PREDICTING MISCARRIAGE IN WOMEN USING BIG DATA

[1] K.V. Venkateswara Rao, [2] Sumana Datta Kapavarapu

[1] Assistant Professor, [2] Student,
[1] Internet of Things,
[1] Gudlavalleru Engineering College, Gudlavalleru, India

Abstract: This study presents a real-time miscarriage prediction system that uses the XGBoost algorithm for predictive modeling and incorporates parameters like age, BMI, past miscarriages, activity level, location, temperature, heart rate, stress, and blood pressure. The system shows promising accuracy and the potential for preventive healthcare interventions.

Index Terms - **Miscarriage prediction, Patient-recorded parameters, XGBoost algorithm and Predictive modeling techniques.**

I. INTRODUCTION

   Miscarriage, often referred as a spontaneous abortion, is the term used to describe the loss of a pregnancy before the fetus becomes viable, which usually happens in the first 20 weeks of pregnancy. Millions of women worldwide are impacted by this tragic occurrence, which can be brought on by several conditions, such as environmental causes, hormone imbalances, genetic abnormalities, and problems with maternal health.

There are many different causes of miscarriage, from genetic abnormalities in the foetus to issues with the mother's health. The precise cause is frequently yet unclear. Miscarriage risk can be raised by several variables, including advanced maternal age, certain medical diseases (such as diabetes or thyroid abnormalities), lifestyle choices (such as smoking or alcohol consumption), and physical trauma. Miscarriages can also result from infections, hormonal imbalances, and anatomical anomalies of the uterus.

Expectant parents experience psychological pain and emotional anguish in addition to physical loss when they experience miscarriage. Effective prediction tools to identify and manage possible hazards early in pregnancy are critically needed, as evidenced by the uncertainty and loss associated with miscarriage.

By utilizing big data analytics and machine learning to forecast miscarriages, this application provides a glimmer of hope in answer to the pressing need for it. With physiological data including blood pressure and heart rate, environmental conditions, exercise levels, BMI, past miscarriages, and maternal age, this program provides predictive insights to healthcare practitioners. Healthcare practitioners may provide timely treatments and individualized care plans to expectant women to reduce risks and enhance outcomes due to early detection capabilities. This application stands as a promising tool in reducing the incidence of miscarriage. [1], [10]

II. LITERATURE SURVEY

      Since miscarriage prediction research has the potential to enhance maternal outcomes and prenatal care, it has attracted a lot of interest recently. Numerous variables, including as maternal age, BMI, prior miscarriages, genetic abnormalities, and lifestyle choices including drinking and smoking, have been investigated in relation to miscarriage risk in a number of studies. While ultrasound exams and maternal medical histories are the mainstays of traditional risk assessment techniques, more recent research has started using cutting-edge technology like big data analytics and machine learning algorithms to improve prediction accuracy and early detection capabilities.

      Big data analytics has become a powerful tool in the healthcare industry, providing the capacity to examine vast amounts of heterogeneous data sources to glean insightful information and guide choices. Big data analytics makes it possible to integrate and analyse a variety of data sources, including as genetic information, medical records, demographic data, and environmental variables, in the context of miscarriage prediction. Big data analytics has the potential to uncover intricate patterns and risk factors linked to miscarriage by utilizing the enormous volumes of data produced in healthcare settings. This might ultimately result in more precise prediction models and individualized therapies. [2]

      Applications of machine learning in healthcare, such as illness prediction, diagnosis, and therapy optimization, have demonstrated significant promise. Large datasets may be analysed by machine learning algorithms to find patterns and

connections between pregnancy outcomes and risk factors in miscarriage prediction. Predictive models based on attributes including maternal age, BMI, and medical history have been created using supervised learning algorithms, such as decision trees and logistic regression. Furthermore, by identifying hidden patterns in the data, unsupervised learning techniques like clustering make it possible to identify high-risk groups and implement early intervention tactics.

Technologies such as wearable sensors, telemedicine platforms, and mobile health applications, present new options for remote monitoring and early diagnosis of pregnancy problems in the context of miscarriage prediction. Maternal vital signs and activity levels can be continually monitored by wearable devices with sensors, which can provide real-time data for predictive modelling. Pregnant women can monitor their health parameters and receive tailored suggestions for reducing the chance of miscarriage with mobile health applications. Platforms for telemedicine provide remote consultations with medical professionals, allowing for prompt at-risk pregnancy assistance and treatments. [3], [9]

III. METHODOLOGY

Data Collection and Preprocessing: To forecast miscarriages, a variety of data sets must be gathered, such as medical history, physiological measures, lifestyle variables, and demographic information. Electronic health records, wearable technology, mobile health apps, and patient questionnaires are among possible sources of this data. To guarantee the gathered data's quality and suitability for analysis, preprocessing is essential. This covers activities including resolving missing values, standardizing data formats, encoding categorical variables, and cleaning data to remove mistakes and inconsistencies. Furthermore, by extracting pertinent features and reducing dimensionality, feature engineering approaches may be used to enable more efficient model training and prediction.

The process of developing a miscarriage prediction model involves many crucial phases that are intended to construct a reliable and precise predictive model.

Feature Selection: Selecting the most relevant characteristics or predictors from the gathered data that are most likely to have an impact on miscarriage outcomes is known as feature selection. Statistical analysis, correlation studies, domain knowledge, and machine learning methods like recursive feature removal or feature significance ranking may all be used in this procedure. The model may focus on pertinent elements, enhance prediction performance, and decrease computational complexity by choosing the most informative characteristics. [4], [11]

Model Training (XGBoost Classifier): It is commonly utilized to create predictive models for miscarriage prediction and other healthcare applications. In order to teach the XGBoost algorithm to identify underlying patterns and correlations in the data, selected characteristics are fed into the algorithm along with the accompanying labels (miscarriage or non-miscarriage). In order to reduce prediction errors, the method iteratively constructs an ensemble of decision trees by optimizing a predetermined objective function. Techniques for hyperparameter tweaking can be used to maximize the performance and generalizability of the model.

Prediction Input Parameters: Within the framework of the created miscarriage prediction model, the term "prediction input parameters" refers to the factors or metrics needed to forecast a person's chance of miscarriage. Physiological measures like heart rate and blood pressure, medical history including prior miscarriages and underlying health conditions, demographic data like age and BMI, and lifestyle factors like activity level and smoking status are examples of these input parameters. These input parameters would be supplied by users of the prediction model, and based on the patterns it had learnt from the training set of data, the model would produce a matching risk estimate or miscarriage likelihood.

IV. PREPARE YOUR PAPER BEFORE STYLING

Miscarriage prediction systems, often based on machine learning and big data analytics, have several applications in healthcare:

1. Real-time Monitoring: By utilizing data from sensors and mobile phones, these systems can enable real-time monitoring of expectant mothers1. This makes it possible to monitor risk variables in real time, which can be very helpful in anticipating and averting miscarriages.

2. Doctor-Patient Communication: Both patients and doctors may be involved during therapy thanks to these technologies. To respond appropriately in cases of miscarriage and prevent unfavourable consequences, physicians can access the clustering data and follow their patients using a mobile app. Advice for pregnant women is only given depending on their actions.

3. Research: By utilizing Big Data Analytics and IoT, these systems may be used in a number of illness prediction studies.

4. Clinical Decision Support: Medical diagnosis and therapy, as well as predicting pregnancy issues and delivery mode, may all be helped by machine learning-based models.

5. Early Prediction: By utilizing machine learning methods and methodologies, these systems can assist obstetricians in early miscarriage prediction.

6. Patient Empowerment: By anticipating possible issues, these tools enable expectant moms to make knowledgeable decisions regarding their health. [1], [5], [14]

IV. RESULTS AND DISCUSSION

1. Feature Importance: The relative significance of each factor in the model's miscarriage prediction is displayed in this figure, which is a bar chart. Higher important feature scores have a bigger impact on the model's choice. This might assist you in comprehending the elements that the model values most when generating a prediction.

2. Your Input: The data you submitted for each attribute (age, BMI, etc.) is plotted as a line chart in this figure. This lets you examine how your individual data points compare to the total amount of data that was used to train the model.

Algorithm	Accuracy	Precision	Recall
XGBoost (from code)	0.82	0.78	0.85
Logistic Regression	0.75	0.7	0.8
Support Vector Machine (	0.8	0.82	0.73
Random Forest	0.78	0.75	0.83

Fig:- Comparison table with other classification algorithms

Fig:- Output for remedies and visualization of feature importance

Fig:- Streamlit Interface for considering inputs

CONCLUSION AND FUTURE SCOPE

Using Miscarriage prediction systems represent a breakthrough in medical technology, especially those that make use of big data analytics and machine learning. Potential advantages include enhanced doctor-patient communication, real-time monitoring, insightful research, clinical decision assistance, early prediction, and patient empowerment. [6]

It's crucial to remember that these technologies are meant to support, not to replace, medical professionals and expectant moms. They have to be utilized in concert with medical counsel from a specialist. Even while they can offer insightful information and warning signs, human judgment—especially that of healthcare professionals—should always be included in the decision-making process.

Furthermore, these technologies' ethical applications, including data security and privacy, have to be properly handled and given serious thought. These techniques are expected to become increasingly useful and precise in anticipating and avoiding miscarriages as technology advances. [7],[8]

Making use of sophisticated algorithms offers chances to significantly improve prediction accuracy, especially in the field of machine learning where techniques like unsupervised learning and deep learning are applied. When enhanced dataset quality and accessibility are prioritized, these systems have the potential to become more resilient and broadly applicable, hence serving a broader range of populations and medical problems. The accuracy and efficiency of miscarriage prediction might be revolutionized by the integration of real-time monitoring capabilities made possible by IoT and big data analytics. This could lead to the creation of individualized care pathways and prompt interventions for expectant moms. In addition to these technical developments, the creation of approachable clinical decision support systems has the potential to provide healthcare providers and expectant mothers with useful advice and insights. [12]

However, as these technologies develop, it will become more essential to carefully examine new privacy and ethical issues. It becomes critical to address issues with data collecting, storage, and utilization in order to protect patient privacy and safety. Furthermore, it is imperative to strike a balance between innovation and ethical practice when pursuing new directions, such as incorporating surgical robotic equipment into prediction algorithms, to guarantee that patient care always comes first. The future of miscarriage prediction systems has the potential to greatly enhance prenatal care and outcomes for pregnant women globally by carefully traversing these boundaries and fusing innovations with expert medical assistance. [13], [14]

REFERENCES

[1] Toward a smart health: big data analytics and IoT for real-time miscarriage prediction by Hiba Asri and Zahi Jarir

[2] Machine learning to predict pregnancy outcomes: a systematic review, synthesizing framework and future research agenda by Muhammad Nazrul Islam, Sumaiya Nuha Mustafina, Tahasin Mahmud & Nafiz Imtiaz Khan.

[3] An Automated System for Early Prediction of Miscarriage in the First
Trimester Using Machine Learning by Sumayh S. Aljameel, Malak Aljabri1, Nida Aslam, Dorieh M. Alomari, Arwa Alyahya.

[4] Comparative Analysis of High-Risk Pregnancy Prediction Using Machine Learning by Priyanka, Sonali Goyal & Ruby Bhatia.

[5] Reality mining and predictive analytics for building smart applications by Hiba Asri, Hajar Mousannif & Hassan Al Moatassime.

[6] A Miscarriage Prevention System Using Machine Learning Techniques by Sarmista Biswas & Samiksha Shukla.

[7] Asri H, Mousannif H, Al Moatassime H, Noel T. 'Big data in healthcare: Challenges and opportunities', in 2015 International Conference on Cloud Technologies and Applications (CloudTech).

[8] Lazer D, Kennedy R, King G, Vespignani A. The parable of google Flu: traps in big data analysis. Science. 2014;

[9] Tsuji K, et al. Book recommendation based on library loan records and bibliographic information. Procedia Soc Behav Sci. 2014;147:478–86.

[10] 'Internet Research Methods - Claire Hewson, Carl Vogel, Dianna Laurent - Google Livres'.

[11] Asri H. IoT and Reality Mining for Real-Time Disease Prediction. In: Azrour M, Irshad A, Chaganti R, editors. In IoT and Smart Devices for Sustainable EnviroCent. Cham: Springer; 2022.

[12] Paul A, Ahmad A, Rathore MM, Jabbar S. Smartbuddy: defining human behaviors using big data analytics in social internet of things. IEEE Wirel Commun. 2016;23(5):68–74.

[13] Construction of machine learning tools to predict threatened miscarriage in the first trimester based on AEA, progesterone and β-hCG in China: a multicentre, observational, case-control study by Jingying Huang, Ping Lv, Yunzhi Lian, Meihua Zhang, Xin Ge, Shuheng Li, Yingxia Pan, Jiangman Zhao, Yue Xu, Hui Tang, Nan Li & Zhishan Zhang.

[14] S. S. Aljameel, M. Aljabri, N. Aslam, D. M. Alomari, A. Alyahya et al., "An automated system for early prediction of miscarriage in the first trimester using machine learning," Computers, Materials & Continua, vol. 75, no.1, pp. 1291–1304, 2023.