AI for Maternal Health Risk Prediction: A Machine Learning Approach to Safer Motherhood

1. Project Scope

In Scope:

- Building a machine learning model to classify maternal health risk as low, medium, or high.
- Utilizing both physiological (e.g., age, blood pressure, sugar level) and psychosocial (e.g., stress level, nutrition, sleep) indicators.
- Training, validating, and interpreting ML models using public datasets (UCI, WHO, RMNCH+A, etc.).
- Designing a basic web/mobile interface for input and risk output.
- Evaluating models using metrics like accuracy, precision, recall, and SHAP explainability.
- Potential collaboration with local clinics/NGOs to validate the solution.

Out of Scope:

- Real-time data from IoT or wearable devices.
- Clinical-grade diagnosis or medical intervention tools.
- Longitudinal maternal tracking systems.
- Hardware deployment or offline servers.
- Government- or NGO-scale rollout in the current phase.

2. Literature Review

Introduction:

Maternal mortality remains a global challenge, with over 800 women dying daily from preventable pregnancy complications (WHO, 2023). In low-resource settings, delays in diagnosis and limited access to medical professionals heighten this risk. A literature review is crucial to understand how machine learning has been applied to maternal risk prediction and identify research gaps this project can address.

Organization:

This review is organized thematically, covering AI in maternal health, risk prediction methodologies, and the role of psychosocial indicators in clinical ML models.

Summary and Synthesis:

1. Smith et al. (2021) – “Machine Learning for Maternal Risk Prediction”
    a. Methodology: Logistic regression and Random Forest applied to clinical records.
    b. Key Findings: Model predicted preeclampsia with 87% accuracy using a small set of vital signs.
    c. Contribution: Demonstrated ML’s potential in low-resource diagnosis but lacked psychosocial data.
2. Kumar & Patel (2022) – “AI in Maternal Healthcare: A Review”
    a. Methodology: Comparative review of 18 AI-based maternal care models.
    b. Key Findings: Emphasized interpretability and localized data as crucial for adoption.
    c. Contribution: Advocated for integrating community-level features like stress and nutrition.

Conclusion:

This review highlights that existing studies support ML for maternal health but show a gap in integrating psychosocial features. Our project will contribute a novel, hybrid approach that merges physiological and behavioral data for a more holistic model. All sources cited will be properly referenced in the final submission.

3. Data Research

Introduction:

To train a meaningful AI model, access to reliable and relevant maternal health data is essential. This section describes the datasets to be used and outlines the rationale for each, including how they support our classification task.

Data Sources:

1. UCI Maternal Health Risk Dataset

1.1. Source: UCI Machine Learning Repository
1.2. Format: CSV, 1,000 samples
1.3. Content: Age, BP, heart rate, sugar, temperature, risk level
1.4. Use: Primary dataset for model prototyping

2. India's RMNCH+A Dataset

    2.1. Source: Government of India
    2.2. Format: CSV/XLS
    2.3. Content: Maternal and child health indicators with regional/socio-demographic variation
    2.4. Use: For testing model generalization and socio-demographic impact

3. Simulated Psychosocial Dataset (project-generated)

    3.1. Content: Stress level, sleep quality, nutrition score
    3.2. Format: CSV
    3.3. Use: For experimenting with expanded features until real data is available

4. MIMIC-III Filtered
   Use: Exploration for extended medical features if necessary

Data Analysis:

Initial preprocessing will include:

- Imputation for missing values
- Normalization/scaling of numeric features
- Encoding of categorical variables
- Handling class imbalance via SMOTE or weighted sampling
- Visualizations and descriptive stats will be used to explore correlations (e.g., stress vs. high-risk classification), and outlier detection will ensure data integrity.

Conclusion:

The combination of structured health records and simulated behavioral data will enable the creation of a more holistic maternal risk model. Data preprocessing and synthetic generation strategies will ensure model readiness even with small samples.

4. Technology Review

Introduction:

This section reviews machine learning tools and platforms for developing and deploying the risk prediction model. The chosen technologies must be interpretable, scalable, and accessible to non-technical users (e.g., healthcare workers).

Technology Overview:

Tool / Library	Purpose	Key Features
scikit-learn	Core ML algorithms	Fast prototyping, interpretable models
XGBoost	High-performance boosting model	Robust for imbalanced classification
SHAP / LIME	Model explainability	Transparent, trust-building for predictions
Flask / FastAPI	Backend API deployment	Lightweight and mobile-friendly
React	Frontend development (web/mobile)	Clean UI for input + output interaction

Relevance to Project:

scikit-learn and XGBoost offer classification models well-suited for tabular data.

Explainability tools ensure ethical use and transparency in healthcare.
Flask/FastAPI + React allow for rapid deployment as a simple web or mobile app.

Comparison:

Deep learning (e.g., TensorFlow) was excluded due to dataset size and the need for interpretability.
XGBoost outperforms Random Forest on small, imbalanced datasets.
FastAPI offers better async performance than Flask for future scaling.

Use Cases:

Johns Hopkins & PATH used interpretable models for rural maternal monitoring.
MediSafe uses a similar mobile-based approach for medication adherence reminders.

Gaps and Opportunities:

Most real-world apps lack psychosocial input—this project fills that gap.
Potential to expand into voice-based interaction (future phase).

Conclusion:

Our tech stack is chosen for ease of use, speed of development, and alignment with the needs of low-resource communities. The ML framework is lightweight and interpretable, making it ideal for maternal healthcare support tools.