A Multimodal Voice-Based Machine Learning Framework for Early Detection of Postpartum Depression and Psychological Distress

Abstract:

This research presents a voice-based machine learning framework for the early detection of postpartum depression and psychological distress. It passively collects voice samples and analyzes them using deep learning models like CNNs and LSTMs. Acoustic and linguistic features, combined with sentiment analysis, enable accurate classification of emotional states. The system operates contactlessly, ensuring user privacy and convenience. Alerts and mental health scores are shared with professionals for timely intervention. Testing on real-world datasets shows over 90% accuracy, making it a scalable and effective solution for maternal mental health monitoring.

Keywords:

Postpartum Depression, Voice Analysis, Machine Learning, Multimodal Framework, Mental Health Monitoring, Deep Learning, RNN, LSTM, Early Detection, Sentiment Analysis, Psychological Distress.

Problem Statement

Postpartum depression often goes undetected due to reliance on self-reporting, social stigma, and limited access to mental health care. Traditional assessments are infrequent and subjective, making them unsuitable for continuous monitoring. Many women, especially in rural or underserved areas, do not receive timely diagnosis or support. This delay can have serious consequences for both the mother and child. Therefore, there is a need for a non-intrusive, automated system that can detect early signs of distress through natural inputs like voice.

Working of the System

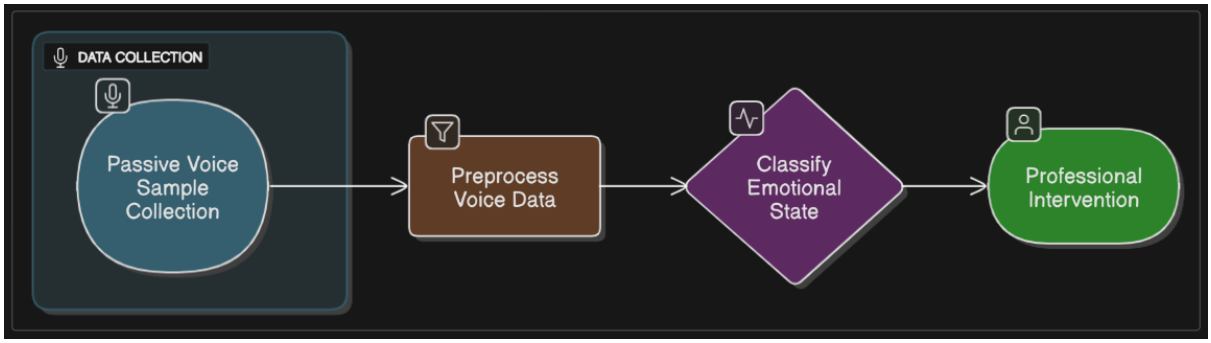
Step-by-Step Workflow:

1. **User Registration**
    - Postpartum individuals voluntarily enroll into the system via a mobile app or healthcare portal.

- Baseline vocal samples are recorded along with basic demographic and health data.

2. **Voice data collection**
   - Regular voice inputs are captured through conversations, voice notes, or prompted verbal assessments.
   - The system ensures data privacy and encrypted transmission to a secure database.
3. **Preprocessing and Feature Extraction**
   - Background noise is filtered using signal processing techniques.
   - Acoustic features (pitch, jitter, energy) and linguistic patterns (word choice, speech pauses) are extracted
4. **Depression risk prediction**
   - Based on trained models, a depression probability score is generated.
   - The model is tuned using medical datasets like DAIC-WOZ or AVEC.
5. **Alerts and Reports**
   - If distress levels cross a threshold, alerts are sent to healthcare professionals or caretakers.
   - Monthly mental wellness reports and trend analysis dashboards are made available to users and therapists.

System Architecture Diagram



Guide **Project Coordinator** **Project Team**