International Journal of Innovative Research and Scientific Studies, 8(4) 2025, pages: 1897-1907

IJIRSS International Journal of Innovative Research & Scientific Studies

ISSN: 2617-6548

URL: www.ijirss.com

Developing a predictive model for stillbirth risk using machine learning algorithms

Mayasee Kuanakewinyoo[1], Arthit Intarasit[2], Jarunee Saelee[3], Jirapond Muangprathub[4*]

[1]College of Digital Science, Prince of Songkla University, Songkhla, 90110, Thailand.
[2,3]Faculty of Science and Technology, Prince of Songkla University, Pattani, 94000, Thailand.
[4]Faculty of Science and Industrial Technology, Prince of Songkla University, Surat Thani, 84000, Thailand.

Corresponding author: Jirapond Muangprathub (*Email: jirapond.m@psu.ac.th*)

Abstract

Stillbirths continue to be a significant global issue, reflecting the quality of obstetric care. In Thailand, stillbirth remains a concern, especially in southern provinces like Pattani. Traditional risk identification focuses on historical obstetric complications, but there is a need for advanced technology to improve early detection and prevention. This study investigates the factors influencing stillbirth risk by analyzing perinatal data from 2018 to 2021 in Pattani Province. A machine learning-based predictive model was developed using various algorithms, including Naïve Bayes, logistic regression, deep learning, decision tree, random forest, and gradient-boosted trees, to handle imbalanced datasets. Performance comparisons of these models were conducted, and the best-performing model was implemented into a web application to provide personalized recommendations to pregnant women. The study highlights the importance of advanced data analytics in stillbirth prevention and offers an innovative, accessible solution for pregnant women to manage and monitor their pregnancies. The system not only improves the prediction of stillbirth risks but also provides an interactive, low-cost platform for delivering real-time suggestions, aiming to reduce stillbirth occurrences and improve maternal health outcomes.

Keywords: Machine learning, Pregnancies, Risk prediction, Stillbirth risk.

DOI: 10.53894/ijirss.v8i4.8259
Funding: This study received no specific financial support.
History: Received: 24 April 2025 / **Revised:** 29 May 2025 / **Accepted:** 2 June 2025 / **Published:** 3 July 2025

Competing Interests: The authors declare that they have no competing interests.
Authors' Contributions: All authors contributed equally to the conception and design of the study. All authors have read and agreed to the published version of the manuscript.
Transparency: The authors confirm that the manuscript is an honest, accurate, and transparent account of the study; that no vital features of the study have been omitted; and that any discrepancies from the study as planned have been explained. This study followed all ethical practices during writing.
Institutional Review Board Statement: The Ethical Committee of the Pattani Provincial Public Health Office, Thailand has granted approval for this study on 28 February 2024 (Ref. No. RECPTN No.038/66).
Acknowledgements: The authors are deeply grateful to the Faculty of Science and Industrial Technology, Prince of Songkla University, Surat Thani Campus, Thailand.
Publisher: Innovative Research Publishing

1. Introduction

Stillbirth refers to the demise of a baby before or during delivery, encompassing both miscarriage and cases where a baby is born without vital signs. It is categorized into early (20-27 weeks), late (28-36 weeks), or term (37 or more weeks) [1-3]. Distinctions in gestational age thresholds for stillbirth versus miscarriage depend on global standards, healthcare access, and infrastructure. High-income and some middle-income nations typically set thresholds from 18 to 22 weeks, while in low-income areas, they may extend up to 28 weeks [1, 2, 4, 5]. For international comparability, the WHO recommends using a cutoff of 1000 g or more at birth (if available) or 28 or more completed weeks of gestation [3]. From previous studies, a stillbirth can be caused by many potential factors, such as slow growth in the womb, the mother's disease, the mother's health issues, and so on. In addition, there are many studies that focus on identifying the factors causative of this problem [1, 4]. Thus, this work will explore and collect factors affecting the condition of stillbirth. These factors were then applied to predict the rate and trend of stillbirths, in order to prevent both fetal and maternal harm.

In developing countries, the risk of death during the neonatal period is six times greater than in developed countries; in the least developed countries, it is over eight times higher. With 41 neonatal deaths per 1,000 live births, the risk of neonatal death is highest in Africa; the sub-Saharan regions of Eastern, Western, and Central Africa have between 42 and 49 neonatal deaths per 1,000 live births. South-Central Asia, with 43 neonatal deaths per 1,000 live births, shows rates close to those registered in sub-Saharan Africa, while the neonatal mortality rate for Latin America and the Caribbean is 15 per 1,000 live births. Most neonatal deaths occur in Asia, where most children are born. Given the high mortality rate in the South-Central Asian subregion, it contributes over 40% of the global neonatal deaths [1, 6, 7]. Global rates and trends of stillbirth have been extensively investigated. Stillbirth is a devastating outcome for affected families and society, accounting for two-thirds of perinatal mortality [1, 2]. Current national guidelines recommend identifying women with any known risk factors for stillbirth as being at high risk. Therefore, developing strategies to prevent stillbirth is a key research priority.

In Thailand, stillbirths remain a significant issue in fetal and maternal health. Several years ago, the stillbirth rate was notably high in many regions, particularly in the southernmost provinces. Consequently, the stillbirth problem is a national concern and requires urgent attention, as it reflects the quality of obstetric services. Therefore, identifying pregnancies at high risk of stillbirth is a critical challenge that must be addressed promptly. Knowledge related to this can be obtained from service facilities, clinics, or hospitals, but it has not yet effectively reached pregnant women directly. This work presents an application that pregnant women can access directly. In this study, machine learning was employed to uncover knowledge about various factors causally related to stillbirth. The results are utilized in a recommender application for direct use by pregnant women.

2. Background

This section describes the stillbirth risk and factors, data preparation, machine learning, and validation.

2.1. Stillbirth Risk and Factors

The minimum thresholds for gestational age and birth weight that differentiate between a second-trimester miscarriage and a stillbirth have shown significant variation, both in official definitions and among different research studies. Various definitions have encompassed gestational age thresholds ranging from 20 to 28 weeks and birth weight thresholds from 350 to 1000 grams [1, 2, 4, 5]. Thus, in this current assessment, the World Health Organization's (WHO) definition of stillbirth will be adopted whenever feasible. This definition entails fetal demise at 20 weeks of gestational age or beyond, or at a birth weight exceeding 350 grams [3]. Traditionally, stillbirths are categorized into three groups: early preterm, often defined with an upper limit of 28 weeks or 1000 grams; late preterm, encompassing the range from 28 to 36 weeks; and full-term, which is 37 weeks or beyond. Additionally, stillbirths are also grouped based on their occurrence in relation to labor, namely as either antenatal or intrapartum.

Stillbirths are commonly classified based on identifiable risk factors and presumed causes. Risk factors, in the context of stillbirth, refer to maternal traits linked to this outcome without a clearly established cause. These risk factors include previous stillbirth experiences, socioeconomic disadvantages, being Black, maternal obesity, limited education, older maternal age, post-term pregnancies, diabetes, smoking, and so on [4, 7]. Significant contributors to stillbirth include congenital anomalies, infections, and instances of asphyxia associated with conditions such as preeclampsia, placental abruption, and accidents involving the umbilical cord. Conditions such as Rh disease and maternal-fetal trauma have also been considered potential causes of stillbirth. Despite thorough investigations in various studies, approximately 50% of stillbirth cases lack a definitively identified cause [8-10].

In nearly every research investigation, previous pregnancy results serve as indicators for forthcoming outcomes. This association holds true even for stillbirths, although due to the infrequent recurrence of such events, substantial sample sizes are required to establish this connection. Studies have revealed that women who experienced a stillbirth in their initial pregnancy had an elevated likelihood of giving birth to low birth weight and preterm babies in subsequent pregnancies. Additionally, they faced an eightfold increase in the occurrence of placental abruptions. In addition, a low socioeconomic status can increase the risk of stillbirth. Shahar et al. studied the factors linked with lower socioeconomic status, such as smoking, alcohol and drug use, advanced maternal age, elevated body mass index (BMI), and medical conditions [11]. However, none of these factors could fully account for the elevated stillbirth risk in this group. Given that the rise in stillbirths among economically disadvantaged women was more apparent in term or near-term pregnancies, they theorized that nuanced variations in medical care might be responsible for these outcome disparities.

In Thailand, many research studies have examined factors consistent with the above risk factors. Specifically, these risk factors include pregnancy and childbirth-related complications, extended pregnancy, maternal infections such as malaria, syphilis, and HIV, maternal health conditions (particularly high blood pressure and diabetes), and fetal growth restrictions [12]. Additionally, factors linked to maternal age and smoking increase the probability of maternal illness and stillbirth. Key demographic elements associated with an elevated risk of stillbirth have been identified. These risk factors include maternal age, body mass index (BMI), ethnicity, low birth weight of the newborn, drug use, lower socioeconomic status, and limited educational attainment. Gross national income and overall access to essential healthcare services are primary predictors [4, 7-10]. In this work, the particular investigation delves into instances of stillbirth within Pattani province, Thailand, utilizing data sourced from the Pattani Provincial Public Health Office spanning the period from 2018 to 2021. Centered on childbirth data and expectant mothers, the study hypothesis revolves around the influence of maternal well-being on the developing fetus. The risk factors for stillbirth derived from the dataset consist of the level of education, age of mother, prepregnancy weight, prepregnancy height, BMI, gravida, maternal infection and parasitic diseases, hematocrit level, chronic disease hypertension, and chronic disease diabetes mellitus. These factors are used to quantify the risk for pregnant women, based on hospital services information.

2.2. Data Preparation

Large real-world databases and data warehouses often contain incomplete, noisy, and inconsistent data. The incompleteness can arise from various reasons, including missing attributes, data not considered important during entry, misunderstandings, equipment malfunctions, and overlooked history or modifications [13]. Noisy data, characterized by incorrect attribute values, may result from faulty data collection instruments, human or computer errors, transmission issues, technological limitations, or inconsistencies in naming conventions and data codes. Data cleaning is necessary to address duplicate tuples and infer missing values for tuples with incomplete data. Preprocessing and data cleaning play crucial roles in ensuring data quality and accuracy for meaningful analysis and reliable decision-making.

Data preparation is a crucial step in the data analysis process, involving the collection, cleaning, and transformation of raw data into a structured format suitable for analysis. This process ensures that the data is accurate, complete, and free from errors or inconsistencies. Data preparation also involves handling missing values, encoding categorical variables, and normalizing data to ensure uniformity and comparability. Proper data preparation is essential for obtaining reliable and meaningful insights during data analysis and model building in various fields, including machine learning, data science, and business analytics [13, 14].

Data preparation involves a series of steps to transform raw data into a format suitable for analysis and modeling. The key steps in data preparation are as follows:

(1) Data Collection: Gather data from various sources, such as databases, files, APIs, or by web scraping.

(2) Data Cleaning: Identify and handle missing values, duplicate entries, and outliers. This step ensures the data's accuracy and consistency.

(3) Data Integration and Transformation: Combine data from multiple sources and consolidate it into a single dataset for analysis, then convert data into a standardized format, for example, by encoding categorical variables, normalizing numerical data, and scaling features.

(4) Data Reduction: Reduce the dimensionality of the dataset to focus on the most important features and optimize computational efficiency.

2.3. Machine Learning

Machine learning is a field of artificial intelligence that focuses on developing algorithms and models capable of learning patterns and making predictions from data without explicit programming [15]. It involves training machines on historical data to enable them to generalize and make informed decisions or predictions on new and unseen data. Machine learning has numerous applications across various industries, from image recognition and natural language processing to healthcare and finance [16]. It continues to advance and revolutionize how computers process and understand information, making it a powerful tool for solving complex problems and enhancing decision-making processes.

Machine learning plays a vital role in the expanding domain of data science, utilizing statistical methods to train algorithms for making classifications and predictions. These valuable insights drive decisions and contribute to crucial growth metrics. As the volume of big data continues to grow, the demand for data scientists will rise, as they play a key role in identifying pertinent business questions and the corresponding data to address them [14]. In this study, we employ various approaches, including Naive Bayes, logistic regression, deep learning, decision trees, random forests, gradient-boosted trees, and support vector machines (SVMs), to build our learning model.

2.4. Evaluation Metrics

Various evaluation metrics are available to assess the performance of an artificial classifier system. While accuracy is commonly used, it should be used with caution. For instance, classifiers like logistic regression may underestimate the probability of rare events significantly, leading to accurate predictions overall but poor predictions for uncommon classes, resulting in a high (but misleading) accuracy score [17]. Accuracy is calculated by dividing the number of correct predictions by the total number of predictions, which can bias the metric towards majority classes in imbalanced datasets. Therefore, accuracy may not be the most suitable metric when dealing with imbalanced datasets [18, 19].

Hence, when dealing with imbalanced datasets, it is more appropriate to use evaluation metrics such as the receiver operating characteristic curve (ROC) and the area under the ROC curve (AUC) instead of accuracy [20, 21]. ROC and

AUC are considered excellent indicators of classifier performance and are visually informative. The ROC curve illustrates model performance by plotting the true positive rate (TPR) against the false positive rate (FPR). TPR, also known as recall, is calculated using equation (1), with TP representing true positives and FN being false negatives. Similarly, FPR is determined by equation (2), involving false positives (FP) and true negatives (TN). AUC, on the other hand, summarizes the performance based on the ROC curve and is valuable for making comparisons between models. A high AUC score signifies superior model performance, while a score close to 0.5 indicates performance comparable to random guessing [22].

$$TPT = \frac{TP}{TP + FN} = Recall. \tag{1}$$

$$FPR = \frac{FP}{FP + FN}. \tag{2}$$

While ROC and AUC are valuable for performance evaluation, they may overestimate the performance of an artificial system when dealing with imbalanced datasets. To address this, it is advisable to complement ROC with other metrics like F-measure, recall, and precision. Recall, which is given by TPT in Equation (1), assesses the correct classification of the positive class by the algorithm. On the other hand, precision in Equation (3) measures the accuracy of the algorithm by examining the correct labeling of positive instances. By using these two metrics, F-measure in Equation (4) can be computed, offering a measure of classification effectiveness that provides more insights than accuracy [23]. The combination of precision and recall is particularly useful for evaluating classifier performance in imbalanced datasets [18, 19].

$$Precision = \frac{TP}{FP + TP}. \tag{3}$$

$$F - measure = \frac{2 \times (Precision \times Recall)}{Precision + Recall}. \tag{4}$$

The classification performance can be visualized using a confusion matrix, shown in Table 1, which provides a clear display of the number counts of true positives (TP), false positives (FP), false negatives (FN), and true negatives (TN) after validation. This matrix offers an easy overview of the algorithm's performance and serves as a complementary tool to other evaluation metrics [24].

Table 1.
An illustrative model of a confusion matrix depicting its usage and parameters.

		Actual Class	
		Positive (P)	**Negative (N)**
Predicted Class	Positive (P)	True Positive (TP)	False Positive (FP)
	Negative (N)	False Negative (FN)	True Negative (TN)

3. Related Works

This section briefly reviews interesting previous works on e-Health systems that involve fetal and maternal concerns. e-Health applications are critically important, especially for patients unable to see a doctor or other health professionals.

A predictive model using machine learning was presented for early assessment of adverse birth risk among pregnant women, as a means to help improve the allocation of social services Pan et al. [25], Bertini et al. [26] and Islam et al. [27]. Akbulut et al. [28] proposed a prediction system with assistive e-Health applications, which both pregnant women and practitioners can utilize. The authors compare the performance of various ML techniques with 96 pregnant women and use these methods to process data to predict fetal anomaly status based on maternal and clinical data, aiming to identify a well-performing classification model [28]. Next, fetal health was improved by using machine learning that made factor-related recommendations to pregnant women [29, 30].

Moreover, Malacova et al. [9] attempted to select the best ML model for building a predictive model of stillbirth. The authors found that almost half of the stillbirths could be potentially identified antenatally based on a combination of current pregnancy complications, congenital anomalies, maternal characteristics, and medical history. Malacova et al. [9]. Koivu and Sairanen [8] used ML methods to predict early stillbirth, late stillbirth and preterm birth pregnancies [8]. The stillbirth risk prediction was focused on preventing and protecting fetal health, Khatibi et al. [5] and Raza et al. [31]. Mohammed et al. [32] have experimented with two resampling techniques: oversampling and under-sampling, using a public imbalanced dataset from the Kaggle website in Santander Customer Transaction Prediction, and have applied a group of machine learning algorithms with different hyperparameters to obtain the best results with both resampling techniques [32]. It is common for medical data to exhibit an imbalance of categories. Therefore, many studies have attempted to develop methods to address this characteristic in order to obtain appropriate models [18, 33, 34]. Oversampling and undersampling will be applied in this current study.

Based on the above review of prior research, this study will compare machine learning techniques for finding a well-performing model in fetal health prediction. In addition, there are many techniques in machine learning to build predictive models, such as a Bayes minimum error rate classifier [35], SVM [36], Random Forest [37] ANN [38] or Decision tree [38]. We expect to apply ML successfully to building a predictive model. The performance of ML depends on the data set, and although ML is effective in modeling for classification, only a few studies have applied it to provide recommendations or suggestions directly to pregnant women. Thus, this work compares the ML approaches Naive Bayes, logistic regression, deep learning, decision tree, random forest, and gradient-boosted trees. These are popular machine learning classifiers. Subsequently, the obtained model was used to develop a web application for suggestions and decision support, aimed at risk prevention and the protection of fetal health and pregnant women.

4. Research Methodology

4.1. Dataset Description and Tools

In Thailand, stillbirth remains a major problem in maternal and child health, as shown in Figure 1. This figure illustrates the level of stillbirths by area. We found that the primary high-risk stillbirth areas consist of three regions: north, middle, and south Thailand. Notably, the stillbirth area in southern Thailand has continued to exhibit a high risk over the past three years. Therefore, this study focused on the southern peninsular region of Thailand, with Pattani province selected as the case study area for this research.

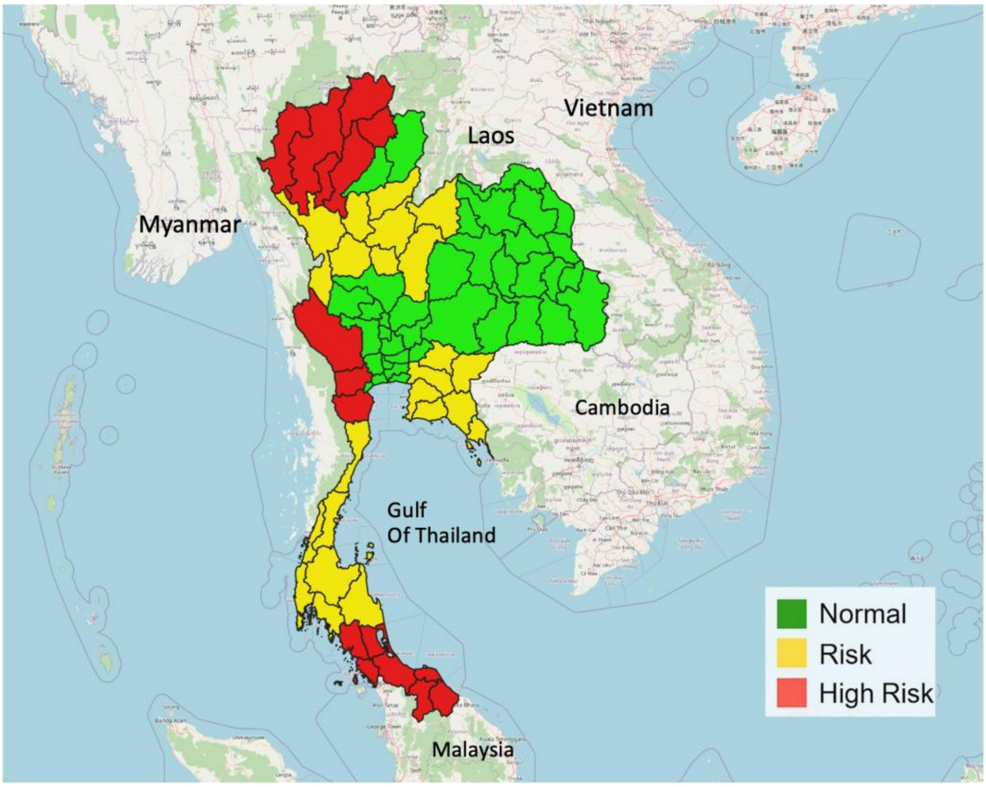


Figure 1.
The stillbirth rates by area in 2020.
Source: Ministry of Public Health [12].

Regarding stillbirths in Pattani province, the statistics for 2018-2021 are summarized in Table 2. The rate exceeds the threshold set by the Ministry of Public Health, which requires no more than 4.00 stillbirths per 1,000 births. For this reason, this province is classified as having a high risk of stillbirths. This presents a key challenge for research to prevent stillbirths as a problem for pregnant women.

Table 2.
The stillbirth statistics for Pattani province.

Year	Stillbirths	Total births	Stillbirth Rate (per 1,000 total births)
2021	53	8,252	6.42
2020	43	8,809	4.89
2019	63	9,965	6.32
2018	58	9,633	6.02
Summary	217	36,659	5.91

Table 2 was derived from childbirth information from the Pattani Provincial Public Health Office, Pattani province, Thailand. Baseline data on pregnant women and childbirths were retrieved from the records of health centers or district hospitals and stored in the database center of Pattani Provincial Public Health Office. The childbirth data involves various factors affecting the stillbirth events, referred to as pregnancy data in foreign key form. However, a problem with these collected data is the missing values. Thus, the relevant data on childbirths and pregnancies has 31,149 birth cases, including 217 stillbirths and 30,932 live births. The dataset content is presented in Table 3. Normally, the main causes of stillbirth are classified into three issues, namely (1) the fetus having chromosomal abnormalities, (2) maternal placental issues, and (3) maternal health problems. This work focuses on the last of these three. The research aims to study and explore the factors affecting the conditions causing stillbirths related to maternal health. Moreover, this work seeks to apply the best machine learning techniques to provide suggestions to pregnant women through the developed application.

In this study, we utilized the widely recognized software, RapidMiner Studio, which offers a comprehensive range of machine learning algorithms for data analysis tasks. RapidMiner Studio is a versatile platform encompassing data preparation, machine learning, deep learning, text mining, and predictive analytics. It serves various purposes such as research, education, training, rapid prototyping, and application development, supporting all stages of the machine learning process, including data preparation, result visualization, model validation, and optimization. For this research, we employed RapidMiner Studio to evaluate and compare the performances of different classification methods, aiming to select the most suitable model for the development application. Subsequently, the chosen model will be integrated into a web-based application, developed using the PHP language, while the MySQL database will store the information related to pregnant women.

Table 3.
The content of the dataset.

Column Name	Detail	Value
Education	Level of education	00 = No study / no degree attained 01= Pre-primary 02 = Primary (elementary school) 03 = Junior high school education/ High school education 04 = Vocational certificate/ Junior diploma/ 05 = Diploma /Graduate 06 = Postgraduate 09 = Unknown
momage	Age of the mother	years
WEIGHT_before	Weight before prepregnancy	kilogram
HEIGHT_before	Height before prepregnancy	Centimeters
BMI	Body mass index (BMI) is a measure of body fat based on height and weight that applies to adult men and women.	Kg/m^2
Gravida	The number of times a woman has been pregnant	Number of times
Infect	Infections in pregnant maternal infection and parasitic diseases, sepsis during labor, and other infections during labor	0 = No infection 1 = infection
HCT_RESULT	Hematocrit level among pregnant women (%)	percent
HT	Chronic disease hypertension	0 = No disease 1 = hypertension (ht)
DM	Chronic disease, diabetes mellitus	0 = No disease 1 = diabetes mellitus (dm)
Birth_st	Stillbirth refers to a baby who dies after 28 weeks of pregnancy. Livebirth is the term used for a newborn who breathes or shows other signs of life after birth.	-stillbirth -livebirth

4.2. The Proposed Stillbirth Risk Prediction System

This research employs system development based on the System Development Life Cycle (SDLC) and CRISP-DM (Cross Industry Standard Process for Data Mining). There are six key steps in this approach, with details as follows.

1. Planning and data preparation: This step involves studying and collecting data on the topic, including background information, related works, exploration of stillbirth data in the study, and machine learning methods with applications in stillbirth. The work is planned to experiment with data from the study area of Pattani province, Thailand.
2. System design and analysis: The requirements for the presented system were derived from a literature review and stakeholders in the study area. The presented predictive model was designed and analyzed according to requirements for directly providing recommendations to pregnant women.

3. Building of the model and its development: The machine learning approach is applied in this step to discover stillbirth knowledge from derived real-world data in the study area. Next, the system will be created in a web application form. An overview of the proposed model is shown in Figure 2.

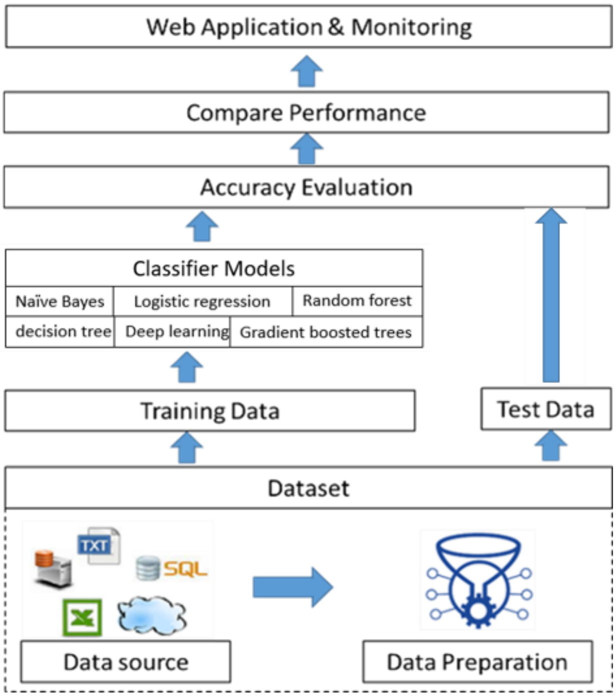


Figure 2.
The proposed framework for stillbirth risk prediction.

4. Testing and evaluation: This step involves testing and evaluating both the methodology used in machine learning and the developed system.
5. Operationalizing and deployment: This step demonstrates the value and benefits of the presented system in the real-world study area.

For building the model and its development, the presented data model refers to a mathematical or computational representation of data and the relationships between variables using machine learning models, to build a predictive model from data after the preparation step. This work applied six algorithms in machine learning to select the best model for stillbirths in the pregnancy dataset. These models consist of Naive Bayes, logistic regression, deep learning, decision tree, random forest, and gradient-boosted trees. The process of the proposed system to obtain the best model is shown in Figure 3. After preparing datasets of three types, original, under-sampled, and over-sampled, we randomly split each dataset into training (70%) and testing (30%) data to evaluate the performance of each classification algorithm.

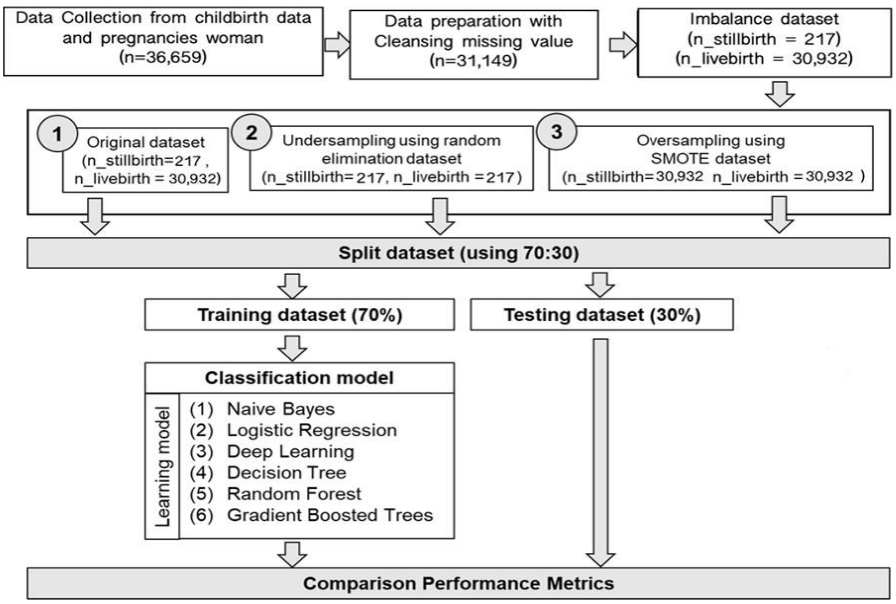
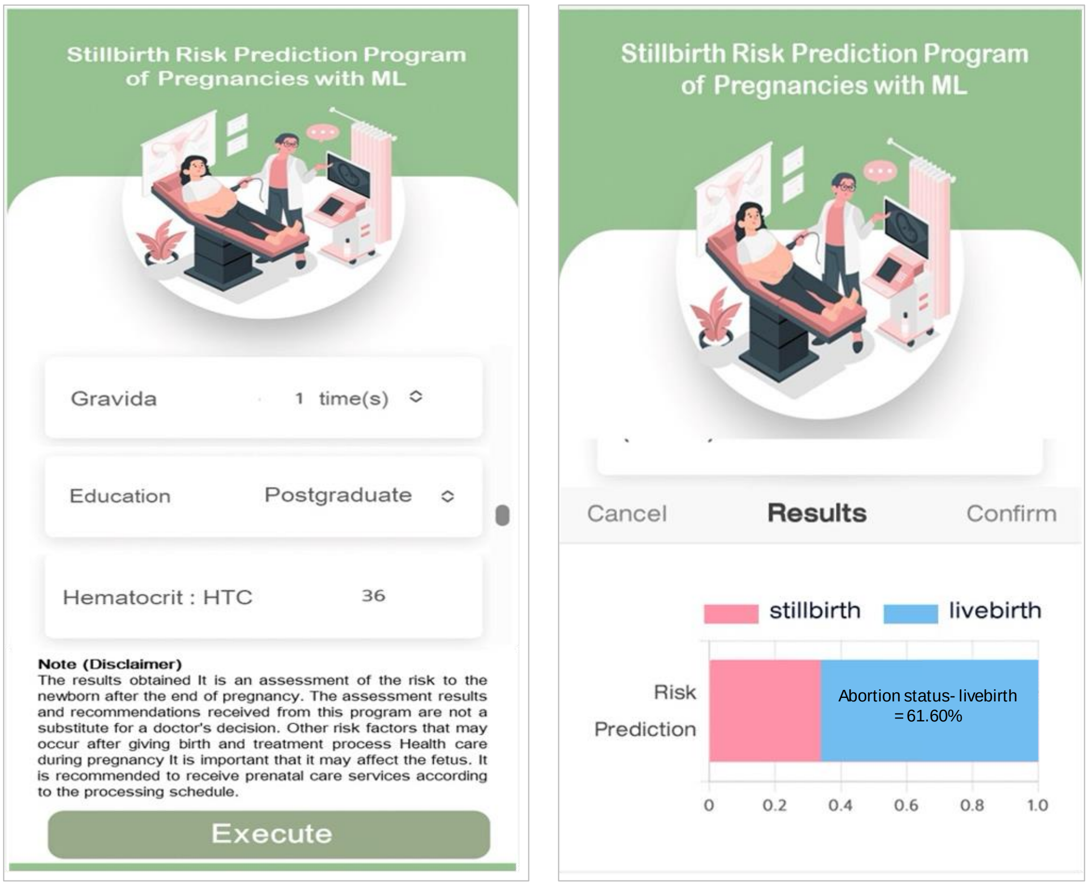


Figure 3.
The proposed data model and evaluation using machine learning.

Following this, the proposed system is set to be executed through a web application. The design and development of the web application are illustrated in Figure 4. This figure provides an illustrative example of the application in action. Users have the capability to input details regarding pregnant women and relevant data. Our system has been crafted as an online website, employing the PHP language for implementation, and utilizing a MySQL database to store the detailed information obtained through the machine learning process mentioned in the preceding section.



(a) Input data in English version (b) Result prediction in English version

Figure 4.
An example of web application.

5. Results and Discussion

The outcomes generated by various machine learning algorithms were classified into three distinct groups. These groups consist of the outcomes derived from both the initial, imbalanced dataset, the outcomes obtained from the undersampled dataset, and, lastly, the outcomes originating from the oversampled dataset. Each of these approaches underwent training utilizing six distinct classifiers: Naive Bayes, logistic regression, deep learning, decision tree, random forest, and gradient-boosted trees. Subsequently, these models were assessed using a set of evaluation metrics, including accuracy, recall, F-measure, precision, and AUC.

5.1. Imbalanced Dataset Performance

The dataset in its pristine form was utilized, devoid of any sampling modifications. The target class exhibited an imbalance, with the majority class accounting for 99.30% and the minority class for 0.70% of instances. The conclusive evaluation scores, achieved through training (70%) and testing (30%), are presented in Table 4. This table encompasses metrics such as F1-measure, recall, precision, AUC score, and accuracy.

The outcome reveals a clear contrast among the metrics (refer to Table 4). A superficial analysis based solely on accuracy might suggest promising results. However, a closer examination of precision and recall reveals notably poorer performance. Consequently, this has led to inaccurate predictions, primarily misclassifying stillbirth cases as normal pregnancies (or live births).

Table 4.
Evaluation metrics for the six classifiers trained with the original imbalanced dataset.

Model	Accuracy	Recall	F Measure	Precision	AUC
Naive Bayes	99.3	0.0	Unknown	Unknown	0.61
Logistic Regression	99.3	0.0	Unknown	Unknown	0.56
Deep Learning	99.3	0.0	Unknown	Unknown	0.58
Decision Tree	98.0	0.0	Unknown	Unknown	0.50
Random Forest	99.0	4.7	Unknown	Unknown	0.80
Gradient Boosted Trees	99.2	0.0	Unknown	Unknown	0.79

5.2. Under Sampling Performance

Following the application of undersampling to the initial imbalanced dataset, 434 entries remained. Exactly half of the dataset was designated as adverse birth outcomes (217 instances of stillbirth), while the other 50% (217 instances of live births) represented healthy pregnancies. The outcome of training (70%) and testing (30%) on the undersampled dataset for the six classifiers is presented in Table 5.

Table 5.
Evaluation metrics for the six classifiers with the undersampling dataset.

Model	Accuracy	Recall	F Measure	Precision	AUC
Naive Bayes	60.18	68.48	58.89	56.11	0.61
Logistic Regression	57.49	21.58	33.33	54.7	0.58
Deep Learning	55.91	41.21	49.02	63.33	0.65
Decision Tree	83.04	84.75	82.25	81.81	0.81
Random Forest	83.04	84.31	82.95	82.42	0.83
Gradient Boosted Trees	84.0	92.5	84.8	78.6	0.913

5.3. Oversampling Performance

Upon oversampling the initial dataset, it encompassed 30,932 records and achieved a balanced distribution of the target class. The evaluation scores resulting from this process are displayed in Table 6. By employing SMOTE, synthetic data points were generated for the underrepresented class, resulting in a balanced dataset. This action expanded the total data points from 31,149 to 61,864 records. The combined effect of increased dataset size and balanced distribution led to the most favorable outcome among the three techniques.

Table 6.
Evaluation metrics for the six classifiers with the oversampled dataset.

Model	Accuracy	Recall	F Measure	Precision	AUC
Naive Bayes	56.5	35.3	44.8	71.3	0.63
Logistic Regression	55.0	23.3	34.1	63.7	0.61
Deep Learning	84.8	74.9	83.1	93.3	0.88
Decision Tree	70.5	80.5	73.2	67.1	0.70
Random Forest	82.2	80.3	81.8	83.4	0.86
Gradient Boosted Trees	58.2	85.1	67.0	55.3	0.53

6. Conclusion

The purposes of this study were to identify key factors influencing the occurrence of stillbirths in pregnant women, with the aim of reducing these events, even if their prevalence is low, and to protect the population. Machine learning techniques were applied to develop a well-performing classification model for stillbirth risk prediction. This experimental study utilized a four-year dataset from a government agency. The data were used to build a predictive model; however, the dataset's class imbalance, common in the medical field, posed a challenge. To find the most effective classifier, three experimental approaches were employed: imbalanced, undersampled, and oversampled data. The original dataset was imbalanced, while undersampling and oversampling involved randomly sampling to match the size of the smallest class and applying the SMOTE method, respectively. These datasets were randomly split into 70% for training and 30% for testing. The training data were used to develop predictive models using six machine learning techniques: Naive Bayes, logistic regression, deep learning, decision tree, random forest, and gradient-boosted trees. The random forest model was selected due to its strong performance across most experimental cases. Subsequently, this model was used to create a web application that provides recommendations to pregnant women to help them take care of themselves and reduce the risk of stillbirths.

References

[1] E. M. McClure, S. Saleem, O. Pasha, and R. L. Goldenberg, "Stillbirth in developing countries: A review of causes, risk factors and prevention strategies," *Journal of Maternal-Fetal and Neonatal Medicine*, vol. 22, no. 3, pp. 183-190, 2009. https://doi.org/10.1080/14767050802559129

[2] R. Townsend *et al.*, "Can risk prediction models help us individualize stillbirth prevention? A systematic review and critical appraisal of published risk models," *BJOG: An International Journal of Obstetrics & Gynaecology*, vol. 128, no. 2, pp. 214–224, 2021. https://doi.org/10.1111/1471-0528.16677

[3] World Health Organization (WHO), "Stillbirth, Maternal, newborn, child and adolescent health, Data, statistics and epidemiology," Retrieved: https://www.who.int/health-topics/stillbirth. [Accessed 2023.

[4] C.-C. Duan, X.-H. Zhang, S.-S. Li, W. Wu, L.-Q. Qiu, and J. Xu, "Risk factors for stillbirth among pregnant women infected with syphilis in the Zhejiang Province of China, 2010–2016," *Canadian Journal of Infectious Diseases and Medical Microbiology*, vol. 2021, no. 1, p. 8877962, 2021. https://doi.org/10.1155/2021/6696587

[5] T. Khatibi, E. Hanifi, M. M. Sepehri, and L. Allahqoli, "Proposing a machine-learning based method to predict stillbirth before and during delivery and ranking the features: Nationwide retrospective cross-sectional study," *BMC Pregnancy and Childbirth*, vol. 21, pp. 1-17, 2021. https://doi.org/10.1186/s12884-021-03939-5

[6] C. J. Hogue *et al.*, "A population-based case-control study of stillbirth: The relationship of significant life events to the racial disparity for African Americans," *American Journal of Epidemiology*, vol. 177, no. 8, pp. 755–767, 2013. https://doi.org/10.1093/aje/kws337

[7] K. Afshan, G. Narjis, and M. Qayyum, "Risk factors and causes of stillbirths among pregnant women in Pakistan," *African Health Sciences*, vol. 19, no. 1, pp. 1507–1516, 2019. https://doi.org/10.4314/ahs.v19i1.6

[8] A. Koivu and M. Sairanen, "Predicting risk of stillbirth and preterm pregnancies with machine learning," *Health Information Science and Systems*, vol. 8, no. 1, p. 14, 2020. https://doi.org/10.1007/s13755-020-0307-3

[9] E. Malacova *et al.*, "Stillbirth risk prediction using machine learning for a large cohort of births from Western Australia, 1980–2015," *Scientific Reports*, vol. 10, no. 1, p. 5354, 2020. https://doi.org/10.1038/s41598-020-74785-x

[10] R. Surendiran, R. Aarthi, M. Thangamani, S. Sugavanam, and R. Sarumathy, "A systematic review using machine learning algorithms for predicting preterm birth," *International Journal of Engineering Trends and Technology*, vol. 70, no. 5, pp. 46-59, 2022. https://doi.org/10.14445/22315381/IJET-V70I5P107

[11] S. Shahar, D. Vanoh, A. F. Mat Ludin, D. K. A. Singh, and T. A. Hamid, "Factors associated with poor socioeconomic status among Malaysian older adults: an analysis according to urban and rural settings," *BMC Public Health*, vol. 19, pp. 1-12, 2019. https://doi.org/10.1186/s12889-019-7102-1

[12] Ministry of Public Health, "Stillbirth," Retrieved: https://dashboard.anamai.moph.go.th/dashboard/stillbirth, 2023.

[13] J. Brownlee, *Data preparation for machine learning: Data cleaning, feature selection, and data transforms in Python*. Melbourne, Australia: Machine Learning Mastery, 2020.

[14] J. Hurwitz and D. Kirsch, *Machine learning for dummies (IBM Limited Edition)*. Hoboken, NJ: John Wiley & Sons, 2018.

[15] D. Mennickent *et al.*, "Machine learning applied in maternal and fetal health: A narrative review focused on pregnancy diseases and complications," *Frontiers in Endocrinology*, vol. 14, p. 1130139, 2023. https://doi.org/10.3389/fendo.2023.1130139

[16] J. Waring, C. Lindvall, and R. Umeton, "Automated machine learning: Review of the state-of-the-art and opportunities for healthcare," *Artificial Intelligence in Medicine*, vol. 104, p. 101822, 2020. https://doi.org/10.1016/j.artmed.2020.101822

[17] D. Sun, J. Xu, H. Wen, and D. Wang, "Assessment of landslide susceptibility mapping based on Bayesian hyperparameter optimization: A comparison between logistic regression and random forest," *Engineering Geology*, vol. 281, p. 105972, 2021. https://doi.org/10.1016/j.enggeo.2020.105972

[18] G. Aguiar, B. Krawczyk, and A. Cano, "A survey on learning from imbalanced data streams: taxonomy, challenges, empirical study, and reproducible experimental framework," *Machine Learning*, vol. 113, no. 7, pp. 4165-4243, 2024. https://doi.org/10.1007/s10994-023-06191-9

[19] P. Kumar, R. Bhatnagar, K. Gaur, and A. Bhatnagar, "Classification of imbalanced data: Review of methods and applications," presented at the IOP Conference Series: Materials Science and Engineering, 2021.

[20] C. X. Ling, J. Huang, and H. Zhang, "AUC: A better measure than accuracy in comparing learning algorithms," in *Advances in Artificial Intelligence: 16th Conference of the Canadian Society for Computational Studies of Intelligence, AI 2003, halifax, Canada, June 11–13, 2003, Proceedings 16*, 2003: Springer, pp. 329-341.

[21] A. M. Carrington *et al.*, "Deep ROC analysis and AUC as balanced average accuracy, for improved classifier selection, audit and explanation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 45, no. 1, pp. 329-341, 2022. https://doi.org/10.1109/TPAMI.2021.3052833

[22] C. Halimu, A. Kasem, and S. S. Newaz, "Empirical comparison of area under ROC curve (AUC) and Mathew correlation coefficient (MCC) for evaluating machine learning algorithms on imbalanced datasets for binary classification," in *Proceedings of the 3rd International Conference on Machine Learning and Soft Computing*, 2019, pp. 1-6.

[23] P. Christen, D. J. Hand, and N. Kirielle, "A review of the F-measure: its history, properties, criticism, and alternatives," *ACM Computing Surveys*, vol. 56, no. 3, pp. 1-24, 2023.

[24] G. Zeng, "On the confusion matrix in credit scoring and its analytical properties," *Communications in Statistics-Theory and Methods*, vol. 49, no. 9, pp. 2080-2093, 2020. https://doi.org/10.1080/03610926.2020.1735961

[25] I. Pan *et al.*, "Machine learning for social services: A study of prenatal case management in Illinois," *American Journal of Public Health*, vol. 107, no. 6, pp. 938–944, 2017. https://doi.org/10.2105/AJPH.2017.303707

[26] A. Bertini, R. Salas, S. Chabert, L. Sobrevia, and F. Pardo, "Using machine learning to predict complications in pregnancy: A systematic review," *Frontiers in Bioengineering and Biotechnology*, vol. 9, p. 780389, 2022. https://doi.org/10.3389/fbioe.2021.780389

[27] M. N. Islam, S. N. Mustafina, T. Mahmud, and N. I. Khan, "Machine learning to predict pregnancy outcomes: A systematic review, synthesizing framework and future research agenda," *BMC Pregnancy and Childbirth*, vol. 22, no. 1, p. 348, 2022. https://doi.org/10.1186/s12884-022-04979-2

[28] A. Akbulut, E. Ertugrul, and V. Topcu, "Fetal health status prediction based on maternal clinical history using machine learning techniques," *Computers in Methods and Programs in Biomedicine*, vol. 163, pp. 87–100, 2018. https://doi.org/10.1016/j.cmpb.2018.06.007

[29] A. Mehbodniya *et al.*, "Fetal health classification from cardiotocographic data using machine learning," *Expert Systems*, vol. 39, no. 6, p. e12899, 2022. https://doi.org/10.1111/exsy.12899

[30] J. Tao, Z. Yuan, L. Sun, K. Yu, and Z. Zhang, "Fetal birthweight prediction with measured data by a temporal machine learning method," *BMC Medical Informatics and Decision Making*, vol. 21, pp. 1-10, 2021. https://doi.org/10.1186/s12911-021-01458-x

[31] A. Raza, H. U. R. Siddiqui, K. Munir, M. Almutairi, F. Rustam, and I. Ashraf, "Ensemble learning-based feature engineering to analyze maternal health during pregnancy and health risk prediction," *Plos one*, vol. 17, no. 11, p. e0276525, 2022. https://doi.org/10.1371/journal.pone.0276525

[32] R. Mohammed, J. Rawashdeh, and M. Abdullah, "Machine learning with oversampling and undersampling techniques: overview study and experimental results," presented at the 2020 11th International Conference on Information and Communication Systems (ICICS), 2020.

[33] S. Shaikh, L. Changan, and M. R. Malik, "An empirical and comparative research on under-sampling & over-sampling defect-prone data-sets model in light of machine learning," *International Journal of Advanced Networking and Applications*, vol. 12, no. 5, pp. 4719-4724, 2021.

[34] T. Wongvorachan, S. He, and O. Bulut, "A comparison of undersampling, oversampling, and SMOTE methods for dealing with imbalanced classification in educational data mining," *Information*, vol. 14, no. 1, p. 54, 2023. https://doi.org/10.3390/info14010054

[35] A. R. Yarlapati, S. R. Dey, and S. Saha, "Early prediction of LBW cases via minimum error rate classifier: A statistical machine learning approach," presented at the 2017 IEEE International Conference on Smart Computing (SMARTCOMP), 2017.

[36] S. Yu, K. K. Tan, B. L. Sng, S. Li, and A. T. H. Sia, "Feature extraction and classification for ultrasound images of lumbar spine with support vector machine," in *Proceedings of the 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), August 2014. https://doi.org/10.1109/EMBC.2014.6944351*, 2014.

[37] H. Sahin and A. Subasi, "Classification of the cardiotocogram data for anticipation of fetal risks using machine learning techniques," *Applied Soft Computing*, vol. 33, pp. 231–238, 2015. https://doi.org/10.1016/j.asoc.2015.03.033

[38] M. L. Huang and Y. Y. Hsu, "Fetal distress prediction using discriminant analysis, decision tree, and artificial neural network," *Computers in Biology and Medicine*, vol. 107, no. 3, pp. 280–287, 2012. https://doi.org/10.1016/j.cmpb.2012.01.004