

MATERNAL RISK ANALYSIS WITH COMPARISON OF MACHINE LEARNING ALGORITHMS AND USE OF DATA AUGMENTATION.

Hernandez Chávez Roberto¹, Chavarría Hidalgo Sandra Guadalupe² Cañedo Figueroa Carlos Eduardo³,
Universidad Autónoma de Chihuahua
Facultad de Medicina y Ciencias Biomédicas
Circuito Universitario 31109, campus UACH II, 31125 Chihuahua, Chihuahua, México
Tel.614-211-7561.
a343514@uach.mx¹, schavarria@uach.mx², ccanedo@uach.mx³,

ABSTRACT.

The health risk during pregnancy is an issue that affects a large part of the female sector, occurring in early stages of adolescence, in healthy adult mothers, to women of an older age. Maternal health is a factor of great relevance in how a pregnancy develops, knowing the contexts that can put at risk the life of the fetus and the mother is one of the reasons why doctors and the health area becomes complicated diagnosis and risk prediction. The objective of the work is to get, based on the characteristics presented in the database, a close prediction of the possible health risk of being pregnant. These characteristics or variables are age, systolic and diastolic blood pressure, body temperature, heart rate and blood glucose in terms of molar concentration. All variables used were obtained from the freely accessible database on the Kaggle platform, where it is mentioned that data collection was in multiple hospitals in Bangladesh from pregnant women. By means of machine learning algorithms, a timely prediction of pregnancy risks is sought with certain values. The use of these (KNN, ANN, SVM and NB) and the increase of data, gave us different percentages of accuracy, being 81.36%, 86.7%, 55.15% and 44.72% respectively.

Keywords: Maternal Risk, Machine Learning, Comparison

1. INTRODUCTION

Maternal health is a subject of high medical interest, given the significant risks faced by women. During pregnancy, childbirth and the postnatal period, maternal health is a crucial issue. The well-being of women during pregnancy, childbirth and the postnatal period focuses on maternal health, which is a fundamental element of public health. Despite significant progress made in recent years, maternal mortality rates remain incredibly high, especially in developing countries. According to published data from the World Health Organization (WHO) [1], approximately 287,000 women died during and after childbirth in 2020. The most common direct causes of maternal injury and death were excessive blood loss, infections, high blood pressure, uninsured abortions and obstructed labour. Recent studies have examined the relationship between maternal health conditions and child health outcomes. These studies have identified a relationship between maternal health and child health outcomes, including the presence of chronic diseases and child health, which are affected by poor maternal physical and mental health during pregnancy or up to 15 months postpartum [2].

One potential avenue for enhancing maternal health outcomes is the incorporation of technologies such as mobile software

architectures and genomics. The purpose of this article is to present the results of research on the risks that may occur during pregnancy or gestational age, and to compare the findings with similar work and/or using the same database, including the challenges women face at this time for their health, and to use recent research findings and potential interventions to improve the way maternal health is predicted and supported. Recent research has investigated the potential of machine learning (ML), a subfield of artificial intelligence (AI), in the field of maternal and fetal health [3]. This research has focused on the use of Machine Learning in the context of diseases, complications, and risk management during pregnancy. Machine learning and AI approaches permit the automatic generation of models to support critical decision-making based on real data.

This has the potential to significantly impact the early detection of risk factors during the early stages of pregnancy, as well as the management, mitigation, and prevention of risks, which can ultimately decrease adverse perinatal outcomes and complications for both mother and fetus. The field of machine learning applied to pregnancy-related diseases and complications is undergoing continual expansion. The use of techniques such as Random Forest and ANNs in machine learning has demonstrated encouraging outcomes, including the capacity to predict the type of delivery and to assess maternal risks [1].

The combination of technological tools allows for the identification of maternal health risks during pregnancy, thereby providing valuable tools for the classification and addressing of such risks.

The aforementioned article examines the use of machine learning in the field of maternal health, with a particular focus on the manner in which it addresses risk factors and complications during pregnancy. The article goes on to examine how these techniques can be used to predict the health risks and outcomes for mothers. Furthermore, the article examines the challenges of data availability and ethical considerations in the medical application of machine learning.

The article commences with an overview of maternal health, highlighting global challenges and disparities. It then proceeds to discuss common diseases such as hypertension and gestational diabetes, their risks, and how they affect the mother and fetus. It subsequently delves into the specific use of machine learning in this field. Techniques such as "random forests" and neural

networks are detailed, and cases are shown where they have identified risks and offered personalized interventions [1]. The text goes on to address current challenges, such as data quality and model interpretation, as well as ethical issues in their clinical use.

A number of studies have investigated the application of machine learning in maternal health risk prediction [1]. The approach is to create hybrid models that combine deep learning models and the predictive capabilities of traditional machine learning. Furthermore, studies have been conducted on machine learning-based maternal health risk prediction models for the Internet of Medical Things (IoMT). These models employ analytical tools and machine learning algorithms to analyze health data and pregnancy-related risk factors [4][5][6].

2. METHODOLOGY

2.1. Dataset description .

The data comes from pregnant women and has been collected from various sources, such as hospitals, community clinics and maternal health centers, through the IoT-based risk surveillance system [7]. The database was obtained from the research “IoT Based Risk Level Prediction Model for Maternal Health Care In The Context Of Bangladesh” [1] and its main or most relevant characteristics are systolic and diastolic pressure, which can give us as a reference factor health problems, such as preeclampsia. The aforementioned work based its methodology of obtaining data as shown in the article “Review and Analysis of Risk Factor of Maternal Health in Remote Area Using the Internet of Things (IoT)”, where their characteristics are presented in the same way blood pressure, blood glucose and heart rate [8]. The database includes various information related to maternal health. To assess the level of risk intensity during pregnancy, characteristics such as age, diastolic pressure, systolic pressure, blood glucose levels, body temperature and heart rate were taken into account, all obtained from 1014 patients sampled. The program developed for the sampling presented specific information and stimuli to assess aspects crucial to maternal and fetal health.

2.2. Data Preprocessing

The data obtained were processed using a MATLAB algorithm which randomly added Gaussian noise to the database and a data augmentation technique was applied to address class imbalance, ensuring an equitable representation of the different maternal health conditions. From the database will be distributed, in the case of artificial neural network, 80% of the data for training and 20% for test, in the rest of the models the data will be distributed in 70% for training and 30% for test and validation, once already applied the increase of data and noise in the data.

2.3. KNN (K-nearest neighbors)

The K-Nearest Neighbors (KNN) is a supervised learning algorithm used for classification and regression. It works on the assumption that similar objects share similar characteristics. Its approach is based on identifying the nearest distances between a

new sample and a set of training samples, allowing it to make predictions. Its implementation in Matlab allows to explore and apply this method easily, adjusting parameters such as k to obtain better results according to the specific context of the problem to be solved. For this we applied the `fitknn` function [9].

2.4. Artificial Neuronal Network (ANN)

An artificial neural network, or simply a neural network, is a mathematical model based on biological brain networks [1][10]. The ANN designed was trained with `patternnet` function and contains 6 inputs, 1 hidden layer with 20 neurons and 3 outputs (Fig. 1). The network's training schedule was carried out using 10 epochs, a validation check of 3, a learning factor of 1×10^{-2} and a minimum error 1×10^{-12} , using the Levenberg-Marquardt backpropagation technique.

2.5. Naive Bayes

The Naive Bayes classifier is a supervised learning algorithm based on Bayes' theorem, which assumes that features are independent of each other for each class. It is a computationally efficient tool that can handle large datasets with multiple dimensions [11][1]. This algorithm was developed using `fitcnb` function on matlab.

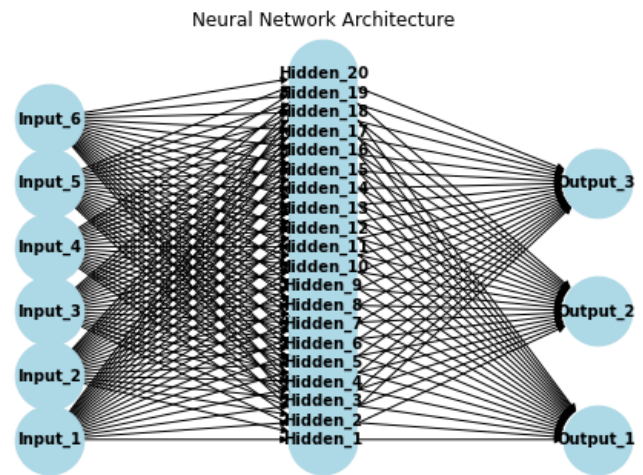


Fig. 1. Artificial Neuronal Network Architecture.

2.6. Support Vector Machine (SVM)

The algorithm of support vector machine (SVM) using a linear kernel for classification data. In this case, we used a SVM trained into MATLAB with `fitsvm` function [12].

2.7. Metrics Assessment

With the remaining 20% and 30% of the data, the efficiency and accuracy of each algorithm was evaluated. The metrics used included the number of true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN). According to

equations (1), (2), (3) and (4) for metrics accuracy, precision, recall and F1- score.

$$Accuracy(\%) = \frac{TP+TN}{TP+FN+FP+TN} * 100 \quad (1)$$

$$Precision(\%) = \frac{TP}{TP+FP} * 100 \quad (2)$$

$$Recall(\%) = \frac{TP}{TP+FN} * 100 \quad (3)$$

$$F1 - Score = 2 * \frac{Precision * recall}{Precision+recall} \quad (4)$$

3. RESULTS.

The development of the algorithms for classification maternal risk can be show on the Fig 2. Where the first point its denoted for the database acquisition, follow for the application of data preprocessing, continue whit the train and evaluation of the algorithms designed.

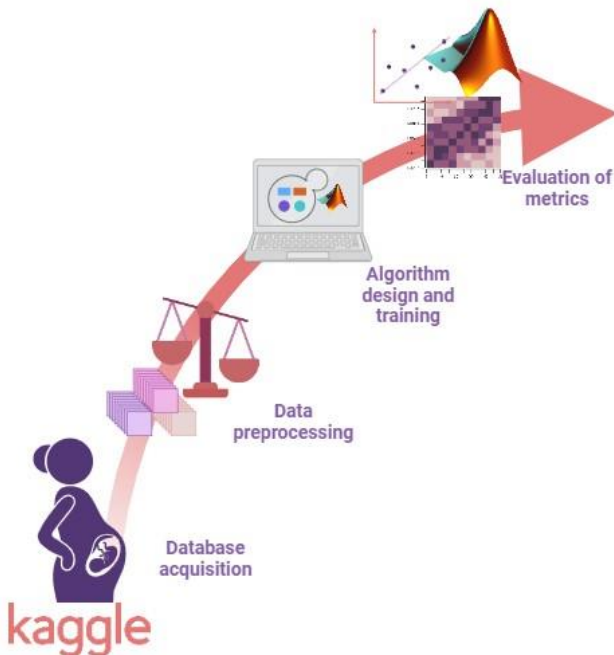


Fig. 2. Methodology and development of the research.

The algorithms that demonstrate the highest efficiency for the analysis of data obtained from are KNN and ANN. In contrast, the SVM and Naive Bayes exhibit the lowest results in F1 score. This is illustrated in Table 1, where the green cell represents the high value and the orange cell the lowest value for each parameter.

Comparing the data with respect to the difference in the high risk and low risk class values, showing in Table 2, the noticeable changes and how these data affect the risk prediction classification, it is observed that both very high and very low variables can be considered high risk and those within regular parameters are low risk.

Table 1. Comparison of values of database characteristics, using high and low risk classes.

Algorithm	Performance			
	F1 Score	Accuracy	Recall	Precision
SVM	55.15 %	83.75 %	59.95 %	51.1 %
ANN	86.7 %	89.71 %	78.57 %	95.06 %
KNN	81.36 %	93.38 %	72.61 %	93.75 %
NB	44.72 %	81.48 %	53.85 %	38.27 %

Table 2. Comparison of values of the database characteristics, using the high and low risk classes, the asterisk indicates the variables with the most significant change to be considered.

Risk class	Age	SysBP	DysBP	BS	Temp	Heart Rate	Value”p”
High Risk	40	140*	100	18 mol/L*	98° F	90*	
Low Risk	32	120*	90	7.5 mol/L*	98° F	70*	

4. DISCUSSION

The field of maternal health and the application of machine learning algorithms has been investigated with various approaches. For instance, the prediction of gestational diabetes has been studied, with the results presented in Table 3. Similarly, research has been conducted on the classification of preeclampsia by means of algorithms, as shown in Table 4.

The application of machine learning algorithms to assist in the early detection of risks during pregnancy would represent a significant advance towards more accurate and efficient prenatal care. These types of models illustrate the potential for a profound transformation in the manner by which maternal health challenges are addressed, through the means of enabling the early detection of potential complications.

This article demonstrates how different machine learning algorithms, including KNN, ANN, SVM and NB, indicate that while there are varying levels of accuracy in predicting risks during gestation, achieving optimal accuracy and effective implementation in clinical settings requires constant improvement in the quality of the data, both in the way it is obtained and in the algorithms.

Furthermore, the beneficial potential of these tools in clinical settings where resources are limited and where early identification of risks could have a crucial impact on the health of mothers and their children becomes apparent. The integration of disparate data sources and the advancement of sophisticated

models for precision interventions during pregnancy could facilitate the attainment of more definitive outcomes.

This technological approach to maternal health represents a significant advance in prenatal care. Despite the challenges that may present themselves, the potential for improving maternal and fetal health outcomes is more than evident, which raises the standard of care to a more accurate and personalized level.

Table 3 Improved results of machine learning models for predicting the risk of gestational diabetes in early pregnancy [13]

Model Type	Accuracy	Recall (Sensitivity)	Precision (Specificity)	F1 Score
SVM	0.7554	0.8205	0.7465	0.7835
MLP	0.7090	0.8462	0.6901	0.7681
LR	0.6006	0.9231	0.5563	0.7397

Table 4 Performance of classification algorithms for detecting maternal risk of preeclampsia with hypertension during pregnancy [14].

Feature Selection	Accuracy	Precision	Recall	F-Measure
Rule Induction	84.22%	84.89%	57.11%	68.67%
K-NN	93.78%	89.14%	88.48%	89.56%
Decision Tree	78.33%	67.16%	44.35%	51.69%
Naïve Bayes	92.61%	85.45%	91.56%	89.33%
Deep Learning	95.12%	89.78%	90.37%	90.37%
SVM	87.15%	79.04%	78.18%	78.32%
SVM (RBF)	85.96%	83.36%	81.11%	82.09%
SVM (LINEAR)	88.10%	81.26%	79.09%	82.79%
Neural Network	93.18%	88.07%	88.66%	89.42%

5. REFERENCES

- [1] T. O. Togunwa, A. O. Babatunde, and K. U. R. Abdullah, "Deep hybrid model for maternal health risk classification in pregnancy: synergy of ANN and random forest," *Front Artif Intell*, vol. 6, 2023, doi: 10.3389/frai.2023.1213436.
- [2] M. Ahmed and M. A. Kashem, "IoT Based Risk Level Prediction Model for

Maternal Health Care in the Context of Bangladesh," in *2020 2nd International Conference on Sustainable Technologies for Industry 4.0, STI 2020*, 2020. doi: 10.1109/STI50764.2020.9350320.

[3] A. Njoku, M. Evans, L. Nimo-Sefah, and J. Bailey, "Listen to the Whispers before They Become Screams: Addressing Black Maternal Morbidity and Mortality in the United States," *Healthcare (Switzerland)*, vol. 11, no. 3, 2023, doi: 10.3390/healthcare11030438.

[4] G. Z. Cheng, A. Chen, Y. Xin, and Q. Q. Ni, "Using the teach-back method to improve postpartum maternal-infant health among women with limited maternal health literacy: a randomized controlled study," *BMC Pregnancy Childbirth*, vol. 23, no. 1, 2023, doi: 10.1186/s12884-022-05302-w.

[5] 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, 2023. doi: 10.3389/fendo.2023.1130139.

[6] G. Kopanitsa, O. Metsker, and S. Kovalchuk, "Machine Learning Methods for Pregnancy and Childbirth Risk Management," *J Pers Med*, vol. 13, no. 6, 2023, doi: 10.3390/jpm13060975.

[7] A. Fredriksson *et al.*, "Machine learning for maternal health: Predicting delivery location in a community health worker program in Zanzibar," *Front Digit Health*, vol. 4, 2022, doi: 10.3389/fdgth.2022.855236.

[8] K. Ahmad, E. Kabir, S. A. Keramat, and R. Khanam, "Maternal health and health-related behaviours and their associations with child health: Evidence from an Australian birth cohort," *PLoS One*, vol. 16, no. 9 September, 2021, doi: 10.1371/journal.pone.0257188.

[9] S. R. Vázquez, A. Vidal, M. Borges, J. Valentín, and L. Ginori, "Classification of the nucleus," *Rev. Cuba. Ciencias Informáticas*, vol. 10, no. 1, pp. 95–109, 2016.

[10] K. Valdez Hernández *et al.*, "Design and Comparison of Artificial Intelligent Algorithms for Breast Cancer Classification," in *XLV Mexican Conference on Biomedical Engineering: Proceedings of CNIB 2022, 6–8 October, Puerto Vallarta, México*, 2022, pp. 46–54.

[11] N. F. Mustamin, F. Aziz, F. Firmansyah y P. Ishak, "Classification Of Maternal Health Risk Using Three Models Naïve Bayes Method", *IJCCS (Indonesian J. Comput. Cybern. Syst.)*, vol. 17, n.º 4, p. 395, octubre de 2023 <https://doi.org/10.22146/ijccs.84242>

[12] P. R. Cillo-Velasco *et al.*, "The Hemodynamics of Aneurysms Treated with Flow-Diverting Stents Considering both Stent and Aneurysm/Artery Geometries", *Appl. Sci.*, vol. 10, n.º 15, p. 5239, julio de 2020. <https://doi.org/10.3390/app10155239>

- [13] Cubillos, G., Monckeberg, M., Plaza, A. et al. Development of machine learning models to predict gestational diabetes risk in the first half of pregnancy. BMC Pregnancy Childbirth 23, 469 (2023). <https://doi.org/10.1186/s12884-023-05766-4>
- [14] Tahir, Muhlis & Badriyah, Tessy & Syarif, Iwan. (2018). Classification Algorithms of Maternal Risk Detection For Preeclampsia With Hypertension During Pregnancy Using Particle Swarm Optimization. EMITTER International Journal of Engineering Technology. 6. 236. 10.24003/emitter.v6i2.287.