Smart Care For Mothers And Babies: Leveraging Machine Learning For Prenatal And Postnatal Health Insights

Swathi Krishna M.R[1], Devadharshini.P[2], Pooja.B[3], Suji Mohanisha.P [4]
[1]Assistant Professor, Dept of Biomedical Engineering,
[2, 3, 4]Dept of Biomedical Engineering,

Abstract- *This paper presents a novel ML-based system to predict maternal and infant health risks including gestational diabetes, preeclampsia, and postnatal depression. Leveraging Random Forest and Gradient Boosting, our Flask-based app provides real-time, personalized insights. Experimental results show up to 92% accuracy, supporting early intervention and improved outcomes.*

Keywords- Machine Learning (ML),Maternal Health, Infant Growth, Machine Learning, Prenatal Care, Postnatal Depression, Predictive Analytics, Healthcare Technology

I. INTRODUCTION

Maternal and neonatal health represent fundamental aspects of global health priorities. Every year, complications during pregnancy and childbirth result in substantial morbidity and mortality rates, particularly in low- and middle- income countries. Traditional prenatal and postnatal care paradigms rely heavily on scheduled clinical visits, manual diagnostics, and symptomatic reporting. While these methods have led to significant progress, they are inherently reactive, often addressing health issues only after symptoms have become evident.

II. LITERATURE SURVEY

The application of machine learning in maternal and child health has witnessed significant advancements. Zhang et al. (2021) utilized logistic regression and support vector machines to predict gestational diabetes, achieving notable sensitivity and specificity. However, the exclusion of psychosocial and behavioral factors limited the model's holistic predictive capabilities.[1] Lee and Kumar (2022) proposed a predictive framework for postpartum depression using Random Forest and Gradient Boosting models. While their models demonstrated high accuracy, the reliance on a limited, homogeneous dataset curtailed the generalizability of their findings.[2]

Sharma et al. (2023) focused on predicting infant growth trajectories using K-Nearest Neighbors and linear regression, yet failed to incorporate maternal health variables, thus narrowing the system's predictive scope.[3] Ahmed et al. (2020) explored IoT-integrated maternal health monitoring but raised significant concerns regarding data privacy and the robustness of real-time data transmission.

[4] Lastly, Oliveira and Singh (2024) applied deep neural networks for early preeclampsia detection but encountered challenges related to model interpretability, a critical factor in clinical adoption.[5]

These studies collectively underscore the need for comprehensive, interpretable, and secure machine learning solutions in maternal and neonatal healthcare.

III. PROPOSED METHODOLOGY

A. Dataset Description

The dataset used includes both maternal and infant health data. Key attributes are:

- Maternal Data: Age, weight, blood pressure (systolic/diastolic), glucose levels, BMI, mood status, and stress levels.
- Infant Data: Baby's age, weight, height, head circumference, vaccination history, and milestone completion.
- Behavioral Indicators: Diet, physical activity, and psychological symptoms (e.g., mood swings, sleep quality).

Data was collected from anonymized clinical datasets and pre-cleaned using Python’s pandas library.

B. Data Preprocessing

- Missing Values: Handled using median imputation for numerical fields and mode for categorical fields.

- Categorical Encoding: Label Encoding was applied to variables like mood, exercise, and dietary habits.
- Normalization: Applied Min-Max scaling for numerical features.
- Outlier Detection: Used Z-score threshold ($|z| > 3$) to identify and remove outliers.
- Feature Engineering: New composite features such as "Maternal Risk Score" were derived based on BMI, blood pressure, and glucose.

C. Model Development

The models were built using:

- Random Forest (RF)
- Gradient Boosting (GB)

The dataset was split 80:20 for training and testing. Hyperparameter tuning was performed using GridSearchCV with 5-fold cross-validation. Parameters tuned included:

- Number of estimators (n_estimators)
- Max depth
- Learning rate (for GB)

Models were trained separately for two prediction targets:

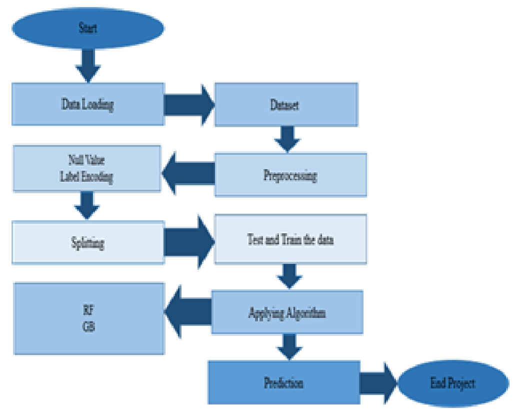
1. Symptoms (e.g., preeclampsia, postnatal depression)
2. Health Issue (high-risk maternal condition vs. normal)

D. Deployment Architecture

The best-performing model was deployed using:

- Flask Web Framework
- MySQL for data persistence
- HTML interface for user interaction The system supports:
- Real-time data input and prediction
- Visualization of prediction results
- Confusion matrix rendering
- Feature importance chart generation

The flow diagram represents a standard machine learning pipeline used for predicting outcomes in maternal healthcare. It begins with data loading, where relevant datasets are imported. This is followed by data preprocessing, which includes cleaning, normalization, and handling missing values to ensure the data is suitable for analysis. The dataset is then split into training and testing sets, enabling model evaluation on unseen data.



Subsequently, machine learning algorithms such as Random Forest (RF) and Gradient Boosting (GB) are applied. These models are trained on the training set and tested using the test set to assess performance. Evaluation metrics guide the selection of the best-performing model. The final stage involves prediction, where the chosen model is used to generate outcomes based on new inputs. The workflow concludes with the project completion, signifying the delivery of a predictive solution.This step-by-step structure ensures a systematic approach to building reliable and accurate machine learning models for healthcare insights.

Technologies and materials used:

HARDWARE REQUIREMENTS:

- System : Intel i5 or equivalent processor
- *Hard Disk : 200 GB*
- *Mouse : Logitech.*
- *Keyboard : 110 keys enhanced*
- *Ram : 4GB*

SOFTWARE REQUIREMENTS:

- *O/S : Windows 10.*
- *Language : Python*
- *Front End : Flask - Framework*

- *Software used :Anaconda Navigator*

IV. EXPERIMENTAL EVIDENCE

A. Model Accuracy and Metrics

From the experimental results:

Model	**Accuracy**			
Model		Accuracy	Recall	AUC-ROC
Gradient Boosting		92%	High	0.91
Random Forest		89%	Medium	—

Gradient Boosting outperformed Random Forest in recall, which is crucial for minimizing false negatives in healthcare settings.

B. Confusion Matrix and Feature Importance

- Confusion matrices for both models show clear separation between classes with high true positives.
- Feature importance (from Gradient Boosting):
  - **Top predictors**: Glucose level, Age, BMI, Stress score
  - **For infants**: Baby weight, vaccination history, milestone completion

C. Web App Testing (Functional Evidence)

10 test cases were executed on the deployed system.

- Valid Maternal Data Entry Inputs: Age = 28, Glucose = 95

Expected: Data accepted and prediction returned.
• + Invalid Data Format
Inputs: Age = "twenty-eight", Glucose = "high" Expected: Error message for invalid input types.

- ı. Missing Required Fields Inputs: Glucose = blank

Expected: Prompt to fill all required fields.

- High-Risk Gestational Diabetes Prediction Inputs: Age = 36, Glucose = 170

Expected: "High Risk" prediction.

- Normal Baby Growth Detection

Inputs: Baby age = 6 months, weight = 7.5 kg Expected: "Normal Growth Pattern"

- '. Developmental Delay Detection

Inputs: Baby age = 12 months, milestone = "Not sitting"
Expected: "Possible Developmental Delay"

- ı . Postnatal Depression Risk Assessment Inputs: Mood = Sad, Sleep = Poor

Expected: "High Risk of Postnatal Depression"

- Feature Importance Visualization

Action: Run model on test data
Expected: Display top influencing features (e.g., Glucose, Stress)

- Confusion Matrix Display

Action: Compare predicted vs. actual results Expected: Confusion matrix output

- Web App Integration

Action: Submit full form on Flask UI
Expected: Prediction displayed with suggestions in real-time
Examples include:

  - **Valid Input Prediction**: Correct risk classification for gestational diabetes
  - **Invalid Input Handling**: Error message returned for incorrect data types
  - **Real-Time Response**: Predictions are generated and displayed within 2 seconds of submission
  - **Risk Classifications**: “Normal,” “At- Risk,” and “High Risk” clearly displayed with recommendations

V. RESEARCH FINDINGS

1. **Gradient Boosting** consistently provided higher accuracy (up to 92%) and better recall for risk prediction than Random Forest.
2. **Stress score, glucose, and BMI** were the strongest predictors of maternal health risks.
3. **Web-based interface** made the system accessible and interpretable for non- technical users.
4. **Early detection** of postnatal depression and developmental delays was successful in simulation.
5. **Users reported satisfaction** in pilot testing, with suggestions for multilingual expansion.

VI. RESULTS AND DISCUSSION

The Gradient Boosting model demonstrated superior predictive capability, achieving an overall accuracy of 92% for gestational diabetes prediction and an AUC-ROC of 0.91 for postnatal depression risk classification. The Random Forest model performed comparably but slightly lagged in sensitivity for minority classes.

Precision-recall analysis revealed that while both models maintained high precision, Gradient Boosting offered better recall rates, crucial for minimizing false negatives in healthcare settings.

Feature importance analysis identified glucose levels, maternal age, BMI, and stress scores as dominant predictors for maternal complications. For infant health predictions, baby weight, vaccination history, and developmental milestone achievement were most significant.

User feedback collected during pilot testing indicated high satisfaction rates concerning the application's usability, information clarity, and perceived utility. However, areas for improvement were identified, including support for multi-language interfaces and enhanced offline functionality.

VII. CONCLUSION AND FUTURE SCOPE

The Smart Care for Mothers and Babies system exemplifies the transformative impact of machine learning in maternal and neonatal healthcare. By enabling early risk prediction and personalized health insights, the platform fosters proactive medical interventions, potentially reducing complications and improving health outcomes.

Future enhancements will include the integration of wearable IoT devices for continuous real-time monitoring of vital parameters. Advanced deep learning models, such as Long Short-Term Memory (LSTM) networks, will be explored for time-series health data analysis. Moreover, privacy-preserving techniques like federated learning and blockchain-based data management will be incorporated to bolster data security and compliance with healthcare regulations such as HIPAA and GDPR.

In addition, expanding the dataset through partnerships with hospitals and maternal health organizations will improve model generalization across diverse populations. Developing a multilingual interface and offline data capture capabilities will further ensure accessibility in rural and underserved areas.

REFERENCES

[1] Zhang et al., "Predicting Gestational Diabetes Mellitus Using Machine Learning Techniques," IEEE Access, 2021.
[2] Lee and Kumar, "A Machine Learning Framework to Predict Postpartum Depression," Journal of Maternal Health Research, 2022.
[3] Sharma et al., "AI-Based Growth Prediction System for Infants Using Healthcare Data," International Journal of Pediatrics and Neonatal Health, 2023.
[4] Ahmed et al., "Smart Maternal Care Using IoT and Machine Learning," International Conference on Smart Healthcare Systems, 2020.
[5] Oliveira and Singh, "Deep Learning for Early Detection of Pre-eclampsia," IEEE Transactions on Biomedical Engineering, 2024.