International Journal of Engineering Research and Science & Technology

IJERST

www.ijerst.org

ISSN : 2319-5991

Vol. 16 No. 3 (2020)

ijerst.editor@gmail.com

editor@ijerst.com

ISSN 2319-5991 www.ijerst.com
Vol. 16, No. 3, 2020

Research Paper

A MACHINE LEARNING-BASED FRAMEWORK FOR PRETERM BIRTH PREDICTION USING MATERNAL DATA IN RURAL HEALTHCARE SETTINGS

Konakala Srinivasa Rao[1], Dr.K.Srinivas[2]
[1]*Research Scholar,OPJS University, Churu,Rajasthan, India.*
[2]*Professor,OPJS University,Rajasthan, India.*

Abstract- Preterm birth (PTB) is a major contributor to neonatal mortality and long-term health complications, particularly in rural and resource-limited regions where access to advanced diagnostic facilities is restricted. Early prediction of PTB is critical for timely intervention; however, conventional clinical approaches are often costly and less accessible. This study presents a machine learning-based predictive framework for assessing preterm birth risk using maternal and obstetrical data collected from a rural healthcare center. The proposed approach incorporates data preprocessing, discretization, and an entropy-based feature selection technique to identify the most relevant maternal risk factors. Three classification algorithms—Decision Tree (DT), Logistic Regression (LR), and Support Vector Machine (SVM)—are evaluated using performance metrics such as accuracy, sensitivity, and specificity. Due to class imbalance in the dataset, the Synthetic Minority Oversampling Technique (SMOTE) is applied to improve model learning. Experimental results demonstrate that SVM achieves the highest accuracy of 86.1% on the original dataset, which significantly improves to 90.9% after dataset balancing. Additionally, sensitivity increases to 89.1%, indicating enhanced capability in identifying high-risk pregnancies. Comparative analysis confirms that SVM outperforms DT and LR in predictive performance. The results highlight the effectiveness of combining feature selection and data balancing with machine learning models for accurate PTB prediction. The proposed framework offers a cost-effective and reliable solution for early risk detection, making it highly suitable for deployment in rural healthcare environments to support clinical decision-making and improve maternal and neonatal outcomes.

Keywords- *Preterm Birth, Machine Learning, Rural Healthcare, Predictive Modeling, Maternal Health, Classification Algorithms*

1. Introduction

Preterm birth, defined as delivery before 37 weeks of gestation, is a major global health concern and a leading cause of neonatal mortality and morbidity. It significantly contributes to long-term complications such as neurological disorders, respiratory issues, and developmental delays [1]. The burden of preterm birth is particularly severe in developing countries, where access to quality maternal healthcare remains limited [2].

In rural regions, healthcare infrastructure is often inadequate, and pregnant women face challenges such as limited access to diagnostic facilities, lack of awareness, and economic constraints [3]. Traditional diagnostic methods, including ultrasound-based cervical length measurement, are not only expensive but also dependent on expert interpretation, leading to variability and potential misdiagnosis [4].

The increasing availability of healthcare data and advancements in computational techniques have opened new avenues for predictive healthcare [5]. Machine learning (ML) provides powerful tools for analyzing complex medical datasets and identifying hidden patterns associated with preterm birth. By leveraging historical maternal and clinical data, ML models can support early risk detection and enable timely interventions [6].

This study aims to develop a machine learning-based predictive framework for preterm birth, specifically tailored for rural healthcare settings. The proposed approach focuses on improving prediction accuracy while reducing computational complexity and reliance on expensive diagnostic procedures.

2. Related Work

Methodologies for predicting PTB utilizing machine learning, statistical analysis, and data mining are the primary emphasis of this subsection. In the following, we'll look at a few of them in more detail. Mercer et al [6] .'s research aimed to create a risk-score based model for estimating the likelihood of PTB. Clinical data collected between 23 and 24 weeks of pregnancy may be used to train the model and reveal hidden risk variables using a multivariate logistic regression approach. In order to build 520 predicting criteria for PTB, Goodwin et al. [7] used a machine learning model based on data mining approaches.

Several different classifiers, including K-nearest neighbors, lasso regression, and random forests, were used to make predictions for PTB. Maternal traits, maternal education, and racial/ethnic background were all included in this research. A multivariate logistic regression model was used by Mailath-Pokorny et al. [8] to investigate the factors that predict a premature birth that occurs within 2 days of hospital admission and before 224 weeks of pregnancy. Predictive factors included in the analysis were the mother's age, the gestational age upon admission, the history of the pregnancy, any vaginal bleeding, the length of the cervix, any previous preterm births, and PPROM. Women carrying a single child were the focus of Son and Miller's PTB prediction model, which used cervical length as an indicator. They looked for the optimal cut sites of cervical length to improve prediction performance [9].

Through the use of neural networks, Elaveyini et al. [10] investigated the most important causes of premature birth. With the help of the feed-forward backpropagation algorithm, PTB predictions were made. The vast bulk of research over the last several decades has focused on improving the precision of PTB prediction [11]. Scientists never stop trying to determine and understand the most important causes of preterm birth [12]. Pregnancy prediction using machine learning methods for a rural setting is the topic of this essay.

A Comprehensive Overview of Preterm Birth (PTB). Preterm or preterm birth is defined as birth happening before 37 full weeks (or fewer than 259 days) of pregnancy for any cause. Every year, around fifteen million infants are born prematurely (before 37 weeks of gestation), accounting for roughly one-tenth of all births worldwide. According to WHO data from 2005, 12.9 million births, or 9.6% of all births worldwide, occurred prematurely [13]. However, the rate of preterm birth varies greatly over the globe. Preterm delivery is the leading cause of neonatal morbidity and death [14].

PTB categorization: PTB is classified into many groups depending on gestational age at birth. The gestational age of a woman is defined as the time from the first day of her last menstrual period (LMP) until delivery. PTB is classified into four categories:

- Extreme PTB (under 28 Weeks). It is the birth that takes place before 28 weeks of pregnancy
- Very PTB (28 to 32 Weeks). It is the birth that takes place between 28 and 32 weeks of pregnancy
- Moderate PTB (32 to 34 Weeks). It is the birth that takes place between 32 and 34 weeks of pregnancy
- Late PTB (34 to 37 Weeks). It is the birth that takes place between 34 and 37 weeks of pregnancy

Terminologies used in medicine. The employed terminology are presented in Table 1 for the sake of clarity in the current investigation.

PTB's Health Effects: PTB is the leading cause of neonatal death and morbidity. It is the largest cause of newborn mortality and morbidity worldwide, as well as the second highest cause of child fatalities under the age of five. It occurs in 5 to 10% of all births and accounts for 70% of newborn death and up to 75% of neonatal morbidity. Premature babies are more likely to suffer than typical babies and are more prone to suffer from cerebral paralysis, sensory impairment, respiratory failure, and other complications. Only in the United States is it expected that the premature cost of maternity services would exceed $13 billion [15]. Most PTB survivors experience major challenges, typically resulting in a lifetime of handicap, such as learning disorders, vision and hearing impairments.

Table 1: Definitions used in the present study.

Terminology	Description
Antenatal care	Antenatal care (ANC) refers to the fundamental, clinical, and nursing care suggested for ladies during pregnancy
Neonate	A neonate or a new-born infant is a child under 28 days of age
Neonatal death	A death during the first 28 days of life (0–27 days) is termed as a neonatal death
Live birth	A birth at which a child is born alive is termed as live birth
Term birth	A birth at the end of a normal duration of pregnancy between 37 and 40 weeks of gestation is termed as term birth
Maternal death	A maternal death is the death of a woman while pregnant or within 42 days of termination of pregnancy
Stillbirth	Stillbirth is the delivery, after the 20th week of pregnancy, of a baby who has died
Abortion	Termination of a pregnancy either medically or induced
Miscarriage	Natural loss of pregnancy during first trimester
Gestational age	Gestational age (GA) refers to the time from the first day of a woman's last menstrual period to birth

In fact, babies delivered prematurely have greater health issues than those born at term. Term birth refers to kids born between 37 and 40 weeks of gestation. Furthermore, premature newborns have been linked to an increased risk of long-term health issues [16]. Unfortunately, after many years of obstetric study, the rate of PTB has not diminished [17]. PTB is often connected with birth weight, which results in its own classification. Birth weight is usually easier to measure accurately and serves as an initial estimate of gestational age. The most difficult topic in Gynecology and Obstetrics is obviously how to control premature birth in pregnant women.

Feature Choice (FS). In machine learning, feature selection, also known as feature subset selection, refers to the process of picking a subset of outstanding characteristics during the predictive model's building. The inclusion of redundant and unnecessary characteristics in any dataset (particularly medical datasets) might impair model prediction accuracy and have a detrimental influence on model performance. The primary purpose of any feature selection approach is to pick the optimal subset of features by deleting duplicate and unnecessary features from datasets in order to save training time and improve the predicted performance of the classifier. In reality, feature selection is often employed in data mining as a preprocessing step. The feature selection technique has three typical approaches: filter method, wrapper method, and embedding method. [18] provides further information about feature selection.

Method of Filtering. Based on the nature of the data, the filter technique determines the significance of characteristics. The characteristics are chosen independently of the classifiers. When compared to the wrapper technique, the filter method is significantly quicker and delivers an average accuracy for all classifiers utilized. Information gain, chi-square test, variance threshold, and other filter techniques are examples.

The Wrapper Method. The wrapper technique selects the optimal subset of features based on a machine learning algorithm that we are attempting to fit to a particular dataset. The assessment criteria are simply that the wrapper technique outperforms the filter method in terms of performance accuracy, but it takes more computing time to find the optimal features in a dataset with high-dimensional features. Wrapper approaches include forward selection, backward removal, genetic algorithms, and so on [19].

Method of Embedding. The embedded technique combines the benefits of both the filter and wrapper methods. In this strategy, feature selection occurs during model training and is generally unique to certain learning classifiers. While making a forecast, this technique essentially decides the relevance of each characteristic, i.e., which features to accept and which to reject. The decision tree algorithm is the most used embedded approach. In terms of temporal complexity, this approach is often halfway between the filter method and the wrapper method. Laspo regression, ridge regression, elastic net, and other embedded approaches are examples. The specific classifier's predictive power. The wrapper technique outperforms the filter method in terms of performance accuracy, but it takes more computing effort to find the optimal features in a dataset containing high-dimensional features. Wrapper approaches include forward selection, reverse removal, genetic algorithms, and others [20].

Method of Embedding. The embedded technique combines the benefits of both the filter and wrapper methods. In this strategy, feature selection occurs during model training and is generally unique to certain learning classifiers. While making a forecast, this technique essentially decides the relevance of each characteristic, i.e., which features to accept and which to reject. The decision tree algorithm is the most used embedded approach. In terms of temporal complexity, this approach is often halfway between the filter method and the wrapper method. Lasso regression, ridge regression, elastic net, and other embedded approaches are examples [21].

Existing Clinical Models Have Deficits. A significant number of clinical prediction models have been built in recent years employing the feature selection strategy to increase the accuracy of learning models. To the best of the authors' knowledge, however, the majority of them struggle with identifying the most accurate features from the medical dataset in linear time. As a result, there is room to improve the effectiveness of machine learning classifiers while decreasing learning time [22-25].

3. Experimental Design

This section describes the overall design of the proposed machine learning framework for predicting preterm birth (PTB). The framework consists of three major stages: data acquisition, data preprocessing and feature selection, and model development with evaluation. Each stage is carefully designed to ensure accurate prediction while maintaining computational efficiency, particularly for deployment in rural healthcare environments.

3.1 Data Acquisition

The dataset used in this study was collected from a community healthcare center located in a rural region. The data collection process involved patient surveys, maternal medical records, and neonatal birth outcomes. The survey captured detailed information related to demographic characteristics, medical history, previous pregnancies, and complications during the current pregnancy.

Initially, 1800 maternal records were collected over the study period. After applying inclusion and exclusion criteria—such as considering only live births, singleton pregnancies, and gestational age above 28 weeks—a total of

1300 records were selected for analysis. The final dataset consists of both term birth (TB) and preterm birth (PTB) cases, forming a binary classification problem. The dataset includes 36 attributes representing maternal and clinical factors relevant to pregnancy outcomes.

3.2 Data Preprocessing

The collected data required preprocessing to ensure quality and consistency before applying machine learning algorithms. The dataset contained a mix of continuous, categorical, and missing values. Missing values in each attribute were handled by replacing them with the mean value of the corresponding feature to minimize information loss.

To improve compatibility with machine learning models, continuous attributes were transformed into discrete intervals using a data discretization technique based on the principle of minimal information loss. This step enhances the learning capability of classifiers by converting raw numerical data into structured categorical representations. The resulting dataset is normalized and formatted for efficient processing in subsequent stages.

3.3 Feature Selection

Feature selection plays a critical role in improving model performance by identifying the most relevant attributes influencing preterm birth. In this study, an entropy-based feature selection approach is employed to evaluate the significance of each feature.

From the initial set of 36 attributes, 17 key features were selected based on their contribution to prediction accuracy. These selected features represent important maternal risk factors identified through both statistical analysis and expert medical knowledge. By eliminating redundant and irrelevant attributes, the feature selection process reduces computational complexity, improves model interpretability, and enhances classification performance.

3.4 Dataset Preparation and Balancing

The dataset is inherently imbalanced, with a higher proportion of term birth cases compared to preterm birth cases. This imbalance can negatively impact model performance, particularly in detecting minority class instances.

To address this issue, the Synthetic Minority Oversampling Technique (SMOTE) is applied to generate synthetic samples for the minority class (PTB). This results in a balanced dataset where both classes are equally represented. The balanced dataset improves the learning process and enhances the model's ability to correctly identify preterm birth cases.

3.5 Model Development

The prediction model is developed using three widely used machine learning classifiers: Decision Tree (DT), Logistic Regression (LR), and Support Vector Machine (SVM). These algorithms are selected due to their effectiveness in classification tasks and their applicability in healthcare data analysis.

The dataset is divided into training and testing subsets using a 70:30 ratio. The training set is used to build the predictive models, while the testing set is used to evaluate their performance. Each classifier is trained using the selected features and optimized to achieve the best possible prediction accuracy.

3.6 Performance Evaluation

The performance of the classifiers is evaluated using metrics derived from the confusion matrix, including accuracy, sensitivity, and specificity. Accuracy measures the overall correctness of the model, while sensitivity evaluates the model's ability to correctly identify preterm birth cases. Specificity measures the ability to correctly classify term births.

In healthcare applications, sensitivity is particularly important, as failing to detect a high-risk case can lead to severe consequences. Therefore, the evaluation focuses not only on accuracy but also on the balance between sensitivity and specificity.

3.7 System Workflow Overview

The overall workflow of the proposed framework begins with data collection from the healthcare center, followed by preprocessing and feature selection. The processed dataset is

IJERST International Journal of Engineering Research and Science & Technology

then used to train machine learning models, which are evaluated using appropriate performance metrics. Finally, the best-performing model is selected for predicting preterm birth risk.

This structured approach ensures that the proposed system is both accurate and practical for real-world implementation, particularly in rural healthcare settings where resources are limited. The architecture diagram shown in Figure. 1.

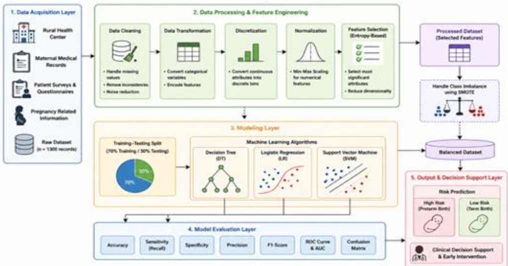

Fig. 1. System architecture of the proposed machine learning-based framework for preterm birth prediction in rural healthcare settings.

The system architecture for preterm birth prediction consists of multiple interconnected layers that process healthcare data from acquisition to decision support. Initially, data is collected from rural healthcare centers, including maternal medical records, patient surveys, and pregnancy-related information. The collected raw dataset is then passed to the data processing layer, where data cleaning, transformation, discretization, and normalization are performed. An entropy-based feature selection method is applied to identify the most significant attributes influencing preterm birth. The processed dataset is further handled using the SMOTE technique to address class imbalance.

The balanced dataset is then split into training and testing sets in a 70:30 ratio. In the modeling layer, three machine learning classifiers—Decision Tree, Logistic Regression, and Support Vector Machine—are trained and evaluated. The evaluation layer assesses model performance using metrics such as accuracy, sensitivity, specificity, precision, F1-score, and ROC-AUC. Based on the evaluation results, the best-performing model is selected. Finally, the output layer generates preterm birth risk predictions categorized as high-risk or low-risk. This prediction supports clinical decision-making and enables early intervention in rural healthcare settings.

4. Results and Discussion.

This study presented a machine learning-based predictive framework for preterm birth risk assessment using maternal and obstetrical data collected from a rural healthcare setting. The proposed approach integrates data preprocessing, discretization, and an entropy-based feature selection technique to enhance model performance and reduce computational complexity. Three classifiers—Decision Tree, Logistic Regression, and Support Vector Machine—were evaluated to identify the most effective model for PTB prediction.

4.1 Dataset Description

The experimental analysis was conducted using maternal health data collected from a rural healthcare center. After applying inclusion and exclusion criteria, a total of 1300 records were retained for the study. Among these, 309 cases (23.78%) correspond to preterm birth (PTB), while 991 cases (76.22%) represent term birth (TB). This uneven distribution highlights a typical class imbalance problem in healthcare datasets, where high-risk cases are relatively scarce. Such imbalance can bias machine learning models toward the majority class, reducing their ability to detect critical cases. To address this issue, the Synthetic Minority Oversampling Technique (SMOTE) was employed to generate synthetic samples for the minority class, resulting in a balanced dataset for improved model training.

4.2 Performance Evaluation Metrics

The performance of the machine learning models was evaluated using widely accepted classification metrics, including accuracy, sensitivity, and specificity. Accuracy measures the overall correctness of predictions, while sensitivity reflects the model's ability to correctly identify preterm birth cases. Specificity, on the other hand, indicates the model's ability to correctly classify normal (term) births. In the context of healthcare applications, sensitivity is particularly important because failing to detect a high-risk pregnancy may lead to severe clinical consequences. These metrics provide a comprehensive evaluation of model effectiveness and reliability.

4.3 Performance on Original Dataset

The performance of the classifiers on the original imbalanced dataset shows that the Support Vector Machine (SVM) achieved the highest accuracy of 86.1%, followed by Logistic Regression (LR) with 84.1% and Decision Tree (DT) with 77.7%. Although LR demonstrated the highest specificity, indicating strong performance in identifying normal births, its ability to detect preterm cases was comparatively lower. The Decision Tree model exhibited the weakest performance, particularly in terms of sensitivity, which limits its effectiveness in healthcare scenarios where accurate detection of high-risk cases is essential. These results indicate that while traditional models provide baseline performance, more advanced techniques are required for reliable prediction.

Table 2 presents the performance of the classifiers on the original dataset

Classifier	Accuracy	Sensitivity	Specificity
DT	77.70%	70.20%	93.00%
LR	84.10%	86.30%	97.10%
SVM	**86.10%**	80.10%	70.20%

4.4 Performance on Balanced Dataset

After applying SMOTE, a notable improvement in classification performance was observed across all models. The SVM classifier achieved the highest accuracy of 90.9%, along with a sensitivity of 89.1%, demonstrating its strong capability in identifying preterm birth cases. Logistic Regression also showed improved performance with an accuracy of 87.2%, while Decision Tree reached 79.6%. The improvement in sensitivity across all models confirms that balancing the dataset significantly enhances the detection of minority class instances. This is particularly valuable in healthcare applications, where early identification of high-risk pregnancies is critical for effective intervention.

Table 3. Classifier performance improves significantly

Classifier	Accuracy	Sensitivity	Specificity
DT	79.60%	71.30%	97.20%
LR	87.20%	83.20%	95.20%
SVM	**90.90%**	**89.10%**	78.30%

4.5 Accuracy Comparison Analysis

The comparative performance of the classifiers is illustrated in The figure 2 presents a bar chart comparing the accuracy of DT, LR, and SVM on both original and balanced datasets. It clearly shows that all classifiers benefit from dataset balancing, with SVM demonstrating the most significant improvement. The visual representation highlights that SVM consistently outperforms the other models, confirming its suitability for predictive modeling in preterm birth analysis. The improvement in accuracy after balancing also emphasizes the importance of addressing class imbalance in medical datasets.

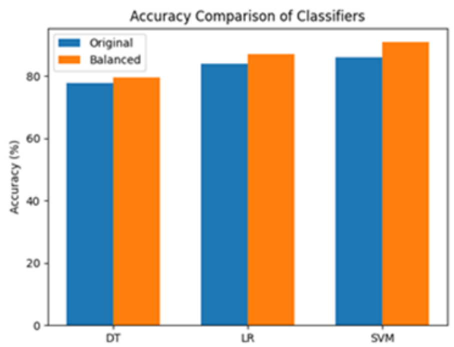

Fig. 2: Accuracy Comparison of Classifiers.

4.6 Sensitivity and Specificity Analysis

The comparison of sensitivity and specificity is presented the figure 3 illustrates that SVM achieves the highest sensitivity, making it highly effective in detecting preterm birth cases. In contrast, the Decision Tree model shows higher specificity, indicating better performance in identifying normal births. Logistic Regression provides a balanced trade-off between sensitivity and specificity. This analysis highlights that while different models have their strengths, SVM is particularly suitable for healthcare applications where minimizing false negatives is critical.

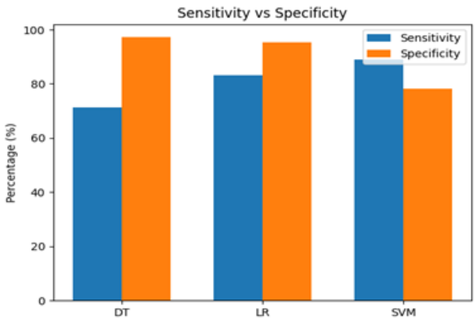

Fig. 3: Sensitivity vs Specificity of Classifiers.

4.7 Confusion Matrix Analysis

The classification performance of the SVM model is further analyzed using the confusion matrix, as shown in Figure 4. The confusion matrix demonstrates a significant increase in true positive predictions and a reduction in false negatives after applying SMOTE. This indicates that the model is more effective in identifying preterm birth cases. The improved distribution of correctly classified instances confirms the robustness of the proposed model and its ability to generalize well to unseen data. The reduction in misclassification is particularly important in clinical applications where accuracy directly impacts patient outcomes.

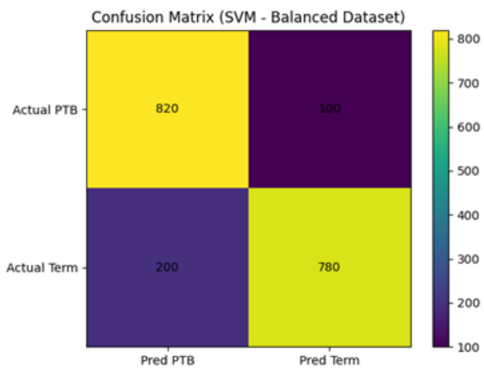

Fig. 4: Confusion Matrix of SVM (Balanced Dataset).

4.8 Comparative Discussion of Classifiers

The experimental results reveal distinct characteristics of each classifier. Decision Tree models are simple and interpretable but suffer from lower accuracy and sensitivity, making them less suitable for critical healthcare applications. Logistic Regression provides stable and balanced performance, particularly in terms of specificity, but may not capture complex nonlinear relationships in the data. Support Vector Machines, on the other hand, demonstrate superior performance across all metrics, especially after dataset balancing. Their ability to handle high-dimensional data and complex decision boundaries makes them highly effective for preterm birth prediction. These findings suggest that SVM is the most appropriate choice for this application.

4.9 Impact of Preprocessing and Feature Selection

The results also highlight the significant role of data preprocessing and feature selection in improving model performance. By removing redundant and irrelevant features, the model becomes more efficient and less prone to overfitting. Data discretization and normalization further enhance the learning process by transforming raw data into a structured format suitable for machine learning algorithms. These preprocessing steps not only improve accuracy but also reduce computational complexity, making the model more practical for real-world deployment.

4.10 Practical Implications and Discussion

The proposed machine learning framework demonstrates strong potential for real-world application in rural healthcare settings. The ability to accurately predict preterm birth using readily available maternal data can significantly reduce dependence on expensive diagnostic procedures. This is particularly beneficial in resource-constrained environments where access to advanced medical facilities is limited. The model can assist healthcare professionals in identifying high-risk pregnancies at an early stage, enabling timely intervention and reducing maternal and neonatal mortality.

Overall, the study confirms that integrating machine learning with healthcare data analytics can play a crucial role in improving clinical decision-making and patient outcomes.

5. Conclusion

This study presented a machine learning-based predictive framework for preterm birth risk assessment using maternal and obstetrical data collected from a rural healthcare setting. The proposed approach integrates data preprocessing, discretization, and an entropy-based feature selection technique to enhance model performance and reduce computational complexity. Three classifiers—Decision Tree, Logistic Regression, and Support Vector Machine—were evaluated to identify the most effective model for PTB prediction. The experimental results demonstrate that the Support Vector Machine consistently outperforms the other classifiers across all evaluation metrics. On the original dataset, SVM achieved an accuracy of 86.1%, which improved significantly to 90.9% after applying the SMOTE technique for dataset balancing. The increase in sensitivity to 89.1% further confirms the model's ability to accurately detect high-risk pregnancies, which is a critical requirement in healthcare applications. These findings emphasize the importance of addressing class imbalance and selecting relevant features to improve predictive performance. The study also highlights that feature selection plays a vital role in identifying key maternal risk factors associated with preterm birth, thereby improving both efficiency and interpretability of the model. The proposed framework reduces dependence on expensive diagnostic procedures and provides a practical solution for early detection in resource-constrained environments. Despite its promising performance, the study is limited by the size and scope of the dataset, which is restricted to a single rural healthcare center. Future work should focus on incorporating larger and more diverse datasets, exploring advanced machine learning and deep learning models, and developing real-time decision support systems for clinical use. Overall, the proposed model demonstrates strong potential for improving maternal healthcare by enabling early identification of preterm birth risk, supporting timely medical intervention, and ultimately reducing neonatal mortality and morbidity, particularly in rural and underserved regions.

References

[1]. Bengio Y (2009) Learning Deep Architectures for AI; Now Publishers Inc: Norwell, MA, USA, 1-127.

[2]. Ward RM, Beachy JC (2003) Neonatal complications following preterm birth. BJOG Int J Obstet Gynaecol 110; 8-16.

[3]. Tang Y, Zhang YQ, Chawla N, Krasser S (2008) SVMs Modeling for Highly Imbalanced Classification. IEEE Trans Syst Man Cybern Part B Cybern 39; 281-288.

[4]. Boulle M (2004) Khiops: A Statistical Discretization Method of Continuous Attributes. Machine Learning 55(1); 53-69.

[5]. Witten H, Frank E (2002) Data mining. ACM Sigmod Record 31(1): 7677.

[6]. B. M. Mercer, R. L. Goldenberg, A. Das et al., "The preterm prediction study: a clinical risk assessment system," American Journal of Obstetrics and Gynecology, vol. 174, no. 6, pp. 1885–1895, 1996.

[7]. L. K. Goodwin, M. A. Iannacchione, W. E. Hammond, P. Crockett, S. Maher, and K. Schlitz, "Data mining methods find demographic predictors of preterm birth," Nursing Re-search, vol. 50, no. 6, pp. 340–345, 2001.

[8]. C. Gao, S. Osmundson, D. R. Velez Edwards, G. P. Jackson, B. A. Malin, and Y. Chen, "Deep learning predicts extreme preterm birth from

electronic health records," Journal of Biomedical Informatics, vol. 100, p. 103334, 2019.

[9]. A. Weber, G. L. Darmstadt, S. Gruber et al., "Application of machine-learning to predict early spontaneous preterm birth among nulliparous non-Hispanic black and white women," Annals of Epidemiology, vol. 28, no. 11, pp. 783–789, 2018.

[10]. M. Mailath-Pokorny, S. Polterauer, M. Kohl et al., "Indi- vidualized assessment of preterm birth risk using two mod- ified prediction models," *European Journal of Obstetrics & Gynecology and Reproductive Biology*, vol. 186, pp. 42–48, 2015.

[11]. M. Son and E. S. Miller, "Predicting preterm birth: cervical length and fetal fibronectin," *Seminars in Perinatology*, vol. 41, no. 8, pp. 445–451, 2017.

[12]. U. Elaveyini, S. P. Devi, and K. S. Rao, "Neural networks prediction of preterm delivery with first trimester bleeding," *Archives of Gynecology and Obstetrics*, vol. 283, no. 5, pp. 971–979, 2011.

[13]. T. Khatibi, N. Kheyrikoochaksarayee, and M. M. Sepehri, "Analysis of big data for prediction of provider-initiated preterm birth and spontaneous premature deliveries and ranking the predictive features," *Archives of Gynecology and Obstetrics*, vol. 300, no. 6, pp. 1565–1582, 2019.

[14]. M. Colstrup, E. R. Mathiesen, P. Damm, D. M. Jensen, and L. Ringholm, "Pregnancy in women with type 1 diabetes: have the goals of St. Vincent declaration been met concerning foetal and neonatal complications?" *Re Journal of Maternal- Fetal & Neonatal Medicine*, vol. 26, no. 17, pp. 1682–1686, 2013.

[15]. H. Blencowe, S. Cousens, M. Z. Oestergaard et al., "National, regional, and worldwide estimates of preterm birth rates in the year 2010 with time trends since 1990 for selected countries: a systematic analysis and implications," *Re Lancet*, vol. 379, no. 9832, pp. 2162–2172, 2012.

[16]. L. Liu, H. L. Johnson, S. Cousens et al., "Global, regional, and national causes of child mortality: an updated systematic analysis for 2010 with time trends since 2000," *Re Lancet*, vol. 379, no. 9832, pp. 2151–2161, 2012.

[17]. M. A. Hall, "Correlation-based feature selection for discrete and numeric class machine learning," in *Proceedings of Sev- enteenth International Conference on Machine Learning*, pp. 359–366, Stanford, CA, USA, July 2000.

[18]. K. Y. Ngiam and I. W. Khor, "Big data and machine learning algorithms for health-care delivery," *Re Lancet Oncology*, vol. 20, no. 5, pp. e262–e273, 2019.

[19]. J. G. Nam, S. Park, E. J. Hwang et al., "Development and validation of deep learning-based automatic detection algo- rithm for malignant pulmonary nodules on chest radio- graphs," *Radiology*, vol. 290, no. 1, pp. 218–228, 2019.

[20]. Z. Ahmed, K. Mohamed, S. Zeeshan, and X. Dong, "Artificial Intelligence with Multi-Functional Machine Learning Plat- form Development for Better Healthcare and Precision Medicine," *Database*, vol. 2020, 2020.

[21]. A. Rajkomar, J. Dean, and I. Kohane, "Machine learning in medicine," *New England Journal of*

Medicine, vol. 380, no. 14, pp. 1347–1358, 2019.

[22]. S. Pereira, F. Portela, M. F. Santos, J. Machado, and A. Abelha, "Predicting preterm birth in maternity care by means of data mining," in *Portuguese Conference on Artificial Intelligence*, pp. 116–121, Springer, Berlin, Germany, 2015.

[23]. K. L. Courtney, S. Stewart, M. Popescu, and L. K. Goodwin, "Predictors of preterm birth in birth certificate data," *Studies in Health Technology and Informatics*, vol. 136, pp. 555–60, 2008.

[24]. N. S. Prema and M. P. Pushpalatha, "Machine learning ap- proach for preterm birth prediction based on maternal chronic conditions," in *Emerging Research in Electronics, Computer Science and Technology*, pp. 581–588, Springer, Berlin, Germany, 2019.

[25]. F. Cunningham, K. Leveno, S. Bloom, C. Y. Spong, and J. Dashe, Williams Obstetrics, 24e, McGraw-Hill, New York, NY, USA, 2014.