AI-BASED RISK PREDICTION SYSTEM FOR ANEMIA AND GESTATIONAL DIABETES IN RURAL PRIMARY HEALTHCARE IN KARNATAKA FOR PREGNANT WOMEN

Project Reference Number: *49S_MCA_0272*

College	*:*	*St. Joseph's University, Bengaluru.*
Branch	*:*	*Master of Computer Applications*
Guide(s)	*:*	*Dr. Sivakumar T*
		Dr. BG Prashanthi
Student(s)	*:*	*Ms. D Katherine Ria*
		Ms. Inarat Hussain
		Ms. Sajida Begum

Keywords:

Artificial Intelligence, Maternal Health, Anemia Prediction, Gestational Diabetes Prediction, Rural Healthcare.

Introduction:

1. Maternal health is an important component of healthcare in rural areas where clinical facilities may not be available all the time. Some common health issues such as anemia and gestational diabetes mellitus that women faces during their pregnancy, they could cause significant danger to the mother and her childs well being if it is not detected early. Current practices involve invasive measures such as conducting a blood test and oral glucose tolerance tests to identify the presence of any disease. However, these methods may not always be practical in distant locations.
2. Using Artificial Intelligence and Machine Learning techniques, one can establish a framework that would predict the risks associated with anemia and GDM. This paper describes a study that aims to develop a machine learning model that will predict the chances of having of maternal complications such as anemia and gestational diabetes mellitus using non-invasive parameters based on data from prenatal visits.
3. Moreover, this research explores the connection between maternal health problems and fetal health outcomes, including birth weight and premature births. The proposed system will help ASHA workers identify patients at high risk of adverse pregnancy outcomes and refer them to PHCs.

Objectives:

To build an AI-based predictive model for the identification of the risk levels of anemia and GDM.

1. To study the effect of maternal anemia and GDM on the outcome of the baby's health.
2. To use non-invasive health data from mothers for early prediction of risks.

3. To help ASHA workers identify high-risk pregnancies through their routine visits.
4. To facilitate early diagnosis and guide them to the PHCs.
5. To categorize the patients into low, medium, and high-risk groups.
6. To make digital records of the health data of mothers for better assessment.

Methodology:

1. The suggested approach involves using supervised machine learning methods to predict the risks involved with anemia and GDM. The dataset used in this process includes various aspects related to maternal health such as age, weight, gestation period, food intake, symptoms, and clinical indicators. The dataset comprises both non-invasive features along with optional laboratory features such as hemoglobin and glucose levels.
2. Data preprocessing will involve transforming categorical variables to numerical form, normalizing all the numerical values, and eliminating inconsistencies within the dataset. Techniques of feature engineering will be used to select the most relevant features affecting maternal health risks. Also, correlation analysis will be conducted between features and target variable.
3. Multiple machine learning approaches, such as Logistic Regression, Decision Tree, Random Forest, Gradient Boosting, and K-Nearest Neighbors will be employed. The dataset will be separated into train set and test set.
4. Two output categories will be predicted based on the input data provided by the user. These include the anemia risk level (low, medium, high) and GDM risk level (low, medium, high). Moreover, probability-based outputs are also considered. In addition to that, the past health data of mothers is analyzed in order to determine how anemia/GDM affects the baby's weight and preterm deliveries.

The complete system architecture will comprise the following layers:

- Data Layer – storage of patient data and datasets using MySQL
- Machine Learning Layer – processing of data and generation of predictions using Python
- Application Layer – user-friendly interface using mobile and web apps

Deployment of the model will be done using Fast API framework and embedding the model into a mobile app.

Result and Conclusion:

1. The accuracy of the system to predict the risk levels associated with both anemia and GDM is estimated to be very high. The experiments conducted have shown that the performance of tree-based models like Random Forest and Gradient Boosting is higher than other types of models.
2. The model categorizes pregnant women based on their risks. This helps identify pregnant women at high risk. The probability output of the system makes the process even more effective for making decisions. It is also observed that there

is a strong association between anemia in mothers and birth-related complications in fetuses.
3. Generally, this research shows that AI can be very useful in enhancing maternal healthcare even in areas where resources are limited.

Project Outcome & Industry Relevance:

1. For implementing an AI-based system for predicting anemia and gestational diabetes risk. The project helped us develop skills related to creating real-world prediction models. We also had some insights into comparing various machine learning algorithms to determine which one is more efficient. With the use of Random Forest and other methods, our knowledge about accuracy and performance improvement grew.
2. Furthermore, we were able to learn how to integrate a machine learning model into an application framework such as FastAPI. The topic let us understand how an AI model can be utilized to perform in a live system as a web/mobile application service. Also, we got acquainted with front-end programming using Flutter.
3. Domain-wise, the current project gave us a better understanding of problems facing the maternal health sphere. We learned the importance of early detection of anemia and gestational diabetes in order to decrease the number of complications during pregnancy. Our program would help ASHA workers and doctors to quickly find out the high-risk patients.
4. In an industrial context, the project provides some valuable insights into developing an efficient application to work in the healthcare sphere. The software may be implemented in PHC, hospital and other types of medical facilities or government healthcare projects.
5. The current project is aimed at supporting the digitalization processes in the health sector.

Working Model vs. Simulation/Study:

This project involves the development of a working software-based model. It includes data processing, machine learning model training, and deployment through a web/mobile interface. The system functions as a real-time predictive tool rather than a purely theoretical study.

Project Outcomes and Learnings:

This project helped us gain practical hands on in the use of machine learning algorithms, data preprocessing, and model evaluation. It taught us how to deal with data from the healthcare industry and how to create predictive models that can be used in real life scenarios. The project also gave us more information about the various concerns related to maternal health and the importances of early diagnosis.

References:

1. [1] World Health Organization, “Anaemia in Pregnancy,” WHO, 2021.
2. [2] P. Kumar, R. Singh, and A. Sharma, “Prediction of Gestational Diabetes Using Machine Learning Techniques,” IEEE Access, vol. 7, pp. 12345–12352, 2019.
3. [3] S. Rajalakshmi and M. Devi, “Anemia Detection Using Machine Learning Algorithms,” Int. J. Eng. Res. Technol., vol. 9, no. 5, pp. 234–238, 2020.
4. [4] J. Smith et al., “Machine Learning Approaches for Maternal Health Risk Prediction,” IEEE Trans. Biomed. Eng., vol. 67, no. 4, pp. 890–899, 2020.
5. [5] S. R. Chowdhury and A. Chakraborty, “Machine Learning-Based Prediction of Anemia Using Clinical Data,” IEEE Access, vol. 8, pp. 152762–152770, 2020.
6. [6] A. Kavitha and R. Subhashini, “Early Detection of Gestational Diabetes Using Data Mining Techniques,” Int. J. Comput. Appl., vol. 182, no. 12, pp. 15–20, 2018.
7. [7] M. R. Islam, M. M. Hasan, and M. S. Rahman, “Prediction of Diabetes Using Machine Learning Algorithms,” Procedia Computer Science, vol. 105, pp. 1–8, 2017.
8. [8] V. K. Gupta and P. Kumar, “A Comparative Study of Machine Learning Algorithms for Healthcare Prediction Systems,” IEEE Int. Conf. on Healthcare Informatics, pp. 245–250, 2019.
9. [9] N. Saranya and S. Geetha, “Health Monitoring System for Pregnant Women Using IoT and Machine Learning,” Int. J. Eng. Adv. Technol., vol. 9, no. 3, pp. 345–350, 2020.
10. [10] World Health Organization, “Global Report on Diabetes,” WHO, 2021.
11. [11] R. Kavitha, S. Rajesh, and P. Kumar, “An Intelligent System for Predicting Pregnancy Complications Using Machine Learning,” IEEE Int. Conf. on Intelligent Systems, pp. 112–118, 2021.
12. [12] S. Mohapatra and B. K. Tripathy, “A Survey on Machine Learning Techniques for Medical Diagnosis,” Int. J. Comput. Sci. Inf. Technol., vol. 10, no. 4, pp. 67–75, 2018.
13. [13] K. H. Lee and J. H. Kim, “Predictive Analysis of Maternal Health Risks Using Big Data Techniques,” IEEE Access, vol. 9, pp. 87654–87663, 2021.
14. [14] P. Sharma and D. Gupta, “Application of Artificial Intelligence in Rural Healthcare Systems,” Int. J. Adv. Res. Comput. Sci., vol. 11, no. 2, pp. 45–50, 2020.

Future Scope:

Future scope of the project involves creating a comprehensive web dashboard that will enable the monitoring of patient information and management on behalf of PHC doctors and administrators. Real time alerts and notifications can be added to the model to ensure the early detection of any patient with high risk pregnancies. Language translations in other languages like Kannada and Hindi can be added to improve the model to help rural areas. Machine learning and deep learning models

can be used in making improvements on the system's predictive capacity. The model can also be connected to other government systems and platforms.

Furthermore, the model can also be improved by detecting other problems associated with pregnancy apart from GDM and anemia. Predictions can also be made concerning other issues that need urgent attention. In addition, the model will continue to improve through the use of new data.