# ML-Driven Pregnancy Care: A Data Science Approach to Maternal Health

Prof. A.M. Ghime [1], Karan Parge[2], Niraj Mahanta[3], Rajat Potgantiwar[4], Mayuresh Sontakke[5]
*[1] Professor, Dept. of Computer Engineering, TSSM BSCOER, PUNE*
*[2-5] Student, Dept. of Computer Engineering, TSSM BSCOER, PUNE*

***Abstract* — AI-powered systems, fueled by data science techniques, offer the ability to analyze diverse sets of health data from medical histories and lab results. By applying AI algorithms to this data, AI can predict potential pregnancy risks. Proposed data-driven approach bridges gaps in traditional maternal care, especially in remote or underserved areas where access to healthcare professionals may be limited. AI-powered pregnancy care represents a significant leap forward in maternal healthcare, offering innovative solutions for predicting risks, providing personalized care, and enhancing patient outcomes. The rise of AI provides a promising solution by offering automated, data-driven approaches for disease detection, leading to faster and more accurate diagnoses.AI serves as a powerful tool in the medical field, enabling the early and accurate detection of various diseases.**

***Key Words:*** **maternal health, data analytics, AI, Data mining, Maternal Risk Factors**

## I. INTRODUCTION

Maternity care is critical for ensuring the health and well-being of both mother and child. Early identification of risk factors during pregnancy is crucial for preventing complications. Traditional methods of maternity risk assessment rely heavily on medical expertise and manual analysis, which can be time-consuming and prone to human error. With the advent of AI, it is now possible to automate and enhance the prediction of maternity risk levels and potential maternal diseases, providing more accurate and timely interventions. Disease detection, a critical aspect of healthcare, involves identifying the presence of medical conditions at an early stage, which is crucial for timely intervention and treatment[1]. However, traditional diagnostic methods often rely on manual analysis of medical data by healthcare professionals, which can be time-consuming, prone to human error, and limited by the availability of medical experts. In recent years, the rapid advancement of AI has transformed various industries, and healthcare is no exception. Disease detection, a critical aspect of healthcare, involves identifying the presence of medical conditions at an early stage, which is crucial for timely intervention and treatment. This has led to significant improvements in both the speed and accuracy of disease detection, enhancing clinical decision-making and enabling early diagnosis of diseases [2]

## II. RELATED WORK

Pregnancy Line: A Visual Analysis System for Pregnancy Care and Risk Communication. Et.al. Li Li, Zhigui Liu, Xiao period for most women and a period that has significant impacts for newborns. Many pregnancies have an increased risk for complications for a variety of reasons[3]. Fortunately, early prenatal care and regular examination can help women have healthy pregnancies and deliver with fewer complications However, current research shows that many expectant mothers do not have a clear understanding of the pregnancy process due to the complexity of medical information and inadequate attention from caregivers. Therefore, in this work, we collected data on expectant mothers requirements and investigated the various medical treatments during pregnancy and the correlations between possible symptoms and newborn diseases[4]. Based on the above knowledge, we then develop visual analytic system named Pregnancy Line, which presents a pregnancy care plan and information on the medical examinations at each pregnancy stage using a series of vivid visual metaphors.[5]

Explainable AI based Maternal Health Risk Prediction using Machine Learning and Deep Learning. Et.al. Anika Rahman, Md. Golam Rabiul Alam. The significance of prenatal care is highlighted since effective maternal health care increases the probability of a successful pregnancy and a healthy baby[6]. One of the most crucial aspects of pregnancy is the diagnosis of high risk pregnancy, which can be extremely beneficial to expectant mothers. Early diagnosis can also lower maternal mortality and morbidity. The goal of this study is to apply machine learning and deep learning algorithms to determine the risk level based on pregnancy risk factors[7]. The analysis of risk factors in this study has been conducted using an existing dataset (Maternal Health Risk Data), and a comparison of various machine learning and deep learning algorithms reveals that the Gradient Boosting algorithm provides the highest accuracy in terms of risk level prediction, with an accuracy of 90.640percent. Additionally, we have implemented Explainable AI (LIME and SHAP) to discover the exact explanation of the prediction that has been made. [8]

Early Prediction of Maternal Health Risk Factors Using Machine Learning Techniques. Et.al. MD Assaduzzaman, Abdullah Al Mamuni. Nowadays, maternal health issues are one of the most challenging issues all over the world. Many women die each year during pregnancy and after delivery, which is a major cause of infant mortality. In rural areas, pregnant women face various difficulties and challenges, including a shortage of doctors, inadequate knowledge, a lack of public clinics, infrastructure issues, and transportation issues[9]. The mother's pregnancy is the major cause of the infant's poor health, rather than any other factors that may have arisen after childbirth. Significant roles are played by maternal risk factors such as the mother's chronic condition, age, nutrition, and other medical assistance during pregnancy. Recent developments in Artificial intelligence methods, particularly machine learning models, have made it easier to make predictions in a variety of disciplines. We can identify the primary maternal risk factors that can lead to newborn child and maternal mortality using machine learning techniques. [10]

Disease Prediction using Machine Learning Algorithms. Et.al. Sneha Grampurohit, Chetan Sagarnal[11].The development and exploitation of several prominent Data mining techniques in numerous real-world application areas (e.g. Industry, Healthcare and Bio science) has led to the utilization of such techniques in machine learning environments, in order to extract useful pieces of information of the specified data in healthcare communities, biomedical fields etc. The accurate analysis of medical database benefits in early disease prediction, patient care and community services[12]. The techniques of machine learning have been successfully employed in assorted applications including Disease prediction[13]. The aim of developing classifier system using machine learning algorithms is to immensely help to solve the health-related issues by assisting the physicians to predict and diagnose diseases at an early stage. A Sample data of 4920 patients' records diagnosed with 41 diseases was selected for analysis. [14]

## III. PROBLEM STATEMENT

To improve maternal healthcare outcomes, there is a critical need for an accurate and timely system that can assess maternity risk levels and detect diseases. Traditional methods for identifying complications during pregnancy rely on manual analysis of clinical data and expert judgment, which can be prone to human error and delays in diagnosis. These challenges are further amplified in resource-limited settings where access to skilled healthcare professionals is scarce.

## IV. OBJECTIVES

Pregnancy risks and complications early on by analyzing a variety of health data sources.

Healthcare interventions tailored to each expectant mother's unique conditions.

Maternal and fetal outcomes by providing timely, data-driven decisions for healthcare providers

## V. SYSTEM ARCHITECTURE

Fig.1 an architecture diagram is a graphical representation of a set of concepts, that are part of an architecture, including their principles, elements, and components. It gives the overall structure of the application.

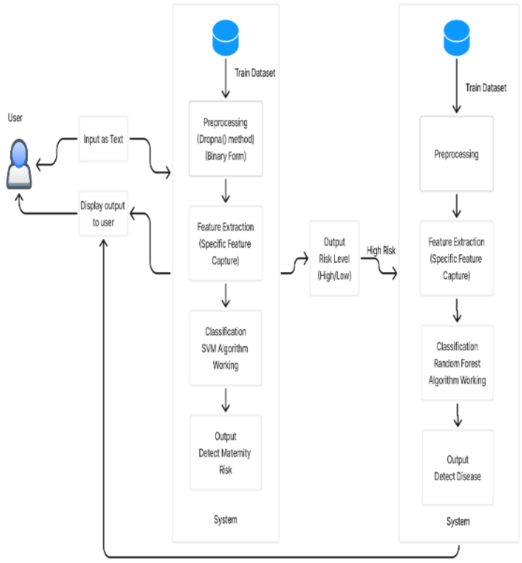


Fig 1. System Architecture

## VI. ANALYSIS AND DESIGN

1. Analysis Model
Our system uses machine learning models for predicting disease presence and pregnancy risk levels based on health parameters (Fig.2).

1.1 ML Models Used
Disease Prediction: Uses SVM to classify diseases like diabetes and hypertension based on blood biomarkers.

Risk Level Prediction: Uses Random Forest to classify pregnancies into low, medium, and high risk based on physiological data.

1.2 Model Evaluation

Accuracy, Precision, Recall, and F1 Score ensure reliable predictions.

2. System Design

2.1 Architecture
The system consists of:
1. Data Collection: Gathers patient data.
2. Preprocessing: Cleans and normalizes data.
3. ML Model Processing: Predicts diseases and risk levels.
4. Results Visualization: Displays insights on a dashboard.

2.2 Data Flow

1. User inputs health data.
2. ML model predicts disease & risk level.
3. Results displayed for healthcare use.

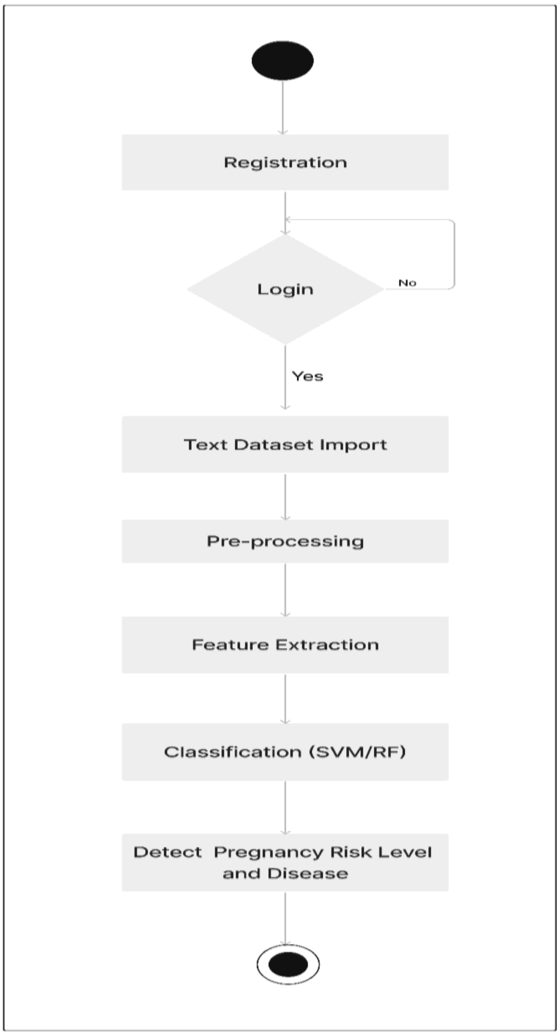


Fig 2. Activity Diagram

## VII. DISCUSSION

Maternity pain encompasses the physical and emotional challenges that women experience during pregnancy, labor, and the postpartum period. This pain can range from the common discomforts of pregnancy, to the intense contractions of labor. Labor pain is often described as one of the most intense forms of pain, driven by powerful uterine contractions and the stretching of tissues during childbirth. Following delivery, postpartum pain can continue, particularly if there were complications, tears, or if a cesarean section was performed. Disease detection is a critical aspect of healthcare, aiming to identify illnesses early to improve patient outcomes and prevent widespread transmission. With advancements in technology, disease detection has evolved significantly from traditional diagnostic methods to include modern approaches and AI-based systems.

## VIII. METHODLOGY

Data Collection:
For our ML-driven pregnancy care system, we have collected data from various hospitals, community clinics, and maternal healthcare facilities using a form-based risk monitoring system. The datasets have been curated to assess maternal health conditions and predict possible complications during pregnancy.

We have used two primary datasets to support our machine learning model:

1. Disease Prediction Dataset – Focuses on identifying specific diseases based on blood biomarkers[15].

- Dataset Dimensions: The dataset consists of 2351 samples with 25 key health-related features.

2. Risk Level Prediction Dataset – Predicts pregnancy risk levels based on key physiological parameters[16].

- Dataset dimensions: The dataset consists of 1014 samples with 6 key features.

Data Preprocessing:
To ensure accurate predictions in our ML-driven pregnancy care system, we performed the following preprocessing steps:

1. Data Cleaning
- Handled Missing Values using mean/mode imputation.
- Removed Duplicates to prevent bias.

2. Data Transformation
Feature Scaling: Min-Max Scaling for normalization.

3. Feature Engineering
Feature Selection: Used Correlation Analysis & Recursive Feature Elimination (RFE).
Feature Used: 14 out of 25 features are used.

- Glucose
- Cholesterol
- Hemoglobin
- Platelets
- WBC
- RBC
- Insulin
- BMI
- Systolic BP
- Diastolic BP
- HbA1c
- Heart Rate
- Troponin
- C-reactive Protein

4. Data Splitting

80% Training | 20% Testing for model evaluation.

Algorithms

1. SVM (Support Vector Machine)

Support Vector Machine (SVM) is a powerful supervised machine learning algorithm used for classification, including predicting pregnancy risk levels (e.g., "High Risk" or "Low Risk"). SVM works by finding the optimal boundary, called a hyperplane, that separates data points from different classes. This hyperplane is determined by maximizing the margin between the closest data points of each class, known as support vectors. The goal is to create the widest possible margin, making the model robust and accurate.
For pregnancy risk prediction, we use health parameters like age, systolic and diastolic blood pressure, blood sugar, body temperature, and heart rate. The data is first preprocessed through normalization to ensure all features are on the same scale. An SVM model is then trained using a labeled dataset, where each patient's risk level is known. Depending on the dataset's nature, we can use different kernel functions (like linear or radial basis function) to handle both linearly and non-linearly separable data.
Once trained, the SVM model can predict the risk level of new patients based on their health metrics. If the model's decision function value is positive, it classifies the patient as "High Risk"; otherwise, it classifies them as "Low Risk". This method is effective in identifying at-risk pregnancies, helping healthcare providers take preventive measures.

2. Random Forest

A supervised machine learning algorithm frequently utilized in classification and regression issues can be Random Forest. It is based on the idea of ensemble learning , which joins together several classifiers to address a complicated problem. By creating decision trees for several classifiers and using the majority vote for ranking, it enhances model performance. We get an accuracy of 90.148% from the train-test split method and 84.402% from the k-fold cross validation method.

## IX. CONCLUSION

The integration of artificial intelligence (AI) and data science into maternal healthcare represents a transformative approach to pregnancy care and disease detection. AI-driven solutions offer significant advantages over traditional healthcare models by enabling early detection of risk and more efficient use of healthcare resources. These technologies enhance the ability of healthcare providers to offer tailored, proactive, and timely interventions, ensuring better health outcomes for both mothers and babies. Maternity pain detection could focus on improving predictive accuracy through advanced AI models and expanding wearable technology integration for continuous, personalized monitoring. Multiple disease detection could explore federated learning for data privacy and improve multimodal data fusion for more accurate diagnostics.

## REFERENCES

[1] A. Singh, M. Bhatia, and A. Garg, “Prediction of ab- normal pregnancy in pregnant women with advanced maternal age and pregestational diabetes using machine learning models,” in 2022 12th International Conference on Cloud Computing, Data Science & Engineering (Con- fluence).

[2] E. P. Thong, D. P. Ghelani, P. Manoleehakul, A. Yesmin, K. Slater, R. Taylor, C. Collins, M. Hutchesson, S. S. Lim, H. J. Teede et al., “Optimising cardiometabolic risk factors in pregnancy: A review of risk prediction models targeting gestational diabetes and hypertensive disorders,” Journal of Cardiovascular Development and Disease.

[3] LI LI , ZHIGUI LIU, XIAO WANG, XIAOQIAN ZHANG, “PregnancyLine: A Visual Analysis System for Pregnancy Care and Risk Communication” 7 August 2022 Digital Object Identifier 10.1109/ACCESS.2022.3204694.

[4] 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.

[5] A.-M. Eloranta, I. Gunnarsdottir, B. Thorisdottir, G. Gunnlaugsson, B. E. Birgisdottir, I.Thorsdottir, and K. Einarsdo´ttir, “The combined effect of pre-pregnancy body mass index and gestational weight gain on the risk of pre-labour and intrapartum caesarean section—the ice- mch study”.

[6] Anika Rahman, Md. Golam Rabiul Alam, “Explainable AI based Maternal Health Risk Prediction using Machine Learning and Deep Learning” 2023 IEEE World AI IoT Congress (AIIoT) | 979-8-3503-3761-7/23/$31.00 ©2023 IEEE.

[7] Sayantan Saha, Argha Roy Chowdhuri et,al “Web Based DiseaseDetectionSystem”,IJERT,ISSN:22780 181,Vol.2 Issue 4, April-2023. P. Bansal, D. Verma, A. Bansal, and A. Verma, “Prenatal risk score in highrisk pregnancy cases and perinatal outcome: a study from south india,” International Journal of Reproduction, Contraception, Obstetrics and Gynecology.

[8] Qulan, J.R. 2021. "Induction of Decision Trees". Mach.Learn. 1,1 (Mar. 2021),81-10MinChen, Yixue Hao et.al “Disease Prediction by Machine Learning over big data from Healthcare Communities”, IEEE[Access 2023].

[9] MD Assaduzzaman, Abdullah Al Mamun "Early Prediction of Maternal Health Risk Factors Using Machine Learning Techniques" 2023 International Conference for Advancement in Technology (ICONAT) Goa, India. Jan 24-26, 2023.

[10] Sayantan Saha, Argha Roy Chowdhuri et,al "Web Based Disease Detection System", IJERT, ISSN:22780181, Vol.2 Issue 4, April-2023.

[11] S. Grampurohit and C. Sagarnal, "Disease Prediction using Machine Learning "Algorithms," 2020 International Conference

for Emerging Technology (INCET), Belgaum, India, 2020, pp. 1-7, 10.1109/INCET49848.2020.9154130.

[12] Shadab Adam et.al "Prediction system for Heart Disease using Naïve Bayes", International Journal of advanced Computer and Mathematical Sciences, ISSN 2230-9624, Vol 3, Issue 3,2022,pp 290 294[Accepted-12/06/2022].

[13] MinChen, Yixue Hao et.al "Disease Prediction by Machine Learning over big data from Healthcare Contaccess 2023]

[14] Mr Chintan Shah, Dr. Anjali Jivani, "Comparison Of Data Mining Classification Algorithms for Breast Cancer Prediction", IEEE-31661.

[15] https://www.kaggle.com/datasets/ehababoelnaga/multiple-disease-prediction

[16] https://www.kaggle.com/datasets/csafrit2/maternal-health-risk-data

[17] E. A. Rodríguez, F. E. Estrada, W. C. Torres, and J. C. M. Santos. "Early Prediction of Severe Maternal Morbidity Using Machine Learn-ing Techniques," Lecture Notes in Computer Science, pp. 259-270, 2016, doi: 10.1007/978-3-319-47955

[18] M. Podda, D. Bacciu, A. Micheli, R. Bellü, G. Placidi. and L Gagliardi, "A machine learning approach to estimating preterm infants survival: development of the Preterm Infants Survival Assessment (PISA) predictor," Scientific Reports, vol. 8, no. 1. Sep. 2018, doi: 10.1038/s41598-018-31920-6.

[19] L. C. Y. Poon, N. A. Kametas, N. Maiz. R. Akolekar, and K. H. Nicolaides, "First-Trimester Prediction of Hypertensive Disorders in Pregnancy," Hypertension, vol. 53, no. 5, pp. 812-818, May 2009, doi: 10.1161/hypertensionaha. 108.127977.

[20] B. Farran, A. M. Channanath, K. Behbehani, and T. A. Thanaraj. "Predictive models to assess risk of type 2 diabetes, hypertension and comorbidity: machine-learning algorithms and validation using national health data from Kuwait-a cohort study." BMJ open, vol. 3. no. 5, May 2013, doi: 10.1136/bmjopen-2012-002457.

[21] A. R. Yarlapati, S. Roy Dey, and S. Saha, "Early Prediction of LBW Cases via Minimum Error Rate Classifier: A Statistical Ma-chine Learning Approach," 2017 IEEE International Conference on Smart Computing (SMARTCOMP), May 2017, doi: 10.1109/smart-comp.2017.7947002.

[22] 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." 2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 2014, pp. 4659-4662, doi: 10.1109/EMBC.2014.6944663

[23] W. Cheng, L. Fang, L. Yang, H. Zhao, P. Wang and J. Yan, "Varying Coefficient Models for Analyzing the Effects of Risk Factors on Pregnant Women's Blood Pressure," 2014 13th International Confer-ence on Machine Learning and Applications, 2014, pp. 55-60. doi: 10.1109/ICMLA 2014.14.

[24] L. K. Woolery and J. Grzymala-Busse. "Machine learning for an expert system to predict preterm birth risk.." Journal of the American Medical Informatics Association, vol. 1, no. 6, p. 439, Nov. 1994, doi: 10.1136/jamia. 1994.95153433.

[25] B. N. Lakshmi, T. S. Indumathi and N. Ravi. "A comparative study of classification algorithms for predicting gestational risks in pregnant women," 2015 International Conference on Comput-ers. Communications, and Systems (ICCCS), 2015, pp. 42-46, doi: 10.1109/CCOMS.2015.7562849.

[26] L. C. Kenny, W. B. Dunn, D. I. Ellis, J. Myers, P. N. Baker. and D. B. Kell, "Novel biomarkers for pre-eclampsia detected using metabolomics and machine learning," Metabolomics, vol. 1, no. 3, pp. 227-234, Jul. 2005, doi: 10.1007/s11306-005-0003-1