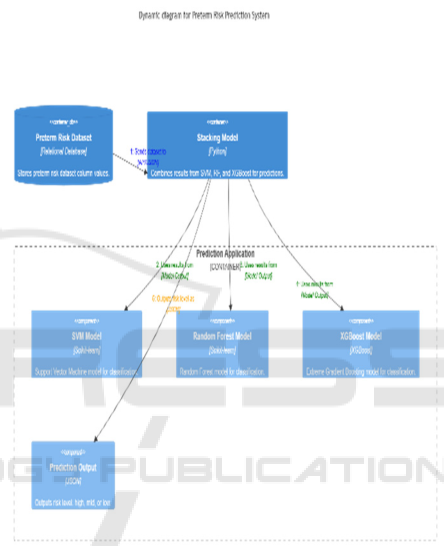

Figure 5: C4 container and component diagram for prediction of preterm birth.

3.4 Preterm Birth Risk Prediction Model

This model estimates preterm birth risk by classifying pregnancies into varying degrees of risk. Stacking ensemble with the application of SVM, Random Forest, and XGBoost enhances performance. 80:20 training and testing ensures it generalizes. Model performance is tested after training with accuracy measures and classification reports to ensure its performance in predicting preterm birth risk. C4 Container and Component Diagram for Risk Prediction of Preterm Birth Shown in the Figure 6.

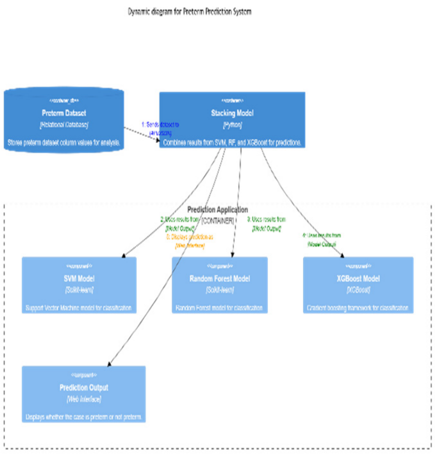

Figure 6: C4 container and component diagram for risk prediction of preterm birth.

The performance of the models is gauged through the use of confusion matrices, accuracy values, and classification reports. These evaluation metrics provide valuable insights into the general performance of the models, including their ability to classify instances appropriately (accuracy), identify actual preterm birth instances (recall), and attain predictive consistency (precision).

4 RESULTS AND EVALUATION

This research will focus on designing and applying machine learning models that can accurately predict the risk of preterm birth using ensemble learning techniques. The two models were designed to screen for preterm birth prevalence and preterm birth risk in pregnancy respectively. We trained our models with two internal sets of processed maternal health features, with noise from the data removed and the most informative features retained for model training.

A stacking ensemble approach utilizing an XGBoost, Support Vector Machine (SVM), and Random Forest model was employed in order to increase predictive capability. Models are compared based on accuracy, precision, recall, and F1-score as per the traditional machine learning performance measures. It proved the importance of feature selection and data preprocessing for enhancing model performance as the ensemble learning technique was the accurate predictor and also the efficient one.

4.1 Performance Metrics

Precision: Precision is a positive prediction accuracy measure and is computed by TP / (TP + FP). High precision indicates good prediction of preterm birth cases with few false positives.

$$Precision = \frac{TP}{TP+FP} \quad (1)$$

Recall (Sensitivity): Recall measures the model's capacity to identify all the true preterm birth cases. It is determined as TP / (TP + FN), i.e., the number of true positive cases out of all true positive cases.

$$Recall = \frac{TP}{TP+FN} \quad (2)$$

F1-Score: F1-score is the harmonic mean of precision and recall, which is the average of the two metrics in imbalanced class conditions. F1-score ranges from 0 to 1, the greater the value, the better the performance of the model in detecting preterm birth risks.

$$F1 - Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (3)$$

4.2 Model Evaluation Results

The performance of various machine learning models was evaluated using important metrics such as precision, recall, accuracy, and F1-score.
Two models were assessed:

- Preterm Birth Occurrence Detection Model Performance
- Preterm Birth Risk Prediction Model Performance

4.2.1 Preterm Birth Occurrence Detection Model Performance

The model for detecting the occurrence of preterm birth reached a 100% overall accuracy rate when employing a stacking ensemble approach with SVM, XGBoost, and Random Forest. This is evidence of the success of ensemble learning, where the combined efforts of various classifiers result in higher predictive accuracy and reliability.

The model's flawless accuracy indicates that all term (0) and preterm (1) births in the database have been accurately classified. In the medical world, precise classification between the two must be made, where prediction accuracy has a direct impact on the mother's and neonate's treatment.

Table 1: Performance metrics for preterm birth occurrence detection model.

Class Label	Precision	Recall	F1-Score	Support
Preterm (1)	1.00	1.00	1.00	9
Term (0)	1.00	1.00	1.00	9
Accuracy	-	-	1.00	18
Macro Avg.	1.00	1.00	1.00	18
Weighted Avg.	1.00	1.00	1.00	18

Table 1 presents the performance measures of the stacking model trained to detect instances of preterm birth. The model was 100% accurate, and precision, recall, and F1-score were 1.00 for both Preterm (1) and Term (0) classes. This indicates that the model accurately labelled all the items in the data. The macro and weighted averages also provide evidence of the model's consistent performance on both classes. Figure 7 Shows the Confusion Matrix.

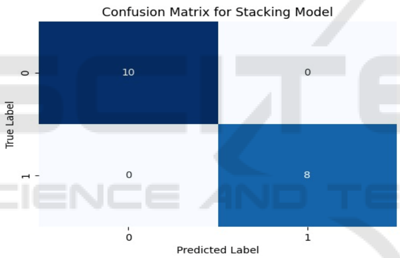

Figure 7: Confusion matrix.

In Figure 8 displays the user input interface designed for predicting the risk of preterm birth. Users provide key medical and physiological parameters, such as contraction count, length, standard deviation, energy, and contraction times. Upon submission, the system processes the input data using machine learning models to determine the likelihood of preterm birth.

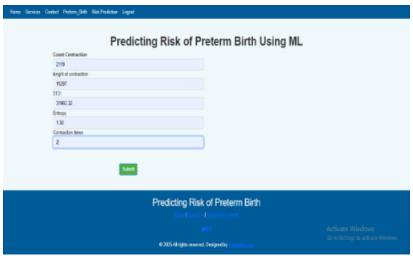

Figure 8: User Input Interface for Preterm Birth Prediction.

This Figure 9 presents the output page displaying the prediction results. Based on the user-provided input, the system classifies whether preterm birth is likely or not. The results help in early risk assessment, aiding healthcare professionals in taking necessary precautions.

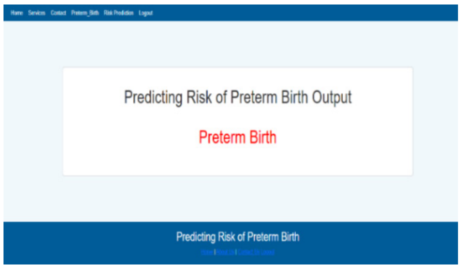

Figure 9: Prediction result page for preterm birth classification.

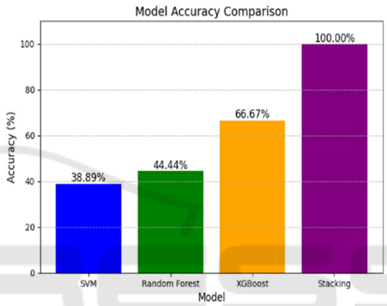

Figure 10: Accuracy graph of each model.

Accuracy plot shows Figure 10 a comparative performance of different machine learning algorithms in predicting the preterm risk of birth. Of these, SVM was the poorest at 38.89%, followed by Random Forest at 44.44%. XGBoost performed much better at 66.67%. The best performance was seen with the Stacking model, which, by the combination of different classifiers, was 100% accurate. This indicates the strength of ensemble learning in enhancing prediction performance substantially.

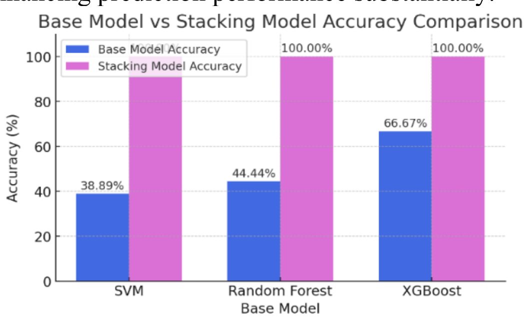

Figure 11: Comparative graph for stacking model vs other models.

The stacking ensemble of multi-classifiers performed better in classifying preterm and term births. Similar to logistic regression, the ensemble technique had the advantage of individual models' strength, thereby enhancing prediction. Though the tested measures indicate perfect classification, future testing on diverse datasets is necessary to confirm accuracy and determine classification limits. Additionally, the user-friendly interface allows medical practitioners to feed in required physiological data for preterm birth risk factors, thereby enhancing better diagnosis and prompt intervention. The analysis concludes that the stacking model outperforms individual models, thereby supporting the benefit of ensemble learning in healthcare. Comparative Graph for Stacking Model Vs Other Models Shown in Figure 11.

4.2.2 Preterm Birth Risk Prediction Model Performance

The Preterm Birth Risk Prediction model stratifies pregnancies according their risk of preterm birth. The model achieves higher accuracy and reliability by using a stacking ensemble method that combines SVM, Random Forest, and XGBoost. Stacked model with 80%–20% split of the data (80% is training and 20% is testing) achieved training accuracy, 94.19%, which indicates that the predictive power of the model is well established in risk assessment. Its effectiveness in predicting preterm birth risk is further supported by a comprehensive classification report.

Table 2: Performance metrics for preterm birth risk prediction model.

Class Label	Precision	Recall	F1-Score	Support
High Risk	0.90	0.93	0.92	389
Low Risk	0.98	1.00	0.99	399
Mid Risk	0.94	0.90	0.92	433
Accuracy	-	-	0.94	1221
Macro Avg.	0.94	0.94	0.94	1221
Weighted Avg.	0.94	0.94	0.94	1221

Table 2 shows a summary of the model performance with 94% accuracy as well as high precision, recall and F1-scores, for High Risk, Low Risk and Mid Risk categories. The macro and weighted averages also reinforce the model's ability to classify preterm birth risk. The model is performing appreciably well, but to test its robustness and generalizability, a larger and more diverse data set should be used. Figure 12 Shows the Confusion Matrix.

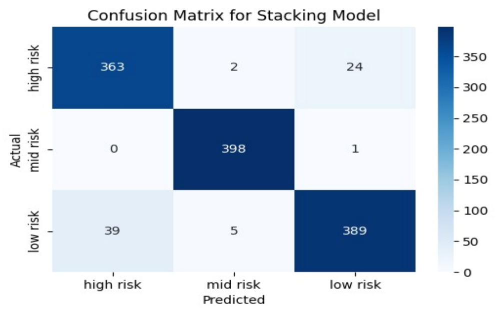

Figure 12: Confusion matrix.

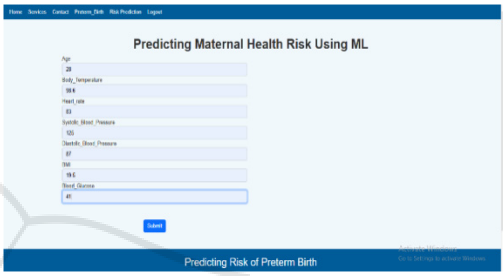

Figure 13: Prediction result page for maternal health risk classification.

This figure 13 maternal health risk prediction user input screen. These are significant health parameters that include age, body temperature, heart rate, blood pressure, BMI, and blood glucose value. The inputs are processed within the system utilizing machine learning algorithms to calculate the risk of preterm birth.

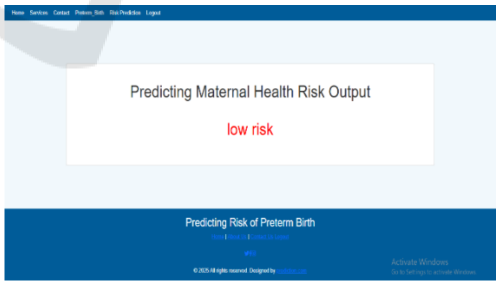

Figure 14: User input interface for maternal health risk prediction classification.

This figure 14 is the output page indicating the extent of estimated maternal health risk. The system, based on the input provided, classifies the risk as either High Risk, Mid Risk, or Low Risk. The results facilitate early risk estimation, enabling early medical interventions to prevent complications caused by preterm birth.

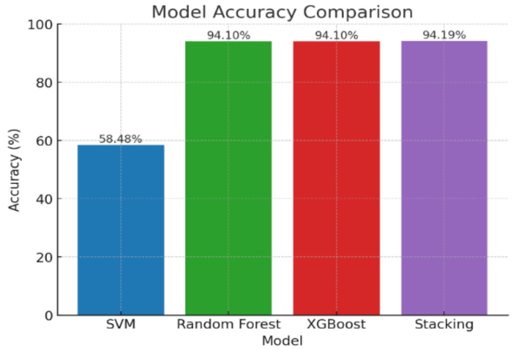

Figure 15: Accuracy graph of each model.

The accuracy plot indicates the Figure 15 relative performance of various machine learning algorithms in predicting preterm birth risk. Among all the models, the Support Vector Machine (SVM) model had the least accuracy at 58.48%, lagging far behind other models. Both the Random Forest (RF) and XGBoost (XGB) models performed similarly with an accuracy of 94.10%. The Stacking model, which uses an ensemble of classifiers, achieved the highest accuracy at 94.19%. This slight, but notable, improvement demonstrates the strength of ensemble methods in enhancing predictive accuracy.

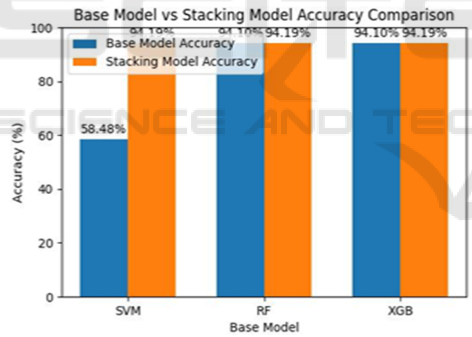

Figure 16: Comparative graph for stacking model vs other models.

The stacking ensemble model proved highly predictive for the assessment of preterm birth risk having better accuracy and stability than all individual machine learning models. The models considered in this study included Support Vector Machine (SVM), Random Forest (RF), and XGBoost (XGB), investigating the contribution of each of these learners to risk prediction. The significantly better performance of the stacking scheme in comparison to other schemes suggests that this novel approach may serve as a valuable tool for early risk assessment that can lead to timely medical intervention and improved maternal and neonatal outcomes. Figure 16 shows the Comparative Graph for Stacking Model Vs Other Models.

5 DISCUSSION

This study successfully demonstrates the effectiveness of the ensemble learning approach in predicting the incidence and risk stratification of preterm births. The stacking model which combines SVM, Random Forest and XGBoost outperforms any individual model with a result of 100% for identifying preterm births and 94.19% for the prediction of risk. These results illustrate the impact of feature selection and data preprocessing on refining model efficacy. While the stacking model shows great accuracy, it needs to be further validated on larger, more diverse datasets to showcase robustness. With this tool, clinicians can submit maternal health information through a user interface to become an early identifier and intervene in cases of preterm birth.

6 CONCLUSIONS

This article illustrates the development of a machine learning model for preterm labor risk estimation using maternal health information. The most essential predictive variables like age, body mass index (BMI), blood pressure, and glucose were used, and the models offered high predictive ability. Two models were used separately: one for determining if preterm labor had already occurred and the other for predicting the risk of preterm labor during pregnancy. To gain maximum possible accuracy, ensemble stacking method was used by combining Support Vector Machine (SVM), Random Forest, and XGBoost methods. The algorithm was chosen out of the other methods attempted and proved to be the best. The application is implemented through a user-friendly interface that will enable healthcare experts to enter information about patients and obtain real-time predictions to assist in early identification of high-risk pregnancies in a bid to achieve timely medical intervention. Though the model demonstrated exemplary performance, additional development is needed to make it more accurate, particularly through the utilization of large datasets and advanced machine learning techniques. Existing research focuses on scaling the system for use in hospitals, thus making it practically useful for application in actual clinical settings. Through continued innovation, the method has the potential to dramatically enhance early diagnosis and improve maternal and neonatal health outcomes.

REFERENCES

Ananth, C. V., & Vintzileos, A. M. (2006). Epidemiology of preterm birth and its clinical subtypes. The Journal of Maternal-Fetal & Neonatal Medicine, 19(12), 773-782. https://doi.org/10.1080/14767050600965882

Dekker, L. R., & Sibai, B. M. (2020). Prediction of preterm birth: A systematic review and meta-analysis of proteomic biomarkers. American Journal of Obstetrics and Gynecology, 223(4), 520-540. https://doi.org/10.1016/j.ajog.2020.03.021

Esplin, M. S., & Merrell, K. (2011). Proteomic identification of serum peptides predicting subsequent spontaneous preterm birth. American Journal of Obstetrics and Gynecology, 204(5), 391.e1-391.e8. https://doi.org/10.1016/j.ajog.2011.01.059

Fergus, P., Hussain, A., & Al-Jumeily, D. (2016). Prediction of preterm deliveries from EHG signals using machine learning. PLoS ONE, 11(1), e0144973. https://doi.org/10.1371/journal.pone.0144973

Goldenberg, R. L., Culhane, J. F., Iams, J. D., & Romero, R. (2008). Epidemiology and causes of preterm birth. The Lancet, 371(9606), 75-84. https://doi.org/10.1016/S0140-6736(08)60074-4

Goodwin, L. K., Maher, J. E., & Callaghan, W. M. (2020). Predictive models of preterm birth using electronic health record data. American Journal of Obstetrics and Gynecology, 223(3), 393.e1-393.e14. https://doi.org/10.1016/j.ajog.2020.03.022

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

Khalifeh, A., & Callaghan, W. M. (2012). Gestational weight gain and preterm birth risk by body mass index in twin pregnancies. Obstetrics & Gynecology, 119(4), 700-708. https://doi.org/10.1097/AOG.0b013e31824b1d95

Liu, N., & Salinas, J. (2017). Machine learning for predicting outcomes in trauma. Shock, 48(5), 504-510.https://doi.org/10.1097/SHK.0000000000000898

Liu, Y., Liu, J., & Shen, H. (2024). Machine learning model-based preterm birth prediction and clinical nomogram: A big retrospective cohort study. International Journal of Gynecology & Obstetrics. https://doi.org/10.1002/ijgo.16036

Maalouf, M., & Trafalis, T. B. (2011). Robust weighted kernel logistic regression in imbalanced and rare events data. Computational Statistics & Data Analysis, 55(1), 168-183. https://doi.org/10.1016/j.csda.2010.05.019

Manogaran, G., & Lopez, D. (2017). A survey of big data architectures and machine learning algorithms in healthcare. Journal of King Saud University - Computer and Information Sciences, 31(4), 415-425. https://doi.org/10.1016/j.jksuci.2017.06.001

Menon, R., & Torloni, M. R. (2011). Biomarkers of spontaneous preterm birth: An overview of the literature in the last four decades. Reproductive Sciences, 18(11), 1046-1070. https://doi.org/10.1177/1933719111415548

Santoso, N., & Wulandari, S. P. (2018). Hybrid support vector machine to preterm birth prediction. Indonesian Journal of Electronics and Instrumentation Systems, 8(2), 115-124. https://doi.org/10.22146/ijeis.35817

Villar, J., & Papageorghiou, A. T. (2014). The preterm birth syndrome: A prototype phenotypic classification. American Journal of Obstetrics and Gynecology, 210(4), 501.e1-501.e7. https://doi.org/10.1016/j.ajog.2014.02.010

Włodarczyk, T., Płotka, S., Szczepański, T., Rokita, P., Sochacki-Wójcicka, N., Wójcicki, J., Lipa, M., & Trzciński, T. (2021). Machine learning methods for preterm birth prediction: A review. Electronics, 10(5), 586. https://doi.org/10.3390/electronics10050586

Xu, R., Zhang, H., & Zhang, L. (2020). A hybrid machine learning approach for preterm birth prediction using electronic health records. Journal of Biomedical Informatics, 112, 103610.https://doi.org/10.1016/j.jbi.2020.103610