



# A Machine Learning-Based Risk Prediction Model During Pregnancy in Low-Resource Settings <sup>†</sup>

Kapil Tomar <sup>1,\*</sup>, Chandra Mani Sharma <sup>1</sup>, Tanisha Prasad <sup>2</sup> and Vijayaraghavan M. Chariar <sup>1</sup>

<sup>1</sup> Indian Institute of Technology Delhi, New Delhi 110016, India; cmsharma.cdd@gmail.com (C.M.S.); vijayaraghavan.m.chariar@rdat.iitd.ac.in (V.M.C.)

<sup>2</sup> Royal College of Surgeons in Ireland, University of Medical and Health Sciences, D02 YN77 Dublin, Ireland; prasadtanisha1701@gmail.com

\* Correspondence: kkapiltomar@gmail.com

<sup>†</sup> Presented at the 2nd International One Health Conference, Barcelona, Spain, 19–20 October 2023.

**Abstract:** Maternal health is a serious concern for many nations due to a lack of appropriate healthcare facilities, healthcare staff, and late diagnoses of life-threatening diseases. Pregnant women suffer with numerous challenges during the pregnancy and childbirth. Non-communicable diseases, a lack of nutrition in diets, and unawareness of the risks associated with pregnancy are the primary reasons for these challenges. Sometimes these reasons become a direct cause of maternal mortality as well. Awareness of the risks and early detection may contribute to a reduction in maternal deaths during pregnancy and childbirth. Various ICTs have been incorporated into the healthcare industry to diagnose the issue as quickly as is feasible and an appropriate remedy can be initiated to treat diseases. Machine Learning (ML) techniques have the potential to predict the probable risk factors for timely interventions; however, challenge arises when the data are limited and unstructured. The Decision Tree (DT), Naive Bayes (NB), Logistic Regression (LR), Support Vector Machine (SVM), K-Nearest Neighbors (KNN), and Linear Discriminant Analysis (LDA) algorithms, with 10-fold cross validation, are used in this study. The dataset utilized in this study included both the present and past medical histories and important vitals of pregnant women. With a test score of 98.8%, the Decision Tree (DT) algorithm outperformed other algorithms, according to the results. Based on the predicted result, pregnant women can consult with medical specialists for their consultation to reduce the potential difficulties in the near future.

**Keywords:** maternal health; rural areas; high-risk pregnancy; machine learning



**Citation:** Tomar, K.; Sharma, C.M.; Prasad, T.; Chariar, V.M. A Machine Learning-Based Risk Prediction Model During Pregnancy in Low-Resource Settings. *Med. Sci. Forum* **2024**, *25*, 13. <https://doi.org/10.3390/msf2024025013>

Academic Editor: Margherita Ferrante

Published: 18 November 2024



**Copyright:** © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

The United Nations (UN) has made maternal health a top focus (UN). To focus on maternal health around the world, the Millennium Development Goals (MDG) were established in 2000, followed by the Sustainable Development Goals (SDG) in 2015 [1]. Maternal mortality rates decreased significantly, but not totally.

Any risk during pregnancy can be detrimental to the health of the mother and unborn child. Although all pregnancies bear risk and necessitate care, a high-risk pregnancy (HRP) always necessitates additional care during and after delivery [2]. Depending on an expectant woman's medical history, any pregnancy may become high-risk as it progresses. Pregnancies at high risk can be avoided if complications are identified early. Regular checkups before and during pregnancy aid many women in having a risk-free pregnancy and delivery free of significant complications [3].

The objective of this research is to evaluate several machine learning algorithms using a primary dataset in order to determine the most effective approach. This study introduces an innovative prediction model that use machine learning algorithms to identify high-risk pregnancies in rural areas of India. The evaluation of the model's performance was

conducted by assessing its accuracy, sensitivity, and specificity in predicting high-risk events. This approach will assist Accredited Social Health Activists (ASHA) and other frontline healthcare workers (FLHW) in detecting high-risk pregnancies by analyzing symptoms and vital signs, including body temperature, blood pressure, heart rate, fetal heart rate, and breathing rate, in order to ensure prompt and appropriate action. Later, based on the predicted risk, the expectant woman can receive appropriate care to ensure a healthy pregnancy and a childbirth without complications.

In this paper, in Section 2, we compare the various machine learning algorithms proposed by other researchers in their papers. Then, in Section 3, we explain the process of our data processing approach in detail and the collected results are discussed and displayed in Section 4. In Sections 5 and 6, we propose our conclusion and recommendations.

## 2. Related Studies

Various machine learning approaches are employed within the healthcare industry to make predictions regarding diseases. Several models have been developed to predict different risk factors associated with pregnancy, such as pre-eclampsia, preterm birth (PTB), diabetes, delivery type (cesarean/vaginal), and newborn infection [4,5]. The utilization of digital devices, such as mobile phones and tablets, in everyday life has generated several prospects. Accessing the current healthcare system through digital devices is an effortless task. However, the availability of data is a significant barrier in developing a model that can accurately predict high-risk pregnancies during gestation.

In several research studies, researchers employed a range of machine learning algorithms, including C.5 Decision Tree (C.5 DT), Naive Bayes (NB), Decision Jungle (DJ), Random Forest Decision Tree (RDT), Linear Discriminant Analysis (LDA), Decision Forest (DF), Support Vector Machine (SVM), Locally Deep Support Vector Machine (L-DSVM), Logistic Regression (LR), Averaged Perceptron (AP), K-Nearest Neighbors (KNN), Bayes Point Machine (BPM), and Neural Network (NN), to forecast the likelihood of encountering risk situations during pregnancy and childbirth [6,7].

The research conducted by Lakshmi [8] employed the C.5 Decision Tree algorithm to forecast the likelihood of risk during pregnancy. In their investigation, the researchers used a total of 12 attributes and out of the 230 data samples, 152 were accurately predicted, while 78 were projected wrongly. Consequently, the model obtained an accuracy rate of 66.08% when applied to the un-standardized dataset. In the same study, with a standardized dataset, out of 230 samples, 164 were accurately predicted, while 66 were forecasted incorrectly, resulting in an overall accuracy rate of 71.03%.

Pereira [9] conducted a study aimed at predicting the type of delivery by leveraging obstetric risk factors during pregnancy through the application of various machine learning algorithms, including DT, GLM, SVM, and NB. The researchers employed a range of factors including body weight, body mass index (BMI), height, blood pressure (HBP and LBP), marital status, number of gestation weeks, and blood group type. In their study, Pereira [9] achieved an accuracy of 83.91% using the Decision Tree (DT) algorithm.

In their study, Akbulut [10] used a total of nine algorithms, including DT, LR, NN, SVM, LD-SVM, DF, B-DT, BPM, and AP, to make predictions regarding the health of the fetus. The researchers reported an accuracy rate of 89.5% for four specific algorithms: AP, B-DT, DJ, and DF. Bautista [11] used a set of nine parameters and conducted a comparative analysis utilizing the DT, RDT, KNN, and SVM algorithms. Bautista [11] reported a 90% accuracy rate using the Random Forest Decision Tree (RDT) algorithm. Raja [12] conducted a study aimed at predicting preterm birth (PTB) in rural India. The researchers applied the DT and LR algorithms for this purpose. According to the authors, the Support Vector Machine (SVM) methods achieved an accuracy of 91%.

In their study, Moreira [13] conducted research on the provision of care for high-risk pregnancies, employing the Multilayer Perceptron (MLP) technique. The authors reported an accuracy rate of 93% in their findings. In their study, Paydar [14] employed a Multilayer Perceptron (MLP) model to forecast pregnancy outcomes in pregnant women, attaining

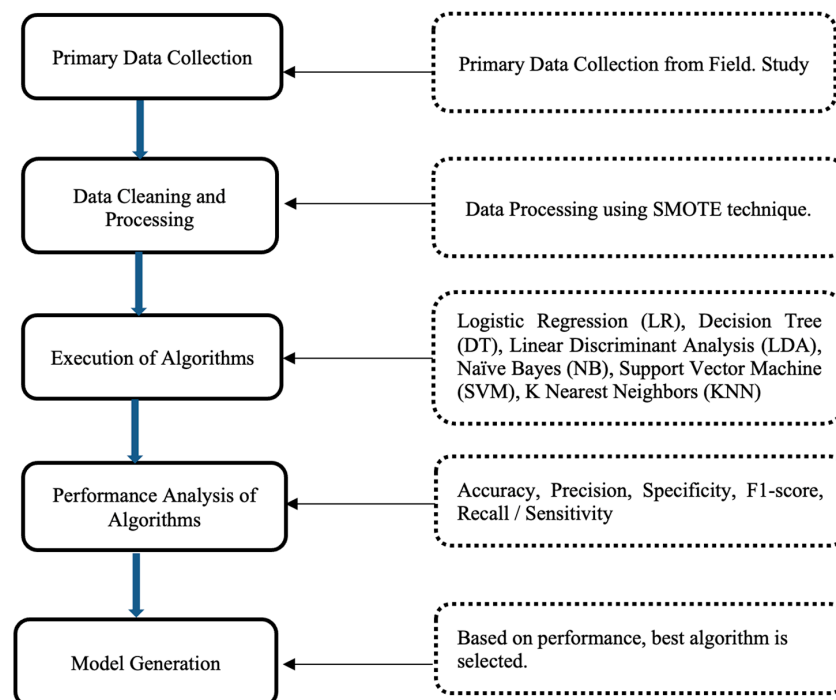
a notable accuracy rate of 97%. In a study conducted by Macrohon [15] various machine learning algorithms, including DT, RF, KNN, NB, and MLP, were employed to forecast high-risk pregnancies. Authors used six included six parameters in their study and achieved a 97.01% accuracy with the Multilayer Perceptron (MLP) algorithm. Table 1 presents a comparison of similar studies, including the authors, the accuracy of the suggested approach, and the performance of the best-performing method.

**Table 1.** Similar studies conducted to predict high-risk pregnancy.

Ref.	Used Machine Learning (ML) Algorithms	Best Performing (ML) Algorithm/s	Maximum Achieved Accuracy
[8]	C.5 DT	C.5 DT	71.30%
[9]	DT, GLM, SVM, NB	DT	83.91%
[10]	AP, DJ, BDT, DF, BPM, LDSVM, LR, SVM, NN	AP, BDT, DF, DJ	89.5%
[11]	DT, RDT, KNN, SVM	RDT	90%
[12]	DT, LR, SVM	SVM	91%
[13]	MLP	MLP	93%
[14]	MLP	MLP	97%
[15]	DT, RF, KNN, SVM, NB, MLP	DT with self-training	97.01%
Proposed Solution [16,17]	LR, LDA, KNN, DT, NB, SVM	DT	98.82%

### 3. Methodology

Initially, primary data were collected during the field study in the designated villages of the district Udham Singh Nagar in Uttarakhand. During the data-cleaning procedure, it was discovered that the data are imbalanced. The Synthetic Minority Over-Sampling Technique (SMOTE) was employed to achieve a balanced dataset. Subsequently, various machine learning algorithms were applied to the dataset in order to identify the optimal algorithm for the high-risk pregnancy (HRP) prediction model. The entire process is depicted in Figure 1.



**Figure 1.** A block diagram for generating the machine learning (ML) model.

### 3.1. Data Collection

Data were gathered from the villages of Mohanpur, Devnagar, Bhuda, Rameshpur, Chutki Devaria, Fulshungi, Khera, Rampura, Bhadipura, Sanjay Nagar, Shaktifarm, and Maharajpur, located in the district of Udham Singh Nagar, Uttarakhand. We presented our work to the ethical committee prior to our field study and obtained their approval. The members of the ethical committee were from the Indian Institute of Technology (IIT) Delhi and the All-India Institute of Medical Science (AIIMS) Delhi. We obtained the consent of all volunteers who participated in the interview and data collection. Their consent was obtained in writing. Consent was obtained from 396 volunteers. After receiving authorization from the authorities, 282 datasets were obtained from the district hospital. Additionally, the SMOTE was employed to construct 159 datasets in order to balance the dataset.

According to figures from the Sample Registration System (SRS), Uttarakhand is the sole state in the country where the Maternal Mortality Ratio (MMR) climbed from 89 in the period of 2015–17 to 103 in the years 2018–20. The Government of India implemented numerous programs and measures to decrease the Maternal Mortality Rate (MMR), despite the fact that the MMR was rising in Uttarakhand. So, these villages were selected for field study of district Udham Singh Nagar, Uttarakhand. The villages indicated above were selected using the stratified random sampling technique. And the respective Sub Center (SC), Primary Healthcare Center (PHC), and Community Healthcare Center (CHC) were visited to conduct interviews with pregnant women. The study involved conducting interviews with pregnant women who had sought care at the Sub Center (SC), Primary Health Center (PHC), and Community Health Center (CHC). These interviews were conducted in collaboration with Accredited Social Health Activists (ASHA) and Auxiliary Nurse Midwives (ANM) in order to gather the necessary data. A description of features of the dataset via the stratified random sampling technique is shown in Table 2. The collected data consist of a total of 837 rows and 13 columns.

The study focuses on the population of women of reproductive age (15–49 years) in the Udham Singh Nagar district in Uttarakhand, which totals 519,972 individuals. Among these, 41,006 pregnant women were registered for Antenatal Care (ANC) [18].

To determine the appropriate sample size, Yamane's formula was applied. Yamane's formula is given as follows:

$$n = \frac{N}{1 + N(e)^2} \quad (1)$$

where  $N$  represents the population size and  $e$  is the margin of error. For this study,  $N = 41,006$  and  $e = 5\%$  (0.05). Plugging these values into the formula, we obtain the following:

$$n = \frac{41006}{1 + 41006(0.05)^2} = 396$$

This calculation yielded a required sample size of 396. However, an additional 282 samples were collected from the District Hospital of Udham Singh Nagar from women who delivered during the same period. Therefore, the total sample size increased to 678. During the classification process, the sample classes were found to be unbalanced. To address this imbalance, the Synthetic Minority Over-sampling Technique (SMOTE) was employed. This technique generated additional samples for the minority class, resulting in a final sample size of 837.

The features listed in Table 2 were chosen after extensive consultation with the gynecologists and surgeons of Jawahar Lal Nehru District Hospital, Udham Singh Nagar, Uttarakhand. They recommended these non-invasive parameters as essential for the monitoring and prediction of high-risk pregnancies (HRP). They assisted in the classification of high- and low-risk classes. Additionally, this dataset was shared and debated with gynecologists from other hospitals to obtain their input in order to establish the risk classification classes. Table 3 provides a summary of the HRP dataset.

**Table 2.** HRP dataset description.

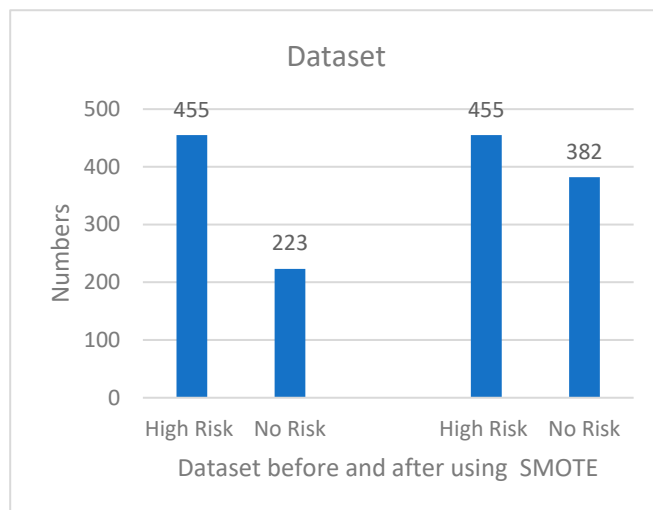
S. No	Feature ID	Feature Name
1	Age	Age of pregnant woman in years
2	G	Gravida (G)—refers to the numerical representation of the total number of times a woman has conceived, including the ongoing pregnancy.
3	P	Para (P)—refers to the delivery of a newborn after a gestational period of 20 weeks or more, irrespective of the infant’s viability at birth.
4	L	Live birth (L)—the total number of living children
5	A	Abortion (A)—the termination of pregnancies (planned or unplanned)
6	D	Death (D)—the number of children dead.
7	SBP	Systolic blood pressure
8	DSP	Diastolic blood pressure
9	RBS	Amount of glucose in blood mg/dL.
10	BT	Pregnant woman’s body temperature in Fahrenheit
11	HR	Heartbeats per minute.
12	Hb	Hemoglobin level of pregnant woman
13	RR	Normal respiration rate at rest per minute.

**Table 3.** Summary of HRP dataset.

Total number of features	13
Total number of classes	2
Total number of instances	837
High Risk	455
No Risk	382

3.2. Data Cleaning and Processing

The data were gathered from several villages during the field visits. The gathered data, which were initially unbalanced, underwent data balancing procedures. At the outset, there were 455 rows designated as High Risk and 223 rows designated as No Risk. The application of Synthetic Minority Oversampling Techniques (SMOTE) [19] resulted in an increase in the number of No Risk rows to 382, as depicted in Figure 2.



**Figure 2.** HRP dataset before and after using SMOTE.

### 3.3. Machine Learning Algorithms

The utilization and prevalence of machine learning are experiencing a notable rise across various domains. Machine learning has proven to be beneficial in the identification and diagnosis of issues within diverse domains of the medical field. The primary aim of this study is to forecast the likelihood of pregnancy-related risks by utilizing the medical history and vital signs of pregnant individuals. Digital technology can assist healthcare workers in identifying pregnancies that have a higher risk. This prediction is accomplished by the utilization of diverse machine learning techniques, including Logistic Regression (LR), Decision Tree (DT), Linear Discriminant Analysis (LDA), Naive Bayes (NB), Support Vector Machine (SVM), and K-Nearest Neighbors (KNN). These algorithms acquire knowledge from the dataset and make predictions for a novel user. The entire dataset was partitioned into an 80:20 ratio, with 80% of the data allocated for training the model and the remaining 20% used for testing [18]. The dataset consists of 837 rows, which belong to two risk classification classes: High Risk, including 455 rows, and No Risk, comprising 382 rows. To achieve optimal accuracy, we employed the 10-fold cross-validation technique for both training and testing purposes.

### 3.4. Performance Metrics for Comparative Analysis

On the collected dataset, multiple machine learning algorithms such as Logistic Regression (LR), Linear Discriminant Analysis (LDA), K-Nearest Neighbors (KNN), Decision Tree (DT), Naive Bayes (NB), and Support Vector Machine (SVM) were used to predict the risk [19,20]. To assess the accuracy and performance of the chosen algorithms, a comparative analysis employing the 10-fold cross validation testing technique was conducted. The algorithms were analyzed using the following criteria and based on their performance, the best model was selected [21],

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \quad (2)$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (3)$$

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (4)$$

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}} \quad (5)$$

$$\text{F1 Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (6)$$

## 4. Results

Logistic Regression (LR), Linear Discriminant Analysis (LDA), K-Nearest Neighbors (KNN), Decision Tree (DT), Naive Bayes (NB), and Support Vector Machine (SVM) algorithms were used on the data collected from the field, and results are presented in this section [22–24].

### Comparative Analysis of Result

All the machine learning algorithms were trained and tested using a k-fold cross validation method. From Table 4 we can see that Support Vector Machine (SVM) has only a 66.67 accuracy, 80% precision, 66% sensitivity, 100% specificity, and a 62% fa-score. Next, the Linear Discriminant Analysis (LDA) algorithm has only an 83.33% accuracy, 84% precision, 83% sensitivity, 80% specificity, and an 83% f1-score. The Naive Bayes (NB) algorithm has an 88.69% accuracy, 89% precision, 89% sensitivity, 86% specificity, and 89% f1-score. On the other hand, Logistic Regression (LR) performed better, with a 91.27% accuracy, 92% precision, 92% sensitivity, 94% specificity, and 92% f1-score. Similarly, on the same dataset, the K-Nearest Neighbors (KNN) algorithm gave better results compared to

the abovementioned algorithms, with a 95.23% accuracy, 95% precision, 95% sensitivity, 94% specificity, and a 95% f1-score. Lastly, the Decision Tree (DT) algorithm gave the best result, with a 98.80% accuracy, 99% precision, 99% sensitivity, 99% specificity, and a 99% f1-score. This comparative analysis gave the conclusion that Decision Tree algorithm performed best among all other algorithms.

**Table 4.** Performance results of machine learning algorithms.

Algorithm.	Accuracy (%)	Precision (%)	Sensitivity (%)	Specificity (%)	f1-Score (%)
Support Vector Machine (SVM)	66.67	80	66	100	62
Linear Discriminant Analysis (LDA)	83.33	84	83	80	83
Naive Bayes (NB)	88.69	89	89	86	89
Logistic Regression (LR)	91.67	92	92	94	92
K-Nearest Neighbors (KNN)	95.23	95	95	94	95
Decision Tree (DT)	98.80	99	99	99	99

Among all the Machine Learning algorithms used to predict High Risk during pregnancy, the modified Decision Tree algorithm produced the best results.

Regarding the use of the Decision Tree (DT) algorithm with a K-fold cross validation technique, Table 5 indicates a 98.82% accuracy in using the 20% unlabeled dataset. The algorithm predicted 84 samples as true positive (TP), 1 sample as false positive (FP), 1 sample as false negative (FN), and 82 samples as true negative (TN). Every time, the algorithm selects 20% of the test data randomly and, similarly, one fold is selected for testing among K folds of the training dataset. The Decision Tree algorithm gave a 99% accuracy for High Risk and No Risk.

**Table 5.** Confusion matrix performance result of the selected model.

Decision Tree	Predicted High Risk	Predicted No Risk	Recall
Actual High Risk	84	1	98.82%
Actual No Risk	1	82	98.80%
Precision	98.82%	98.8%	

### 5. Discussion

Numerous studies and pieces of research have been conducted on the utilization of machine learning algorithms for the prediction of risk during pregnancy. The objective of this study was to enhance the precision of the machine learning algorithm. However, the performance of the utilized algorithms can be influenced by several restrictions associated with the available dataset, including but not limited to accuracy, biasing, and weaknesses [24]. The objective of the design and development of this model was to proactively detect pregnancies with a high risk of complications in order to mitigate maternal mortality rates. Furthermore, this intervention is expected to contribute to the enhancement of maternal health outcomes in rural regions. Due to the limited resources available, numerous healthcare centers in rural areas often consider it a viable alternative. The methodology can be effectively utilized by front-line healthcare workers (FLHWs) operating in remote regions for the purpose of identifying pregnancies that are at a heightened risk.

The objective of this study was to create and implement a robust predictive model for identifying high-risk pregnancies in rural regions. This was achieved by employing a range of machine learning techniques, including Logistic Regression (LR), Linear Discriminant Analysis (LDA), K-Nearest Neighbors (KNN), Decision Tree (DT), Naive Bayes (NB), and Support Vector Machine (SVM). The dataset was partitioned into an 80:20 ratio for the purpose of training and evaluating the models. For each method, a random selection of 20% of the unlabeled data were used for validation purposes. The evaluation of all the aforementioned algorithms

was conducted using a 10-fold cross-validation technique. In each cycle, a one-fold validation approach was employed. Upon conducting an analysis of the algorithms, it was observed that the Decision Tree (DT) method yielded the most favorable outcome, exhibiting an accuracy rate of 98.80%, a precision of 99%, sensitivity of 99%, specificity of 99%, and an f1-score of 99%. In our investigation, it was determined that K-Nearest Neighbors (KNN) and Logistic Regression (LR) exhibit notable performance as the second-best algorithms, achieving accuracy, precision, sensitivity, and f1-score values of 95% and 92%, respectively. Linear Discriminant Analysis (LDA) and Naive Bayes (NB) had the third highest levels of accuracy, achieving 89% and 83%, respectively. During the comparative analysis of these machine learning methods, it was observed that the Support Vector Machine (SVM) approach exhibited the lowest accuracy rate of 67%. The Python language was used as a programming language to implement all of the aforementioned Machine Learning (ML) algorithms and to develop the prediction model.

In this study, various machine learning algorithms were used to predict the risk during pregnancy. One limitation of this study is its sample size of 837, which potentially affects the findings' generalizability. The dataset's regional specificity may not fully represent broader demographic variations, possibly limiting the model's applicability across different populations. Additionally, the predictive accuracy of the algorithm hinges on the selected features of the dataset, where some other relevant variables are absent due to data constraints, which could impact the outcome. Furthermore, the complexity of machine learning algorithms might obscure the interpretability of predictions, posing challenges for clinical integration, without clear, actionable insights derived from the model's decision-making process.

## 6. Conclusions

In this study, it was determined that the Decision Tree (DT) algorithm can serve as a foundational algorithm for the development of a high-risk pregnancy prediction (HRPP) system. The aforementioned approach can prove to be beneficial in geographically isolated regions with limited resources, where healthcare professionals are burdened with additional responsibilities. This system is anticipated to contribute to a reduction in the burden and an enhancement in the efficiency of healthcare professionals, including ASHA and ANMs, by enabling them to evaluate high-risk pregnant women inside their respective healthcare facilities. Pregnant individuals can also utilize this system to evaluate their condition with appropriate training. For the future, there are plans to create a mobile application that may be utilized by healthcare professionals and expectant mothers through their mobile devices. Additionally, the application has the capability to provide recommendations for potential solutions. The successful implementation of this intervention necessitates substantial advice and support from healthcare practitioners. In order to decrease maternal mortality rates and enhance the health of pregnant women, the implementation of information and communication technology (ICT) interventions within the medical field could be beneficial.

**Author Contributions:** Conceptualization, K.T. and C.M.S.; methodology, K.T.; software, K.T.; validation, K.T., V.M.C. and T.P.; formal analysis, K.T., C.M.S. and T.P.; investigation, K.T., V.M.C. and T.P.; resources, K.T., V.M.C. and T.P.; data curation, K.T., C.M.S., V.M.C. and T.P.; writing—original draft preparation, K.T., V.M.C. and T.P.; writing—review and editing, K.T., C.M.S., V.M.C. and T.P.; visualization, K.T., C.M.S., V.M.C. and T.P.; supervision, V.M.C.; project administration, K.T., V.M.C. and T.P. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Institutional Review Board Statement:** The study was conducted in accordance with the Declaration of Helsinki, and approved by the Ethics Committee of IIT DELHI (under the protocol number 2021/P020).

**Informed Consent Statement:** Written informed consent was obtained from the participants before data collection. No personally identifiable information is being published in this paper.

**Data Availability Statement:** The data presented in this study are available upon reasonable request from the corresponding author.

**Conflicts of Interest:** The authors declare no conflict of interest.



## References

1. Tomar, K.; Sharma, C.M.; Sharma, P.; Gupta, D.; Chariar, V.M. Impacts of Environmental Factors on Maternal Health in Low Resource Settings. In Proceedings of the 6th International Conference on Resources and Environment Sciences (ICRES 2024), Bangkok, Thailand, 7–9 June 2024. [CrossRef]
2. Cleveland Clinic. High-Risk Pregnancy. 14 December 2021. Available online: <https://my.clevelandclinic.org/health/diseases/22190-high-risk-pregnancy> (accessed on 10 April 2023).
3. US National Institutes of Health. What Is a High-Risk Pregnancy? 31 January 2017. Available online: <https://www.nichd.nih.gov/health/topics/pregnancy/conditioninfo/high-risk> (accessed on 10 April 2023).
4. Ebrahimzadeh, F.; Hajizadeh, E.; Vahabi, N.; Almasian, M.; Bakhteyar, K. Prediction of unwanted pregnancies using logistic regression, probit regression and discriminant analysis. *Med. J. Islam. Repub. Iran* **2015**, *29*, 828–832.
5. Montella, E.; Ferraro, A.; Sperli, G.; Triassi, M.; Santini, S.; Improta, G. Predictive Analysis of Healthcare-Associated Blood Stream Infections in the Neonatal Intensive Care Unit Using Artificial Intelligence: A Single Center Study. *Int. J. Environ. Res. Public Health* **2022**, *19*, 2498. [CrossRef] [PubMed]
6. Yiu, T. Understanding Random Forest. 12 June 2019. Available online: <https://towardsdatascience.com/understanding-random-forest-58381e0602d2> (accessed on 14 April 2023).
7. Zhu, W.; Zeng, N.; Wang, N. *Sensitivity, Specificity, Accuracy, Associated Confidence Interval and ROC Analysis with Practical SAS® Implementations*; Northeast SAS Users Group 2010; Health Care and Life Sciences: Baltimore, MD, USA, 2010; pp. 1–9.
8. Lakshmi, B.N.; Indumathi, T.S.; Ravi, N. A Study on C.5 Decision Tree Classification Algorithm for Risk Predictions During Pregnancy. *Procedia Technol.* **2016**, *24*, 1542–1549. [CrossRef]
9. Pereira, S.; Portela, F.; Santos, M.F.; Machado, J.; Abelha, A. Predicting Type of Delivery by Identification of Obstetric Risk Factors through Data Mining. *Procedia Comput. Sci.* **2015**, *64*, 601–609. [CrossRef]
10. Akbulut, A.; Ertugrul, E.; Topcu, V. Fetal health status prediction based on maternal clinical history using machine learning techniques. *Comput. Methods Programs Biomed.* **2018**, *163*, 87–100. [CrossRef] [PubMed]
11. Bautista, J.M.; Quiwa, Q.A.I.; Reyes, R.S.J. Machine learning analysis for remote prenatal care. In Proceedings of the IEEE Region 10 Annual International Conference, Proceedings/TENCON, Osaka, Japan, 16–19 November 2020; pp. 397–402. [CrossRef]
12. Raja, R.; Mukherjee, I.; Sarkar, B.K. A Machine Learning-Based Prediction Model for Preterm Birth in Rural India. *J. Healthc. Eng.* **2021**, *2021*, 6665573. [CrossRef] [PubMed]
13. Moreira, M.W.; Rodrigues, J.J.; Kumar, N.; Al-Muhtadi, J.; Korotaev, V. Nature-Inspired Algorithm for Training Multilayer Perceptron Networks in e-health Environments for High-Risk Pregnancy Care. *J. Med. Syst.* **2018**, *42*, 51. [CrossRef] [PubMed]
14. Paydar, K.; Niakan Kalthori, S.R.; Akbarian, M.; Sheikhtaheri, A. A clinical decision support system for prediction of pregnancy outcome in pregnant women with systemic lupus erythematosus. *Int. J. Med. Inform.* **2017**, *97*, 239–246. [CrossRef] [PubMed]
15. Macrohon, J.J.E.; Villavicencio, C.N.; Inbaraj, X.A.; Jeng, J.-H. A Semi-Supervised Machine Learning Approach in Predicting High-Risk Pregnancies in the Philippines. *Diagnostics* **2022**, *12*, 2782. [CrossRef] [PubMed]
16. Yadav, A. Support Vector Machines (SVM)—20 October 2018. Available online: <https://towardsdatascience.com/support-vector-machines-svm-c9ef22815589> (accessed on 14 April 2023).
17. Antonogeorgos, G.; Panagiotakos, D.B.; Priftis, K.N.; Tzonou, A. Logistic Regression and Linear Discriminant Analyses in Evaluating Factors Associated with Asthma Prevalence among 10- to 12-Years-Old Children: Divergence and Similarity of the Two Statistical Methods. *Int. J. Pediatr.* **2009**, *2009*, 1–6. [CrossRef] [PubMed]
18. Singh, N.; Nguyen, P.H.; Jangid, M.; Singh, S.K.; Sarwal, R.; Bhatia, N.; Johnston, R.; Joe, W.; Menon, P. *District Nutrition Profile: Udham Singh Nagar, Uttarakhand*; International Food Policy Research Institute: New Delhi, India, 2022.
19. Hernandez, M.; Epelde, G.; Beristain, A.; Álvarez, R.; Molina, C.; Larrea, X.; Alberdi, A.; Timoleon, M.; Bamidis, P.; Konstantinidis, E. Incorporation of Synthetic Data Generation Techniques within a Controlled Data Processing Workflow in the Health and Wellbeing Domain. *Electronics* **2022**, *11*, 812. [CrossRef]
20. Baratloo, A.; Hosseini, M.; Negida, A.; El Ashal, G. Part 1: Simple Definition and Calculation of Accuracy, Sensitivity and Specificity. *Emergency* **2015**, *3*, 48–49. [PubMed]
21. Villavicencio, C.N.; Macrohon, J.J.E.; Inbaraj, X.A.; Jeng, J.H.; Hsieh, J.G. COVID-19 prediction applying supervised machine learning algorithms with comparative analysis using weka. *Algorithms* **2021**, *14*, 201. [CrossRef]
22. Chauhan, N.S. Decision Tree Algorithm, Explained. 9 February 2022. Available online: <https://www.kdnuggets.com/2020/01/decision-tree-algorithm-explained.html> (accessed on 14 April 2023).
23. IBM K-Nearest Neighbors Algorithm. Available online: <https://www.ibm.com/topics/knn> (accessed on 14 April 2023).
24. Raschka, S. STAT 479: Machine Learning. 2018. Available online: [https://sebastianraschka.com/pdf/lecture-notes/stat479fs18/02\\_knn\\_notes.pdf](https://sebastianraschka.com/pdf/lecture-notes/stat479fs18/02_knn_notes.pdf) (accessed on 14 April 2023).

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.