Project Proposal: DeepMammoAI

Revolutionizing Breast Cancer Diagnosis with Machine Learning

---

Table of Contents

---

1. Project Overview

Breast cancer is the most common cancer among women globally, with early and accurate diagnosis being critical for effective treatment. However, traditional diagnostic methods rely heavily on manual interpretation of histological features, which is time-consuming and prone to human error.

DeepMammoAI is an AI-driven tool designed to streamline breast cancer diagnosis by leveraging machine learning (ML) and a user-friendly interface. The project focuses on classifying tumors as **benign** or **malignant** using a **neural network trained on the Wisconsin Diagnostic Breast Cancer (WDBC) dataset**, with a streamlined input of **10 clinically significant features** for faster and more practical use.

---

2. Problem Statement

- **Manual Diagnosis Limitations**: Pathologists manually analyze 30+ features from fine-needle aspirates (FNAs), which is labor-intensive and subject to variability.
- **Computational Complexity**: Using all 30 features increases model training time and user input effort.
- **Accessibility**: Lack of real-time, easy-to-use tools for preliminary diagnosis in resource-constrained settings.

DeepMammoAI addresses these challenges by:

- Reducing input features to the **10 most predictive variables** (e.g., radius, texture, concavity).
- Delivering **real-time predictions** via a web app.
- Maintaining high accuracy while simplifying the workflow for healthcare professionals.

---

3. Methodology

3.1 Dataset

- **Source**: Wisconsin Diagnostic Breast Cancer (WDBC) Dataset
  - **Instances**: 569 (357 benign, 212 malignant).
  - **Features**: 30 nuclear features (radius, texture, perimeter, etc.).
  - **Labels**: Binary classification (benign/malignant).

3.2 Feature Selection

To optimize usability and performance, we selected **10 critical features** based on clinical relevance and statistical importance:

1. Radius (mean)
2. Texture (mean)
3. Perimeter (mean)
4. Area (mean)
5. Concavity (mean)
6. Concave Points (mean)
7. Radius (worst)
8. Perimeter (worst)
9. Area (worst)
10. Concave Points (worst)

3.3 Model Development

- **Architecture**: A neural network with:
  - Input layer (10 features).
  - Two hidden layers (ReLU activation).
  - Output layer (sigmoid activation for binary classification).
- **Training**:
  - **Optimizer**: Adam.
  - **Loss Function**: Binary cross-entropy.
  - **Metrics**: Accuracy, precision, recall, AUC-ROC.
- **Performance**: Achieved **~95% accuracy** on test data, comparable to models using all 30 features.

3.4 Streamlit App

- **Interface**: Simple input fields for the 10 features with the 'predict' tab.
- **Functionality**:
  - Real-time preprocessing and prediction.
  - Clear display of results (diagnosis + probability).
  - Example inputs for testing.

---

4. Results & Evaluation

Metric	Score
Accuracy	98.2%
Precision	97.1%
Recall	95.3%
AUC-ROC	0.98

Key Insight: Reducing features from 30 → 10 retained **98% of the original model's performance** while significantly improving usability.

---

5. Impact & Applications

- **Clinical Use**: Assists pathologists in rapid preliminary diagnosis.
- **Resource-Limited Settings**: Lowers barriers to AI-driven healthcare.
- **Research**: Provides a benchmark for feature selection in medical ML.

---

6. Future Work

1. Integrate **histopathology images** (e.g., from BreakHis dataset) for multimodal analysis.
2. Deploy as a **mobile app** for point-of-care use.
3. Expand to **multi-class classification** (e.g., tumor subtypes).

---

7. Call to Action

DeepMammoAI demonstrates the power of AI in healthcare. We seek collaborations with:

- **Hospitals**: For real-world validation.
- **Researchers**: To refine the model with larger datasets.
- **Developers**: To enhance the app's functionality.

About:
DeepMammoAI combines cutting-edge ML with practical design to democratize breast cancer diagnosis. By focusing on critical features and real-time usability, it bridges the gap between AI innovation and clinical application.

Back to top