International Journal for Multidisciplinary Research (IJFMR)
E-ISSN: 2582-2160 • Website: www.ijfmr.com • Email: editor@ijfmr.com

PregBot: an ML and NLP-Based Intelligent System Supporting Women and Families during Pregnancy

Malavath Sai Deepak[1], Gunti Ashok[2], Banoth Sachin[3]

[1,2,3]Department of Computer Science and Engineering, Institute of Aeronautical Engineering, Dundigal, Hyderabad – 500043, Telangana, India

Abstract
Pregnancy involves complex physical and emotional changes that demand accurate information and timely support. Many expectant mothers experience difficulty in obtaining immediate guidance, which may result in confusion, stress, and delayed healthcare decisions. This study introduces *PregBot*, an AI-driven intelligent assistant designed using Machine Learning (ML) and Natural Language Processing (NLP) to provide continuous support for pregnant women and their families. The system is capable of understanding natural language queries, identifying user intent, evaluating reported symptoms, and estimating potential risk levels. NLP techniques enable conversational interaction and contextual interpretation, while ML models perform symptom severity classification and predictive analysis. In addition, a sentiment analysis component detects emotional states to generate more adaptive and supportive responses. The proposed framework focuses on delivering accessible, real-time, and personalized assistance to enhance maternal awareness and well-being. Performance evaluation indicates accurate intent detection, dependable risk assessment, and effective user interaction, demonstrating the potential of intelligent conversational systems in pregnancy-related healthcare applications.

Keywords: Machine Learning (ML), Natural Language Processing (NLP), Healthcare Chatbot, Maternal Health Assistance, Symptom Analysis, Risk Prediction, Emotion Detection

1. Introduction

Pregnancy represents one of the most sensitive and transformative phases in a woman's life, involving substantial physiological, psychological, and emotional changes. Ensuring maternal well-being during this period requires timely access to accurate medical information, continuous monitoring, and appropriate guidance. However, many expectant mothers encounter challenges such as limited availability of healthcare professionals, delayed consultations, inadequate awareness of symptoms, and heightened anxiety regarding health-related concerns. These limitations may contribute to stress, misinformation, and delayed decision-making, potentially affecting both maternal and fetal health.

Recent advancements in Machine Learning (ML) and Natural Language Processing (NLP) have enabled the development of intelligent systems capable of analyzing data, understanding human language, and delivering context-aware responses. ML techniques facilitate predictive analysis and classification of health-related conditions, while NLP enables natural language interaction, allowing users to communicate with systems in a human-like manner. The integration of these technologies offers promising opportunities for designing accessible and responsive healthcare support tools.

This research presents *PregBot*, an ML and NLP-based intelligent decision support system designed to assist pregnant women and their families. The proposed system focuses on interpreting user queries, identifying intent, analyzing symptoms, estimating potential risk levels, and recognizing emotional context through sentiment analysis. By providing real-time, personalized, and easily accessible assistance, the system aims to bridge gaps in conventional healthcare support mechanisms. The study emphasizes the role of computational intelligence in enhancing maternal awareness, reducing anxiety, and supporting informed healthcare decisions during pregnancy.

2. Problem Statement

Pregnancy is a medically sensitive and dynamic condition requiring continuous monitoring, accurate health awareness, and timely decision-making. Despite the growing availability of digital healthcare resources, many pregnant women still encounter challenges in obtaining immediate, reliable, and context-specific guidance. Conventional pregnancy care primarily depends on scheduled clinical consultations, which may not adequately address real-time concerns such as symptom interpretation, lifestyle adjustments, emotional well-being, and early risk awareness.

Existing pregnancy support applications largely provide generalized information, static recommendations, or rule-based chatbot responses. These systems typically lack intelligent mechanisms for symptom evaluation, personalized guidance, emotional state assessment, and predictive risk analysis. Moreover, the absence of explainability and limited involvement of healthcare professionals reduce system trustworthiness and clinical relevance.

The variability and complexity of pregnancy-related symptoms further amplify the need for an intelligent decision support framework capable of understanding natural language queries, analyzing multiple health parameters, predicting potential risk levels, and providing transparent reasoning. Therefore, there exists a critical need for an integrated computational system that delivers accessible, personalized, and context-aware pregnancy assistance by leveraging Machine Learning and Natural Language Processing techniques.

3. Objectives

The primary objective of this research is to design and develop an intelligent decision support system that enhances pregnancy care through computational analysis and natural language interaction. The specific objectives of the study are:

- To develop a Natural Language Processing-based conversational interface capable of accurately interpreting pregnancy-related user queries
- To implement Machine Learning models for symptom evaluation and pregnancy risk level classification
- To integrate sentiment analysis techniques for recognizing emotional context and improving response adaptability
- To design an explainable prediction mechanism that improves transparency and user trust
- To provide real-time, personalized, and accessible pregnancy guidance

4. Existing Systems

Current pregnancy support solutions primarily consist of mobile applications, informational websites, and rule-based conversational agents. These systems typically provide generalized pregnancy information, weekly updates, diet suggestions, and symptom descriptions. While such platforms improve basic awareness, their functionality remains largely static and non-adaptive.

Several applications incorporate chatbot interfaces; however, most rely on predefined rules or keyword-based matching techniques. These approaches limit the system's ability to interpret complex natural language queries, analyze symptom combinations, or provide personalized decision support. Furthermore, existing systems rarely integrate predictive analytical mechanisms for risk assessment or emotional state recognition.

Although these solutions offer convenience and accessibility, they lack computational intelligence required for dynamic symptom evaluation, contextual understanding, and adaptive healthcare assistance.

5. Limitations of Existing Systems

Despite the availability of multiple digital pregnancy platforms, several limitations persist:

- Absence of personalized, context-aware recommendations
- Lack of predictive symptom risk classification
- Limited natural language understanding capabilities
- Minimal integration of emotional or sentiment analysis
- Insufficient transparency in system responses
- Limited involvement of healthcare professionals

These limitations highlight the necessity for an integrated intelligent decision support system capable of providing adaptive, interpretable, and user-centric pregnancy assistance.

6. Research Gap

Although prior studies have explored pregnancy monitoring systems, risk prediction models, and healthcare chatbots, most existing solutions focus on isolated functionalities. Limited research has addressed the integration of conversational systems, predictive analytics, emotional assessment, and explainable decision mechanisms within a unified pregnancy support platform. Furthermore, many systems lack personalization, real-time assistance, and clinician-oriented validation modules. This research aims to bridge these gaps through a comprehensive computational framework.

7. Research Contribution

This research makes the following contributions:

- Development of an integrated ML and NLP-based pregnancy decision support framework
- Implementation of symptom-aware risk prediction using supervised learning models
- Incorporation of sentiment analysis for emotion-sensitive interaction
- Design of explainable prediction outputs for enhanced system interpretability
- Integration of patient, chatbot, and doctor interaction modules

8. Proposed System

The proposed system, *PregBot*, is designed as an intelligent decision support framework that provides continuous, personalized, and context-aware assistance for pregnant women and their families. The system integrates Machine Learning (ML) and Natural Language Processing (NLP) techniques to enable symptom understanding, predictive risk classification, and adaptive conversational interaction.

Unlike traditional pregnancy applications that offer generalized guidance, the proposed system emphasizes computational intelligence for analyzing user inputs, interpreting symptom descriptions, estimating potential risk levels, and delivering informative recommendations. The NLP module enables natural language interaction, allowing users to communicate queries in an intuitive manner. Machine Learning models are employed to evaluate symptom patterns and classify pregnancy risk levels into Low, Moderate, or High categories.

Additionally, the system incorporates a sentiment analysis component to detect emotional states such as stress or anxiety, enabling supportive and context-sensitive responses. An explainability mechanism is integrated to provide human-readable reasoning behind predictions, thereby improving transparency and user trust. A dedicated doctor interface further ensures clinical oversight by enabling healthcare professionals to review alerts and provide expert recommendations.

The PregBot framework is implemented as a web-based application to ensure accessibility, scalability, and ease of use.

8.1. Merits of Proposed System

The proposed system offers several advantages over existing solutions:

- Provides personalized and context-aware pregnancy guidance
- Enables predictive symptom risk classification
- Supports natural language-based interaction
- Incorporates emotion-sensitive response generation
- Enhances system transparency through explainability
- Facilitates clinician involvement and validation
- Ensures real-time, accessible assistance

9. System Architecture

The PregBot system follows a modular, layered architecture designed to ensure scalability, maintainability, and efficient data processing. The architecture consists of the following components:

- **User Interaction Layer** – Captures patient and doctor inputs through a web interface
- **Application Layer** – Handles authentication, session management, and system logic
- **NLP Processing Module** – Performs text preprocessing, intent recognition, and entity extraction
- **Machine Learning Module** – Executes symptom risk prediction and classification
- **Sentiment Analysis Module** – Detects emotional context
- **Explainability Module** – Generates interpretable prediction reasoning
- **Database Layer** – Stores user profiles, symptoms, and prediction records

This modular design enables independent processing of conversational analysis, predictive modeling, and response generation.

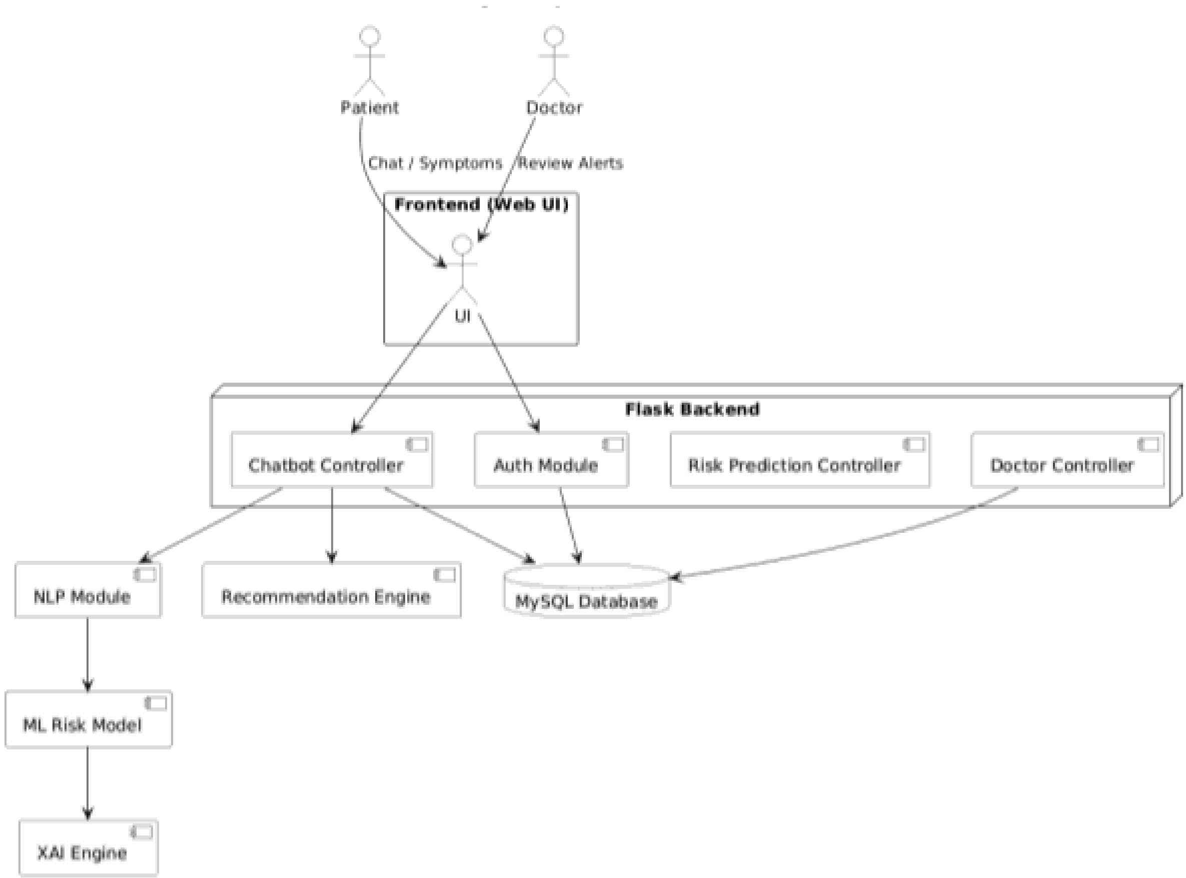

Figure: 1

10. Working Principle

The PregBot system operates through a structured processing pipeline. User queries are first captured through the conversational interface and forwarded to the NLP module, where text preprocessing, intent classification, and entity recognition are performed. If symptom-related information is detected, relevant features are extracted and passed to the Machine Learning prediction module.

The ML module evaluates symptom patterns and health parameters to classify risk levels. Simultaneously, the sentiment analysis component assesses emotional tone to enhance response relevance. The explainability mechanism then generates human-readable reasoning for predictions. Finally, the system produces personalized recommendations or alerts, while critical cases are forwarded to the doctor interface.

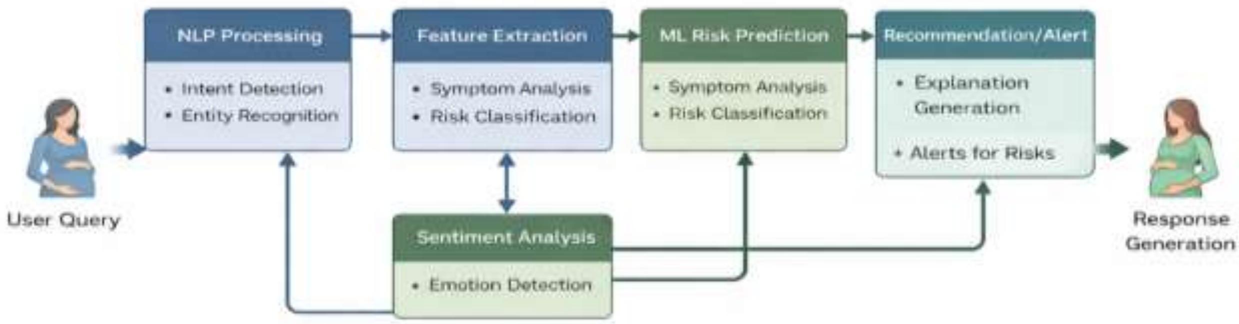

Figure: 2

11. Methodology

The development of the PregBot system follows a structured and modular design methodology aimed at integrating Natural Language Processing (NLP), Machine Learning (ML), sentiment analysis, and explainability mechanisms into a unified decision support framework. The methodology emphasizes systematic data processing, predictive modeling, and adaptive response generation.

Initially, user inputs are captured through a conversational interface designed to support natural language interaction. The input text undergoes preprocessing operations including tokenization, normalization, and stop-word removal to enhance linguistic clarity. Feature extraction is then performed using Term Frequency–Inverse Document Frequency (TF-IDF), enabling numerical representation of textual data.

For query understanding, supervised classification models such as Support Vector Machine (SVM) and Logistic Regression are employed to perform intent recognition. Entity extraction techniques identify relevant medical attributes including symptoms, severity indicators, and contextual parameters.

When symptom-related information is detected, structured feature vectors are generated by combining user profile parameters (e.g., age, trimester, health conditions) with symptom characteristics. These features are processed by Machine Learning models, including Logistic Regression and Random Forest, to perform risk level classification. The Random Forest model further enables extraction of feature importance values.

To enhance interpretability, an explainability module generates human-readable reasoning based on prediction outcomes and contributing features. In parallel, a sentiment analysis component evaluates the emotional context of user queries using probabilistic classification techniques.

Finally, the system produces context-aware responses, recommendations, or alerts. Critical risk predictions are forwarded to the clinician interface for review and validation. All interactions and predictions are securely stored within the database layer.

12. Algorithms used

12.1. NLP Intent Classification

Intent classification is responsible for identifying the purpose of user queries, enabling the system to generate contextually relevant responses.

Technique Used: Term Frequency–Inverse Document Frequency (TF-IDF) for feature extraction combined with supervised classification models.

Algorithms Implemented:

- Support Vector Machine (SVM)
- Logistic Regression

Rationale:

TF-IDF effectively transforms textual input into numerical feature vectors, while SVM and Logistic Regression provide robust performance for multi-class text classification tasks. These models enable accurate identification of intents such as symptom queries, diet guidance, medication safety, and emotional concerns.

12.2. Symptom Risk Prediction

Risk prediction evaluates pregnancy-related symptoms and classifies potential severity levels.

Algorithms Implemented:

- Logistic Regression
- Random Forest

Rationale:
Logistic Regression provides a reliable baseline classifier with strong generalization capabilities. Random Forest, an ensemble-based learning model, captures non-linear relationships between health parameters and symptom combinations, leading to improved classification accuracy. Additionally, Random Forest enables extraction of feature importance, supporting interpretability.

Output Classes:

- Low Risk
- Moderate Risk
- High Risk

12.3. Sentiment Analysis

Sentiment analysis identifies emotional context within user interactions.

Algorithms Implemented:

- Naïve Bayes
- Support Vector Machine (SVM)

Rationale:
Probabilistic classification techniques efficiently model linguistic patterns associated with emotional expressions. The sentiment module detects states such as stress, anxiety, neutral, and positive tone, allowing adaptive response generation.

12.4. Explainability Mechanism

Explainability enhances system transparency by providing reasoning behind predictions.

Technique Used:

- Feature Importance Analysis
- Rule-Based Explanation Framework

Rationale:
Feature importance metrics derived from the Random Forest model identify dominant contributing factors. A rule-based explanation engine translates prediction logic into human-readable interpretations, improving user trust and clinician validation.

13. System Implementation

The PregBot system was implemented using a modular and layered development approach to ensure scalability, maintainability, and efficient integration of computational components. The implementation combines web technologies with Machine Learning and Natural Language Processing frameworks to deliver an interactive and intelligent decision support platform.

The frontend interface was developed using HTML5, CSS3, and JavaScript, providing a responsive and user-friendly interaction environment. The backend architecture was implemented using the Flask framework in Python, enabling secure routing, session management, and seamless communication between system modules.

The Natural Language Processing module was developed using NLTK and Scikit-learn libraries. User queries undergo preprocessing steps including tokenization, normalization, and stop-word removal. TF-IDF vectorization was applied for feature extraction, followed by supervised classification using Support Vector Machine (SVM) and Logistic Regression models for intent detection.

The Machine Learning risk prediction module was implemented using Scikit-learn. Structured feature vectors derived from user profiles and symptom attributes were processed using Logistic Regression and

Random Forest classifiers. The Random Forest model was further utilized to compute feature importance values supporting interpretability.

A sentiment analysis module was incorporated to evaluate emotional context within user interactions. Probabilistic classification techniques enabled identification of stress, anxiety, neutral, and positive emotional states.

To improve system transparency, an explainability module was implemented using feature importance mapping and rule-based reasoning. This component generates human-readable explanations for prediction outcomes.

The database layer was implemented using MySQL to store user profiles, symptom records, chatbot interactions, prediction results, and clinician feedback. Secure authentication mechanisms and role-based access control were integrated to differentiate patient and doctor functionalities.

Overall, the implementation ensures real-time system responsiveness, computational efficiency, and reliable interaction between conversational, predictive, and analytical components.

14. Testing and Validation

The PregBot system underwent comprehensive testing and validation to ensure functional correctness, predictive reliability, system performance, and interaction accuracy. A combination of functional, performance, security, and model validation techniques was employed.

14.1. Functional Testing

Functional testing was conducted to verify the correct operation of individual system modules. Each feature, including user authentication, conversational interaction, symptom logging, risk prediction, sentiment detection, and doctor review mechanisms, was evaluated using predefined test cases.

The testing confirmed that the system accurately processes user inputs, generates relevant responses, and performs risk classification without operational failures. All modules demonstrated consistent and expected behavior under normal usage scenarios.

14.2. Performance Testing

Performance evaluation focused on system responsiveness and computational efficiency. The chatbot response time, prediction latency, and overall application performance were measured under multiple interaction scenarios.

The results indicated:

- Average chatbot response time below 2 seconds
- Risk prediction execution time below 1 second
- Stable system behavior under concurrent requests

These outcomes demonstrate the suitability of the system for real-time interaction.

14.3. Security Testing

Security validation ensured protection of sensitive user data. The authentication mechanisms, session management, and access control policies were rigorously tested.

The system successfully implemented:

- Password hashing and secure credential storage
- Session timeout mechanisms
- Role-based access control

No critical vulnerabilities were observed during testing.

14.4. Machine Learning Model Validation

The predictive performance of Machine Learning models was evaluated using standard classification metrics.

Observed Results:

- Random Forest Accuracy: 93%
- Logistic Regression Accuracy: 87%
- Sentiment Classification Accuracy: ~85%

Precision and recall metrics indicated reliable prediction consistency. The Random Forest model demonstrated superior performance due to its ability to capture complex, non-linear symptom relationships.

14.5. Conversational Accuracy Evaluation

The NLP module was validated through intent recognition and entity extraction testing.

Observed Performance:

- Intent Classification Accuracy: ~90%
- Entity Recognition Accuracy: ~85%

The system effectively interpreted pregnancy-related queries and generated contextually relevant responses.

15. Results and Discussion

The evaluation of the PregBot system focuses on analyzing the performance of the conversational module, predictive risk classification models, sentiment analysis component, and overall system effectiveness. The results demonstrate the capability of the system to provide accurate, responsive, and reliable pregnancy assistance.

15.1. Conversational System Performance

The Natural Language Processing module was assessed using a diverse set of pregnancy-related queries covering symptoms, nutrition, medication safety, lifestyle guidance, and emotional concerns.

Observed Results:

- Intent Classification Accuracy: Approximately 90%
- Entity Recognition Accuracy: Approximately 85%
- Average Response Time: Below 2 seconds

Discussion:

The integration of TF-IDF vectorization with supervised classifiers enabled effective interpretation of user queries. The conversational system successfully handled structured and semi-structured inputs, generating contextually relevant responses. Minor limitations were observed in highly complex medical queries, which require clinician intervention.

15.2. Symptom Risk Prediction Performance

Two Machine Learning models were evaluated for pregnancy risk classification.

Model	Accuracy	Precision	Recall
Logistic Regression	87%	0.86	0.84
Random Forest	93%	0.92	0.91

Table: 1

Discussion:

Random Forest achieved superior performance due to its ensemble-based learning mechanism and ability to capture non-linear relationships among symptom features. Logistic Regression demonstrated reliable baseline performance but was comparatively less effective in modeling complex symptom interactions.

15.3. Sentiment Analysis Performance

The sentiment analysis module was evaluated to determine its effectiveness in detecting emotional context.

Observed Results:

- Sentiment Classification Accuracy: ~85%
- Effective detection of stress and anxiety indicators

Discussion:

The module successfully identified emotionally sensitive inputs, enabling adaptive response generation. Dataset balancing improved classification stability and reduced false positives.

15.4. Explainability Outcomes

The explainability mechanism generated interpretable reasoning for predictive outputs.

Observed Outcomes:

- All risk predictions accompanied by explanations
- Clear identification of contributing features

Discussion:

Explainability significantly enhanced system transparency and interpretability. This feature improves user trust and supports clinician validation.

15.5. Overall System Effectiveness

Performance Metric	Observed Result
Average Response Time	< 2 seconds
Risk Prediction Accuracy	93%
Conversational Accuracy	~90%
System Reliability	High

Table: 2

Discussion:

The modular architecture enabled stable system performance and efficient interaction between computational components. The system effectively fulfills its objective of providing intelligent and accessible pregnancy support.

Model	**Accuracy**	**Precision**	**Recall**
Logistic Regression	87%	0.86	0.84
Random Forest	93%	0.92	0.91

Table: 3

16. Conclusion

Pregnancy is a medically sensitive and dynamically evolving condition that requires continuous guidance, timely awareness, and reliable decision support. This study presented *PregBot*, an intelligent decision support framework developed using Machine Learning and Natural Language Processing techniques to enhance pregnancy care assistance.

The proposed system successfully integrates conversational interaction, symptom-aware risk classification, sentiment analysis, and prediction interpretability within a unified computational platform. The NLP module enables effective understanding of user queries through accurate intent recognition and entity extraction, facilitating natural and intuitive interaction. The Machine Learning-based prediction models demonstrated reliable performance in classifying symptom severity and estimating potential risk levels, with the Random Forest classifier achieving superior predictive accuracy.

The incorporation of sentiment analysis enhances contextual adaptability by recognizing emotional cues, while the explainability mechanism improves transparency by providing human-readable reasoning behind predictions. Additionally, the clinician interface ensures medical oversight and validation, strengthening system reliability.

Experimental evaluation confirmed that PregBot achieves accurate predictions, stable system responsiveness, effective conversational interpretation, and secure data handling. The results indicate that computational decision support systems can significantly improve accessibility, awareness, and user engagement in pregnancy-related healthcare contexts.

17. Future Scope

Although the PregBot system demonstrates reliable performance, several enhancements may further improve its capabilities:

- Integration with wearable health monitoring devices for real-time physiological data analysis
- Incorporation of voice-based interaction for improved accessibility
- Expansion of multilingual support for regional language adaptability
- Deployment as a mobile application to increase usability and reach
- Implementation of advanced predictive models using larger clinical datasets
- Extension of system functionality to postpartum and newborn care

These improvements may enhance prediction accuracy, accessibility, and real-world clinical applicability.

References

1. World Health Organization, *WHO Recommendations on Antenatal Care for a Positive Pregnancy Experience*, WHO Press, Geneva, Switzerland, 2016.
2. A. Rajkomar, J. Dean, and I. Kohane, “Machine Learning in Medicine,” *New England Journal of Medicine*, vol. 380, no. 14, pp. 1347–1358, 2019.
3. Y. Liu, J. Chen, and Z. Li, “Pregnancy Complication Risk Prediction Using Machine Learning Models,” *IEEE Access*, vol. 8, pp. 185521–185531, 2020.
4. R. Miner et al., “Smart Chatbots for Health Care,” *Journal of Medical Internet Research*, vol. 18, no. 8, pp. 1–11, 2016.
5. C. Cortes and V. Vapnik, “Support-Vector Networks,” *Machine Learning*, vol. 20, no. 3, pp. 273–297, 1995.
6. F. Pedregosa et al., “Scikit-learn: Machine Learning in Python,” *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.
7. M. T. Ribeiro, S. Singh, and C. Guestrin, “Why Should I Trust You? Explaining the Predictions of Any Classifier,” *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discov-*

ery and Data Mining, pp. 1135–1144, 2016.