**Project Review: Multi-modal AI System for Breast Cancer Risk Prediction**

**Slide 1: Title Slide**

* **Project Title:** Synergistic AI Diagnostics for Breast Cancer
* **Presented By:** [Your Names/Team Members]
* **Date:** [Current Date]
* **Under the Guidance:** Dr. Athota Kavitha, Associate Professor, CSE Department, JNTUH

**Slide 2: Project Overview (From Abstract)**

* **Goal:** Develop a robust diagnostic approach integrating heterogeneous data sources for personalized and timely breast cancer insights.
* **Proposed Framework:** Novel, multi-paradigm framework combining:
  + Ensemble Learning
  + Deep Convolutional Neural Networks (CNNs) with Transfer Learning
  + Reinforcement Learning (RL)
  + Graph-based methods
* **Data Modalities:**
  + Multi-view imaging data (craniocaudal and mediolateral mammograms)
  + Clinical information
  + Genomic data (explicitly leveraging TCGA-BRCA, GEO, METABRIC)
* **Key Techniques from Abstract:**
  + Graph Neural Networks (GNNs) for modeling interrelationships.
  + RL (Stable Baselines3) to dynamically optimize hyperparameters and decision thresholds.
* **Tools Mentioned in Abstract:** 3D Slicer, ITK-SNAP, SimpleITK, PyTorch, MONAI, Stable Baselines3, PyTorch Geometric, NetworkX, Bioconductor (R), Scanpy (Python).

**Slide 3: Progress Made (Based on Research Report)**

* **System Architecture:**
  + Implemented client-server architecture with Python Flask backend, web frontend (HTML, CSS, JS), and SQLite database.
  + Environment management using .env.
* **Data Integration:**
  + Framework for integrating Patient Questionnaire Data, Mammographic Image Data, and Gene Expression Data is established.
* **Image Analysis Implementation:**
  + Preprocessing using OpenCV and Scikit-image (Resizing, AHE, Inversion, Normalization).
  + Implementation of Traditional Feature-Based Models (HOG, LBP, SIFT-BoVW) using Scikit-image.
  + Implementation of Deep Learning Models (ResNet) using TensorFlow/Keras.
  + Ensemble methodology for combining image model predictions.
* **Gene Expression Analysis Implementation:**
  + Preprocessing using Pandas and NumPy (Transposition, Cleaning, Scaling).
  + Implementation of Logistic Regression model using Scikit-learn.
  + Persistence of model, scaler, and feature names.
* **Reinforcement Learning Implementation (Revised Scope):**
  + Implemented RL for **automated hyperparameter tuning** of individual gene expression and image models using **Stable Baselines3 and Gymnasium**.
  + Custom Gymnasium environments (GeneModelTuningEnv, ImageModelTuningEnv) created.
  + PPO agent from Stable Baselines3 used for tuning.
  + Capability to save best-performing tuned models.
* **Automated Medical Report Generation:**
  + Integration of Llama 3 via Groq API using llm\_utils.py.
  + Programmatic prompt construction synthesizing multi-modal data.
  + Report formatting into HTML.
* **Data Management:**
  + SQLite database (bcrrp\_data.db) for persistent storage.
  + File system storage for patient-specific files.
* **System Workflow:**
  + Implemented workflow for new patient assessment (questionnaire, image prediction, gene prediction, report generation).
  + Implemented patient records management (view all patients, view patient details).

**Slide 4: Features Missing / Remaining Work**

* **Features from Abstract NOT fully implemented/detailed in Report:**
  + Analysis of **Multi-View Imaging Data** (specifically craniocaudal and mediolateral mammograms mentioned in abstract). The report focuses on processing individual images.
  + Explicit leveraging of **TCGA-BRCA and METABRIC** public gene expression datasets (report mentions GEO and potential Kaggle, but not these specifically).
  + Use of **Graph Neural Networks (GNNs)** to model latent interrelationships.
  + Specific Image Processing Tools: **3D Slicer, ITK-SNAP, and SimpleITK** (report uses OpenCV and Scikit-image).
  + Deep Learning Framework: **PyTorch with MONAI** (report uses TensorFlow/Keras).
  + Graph Analysis Libraries: **PyTorch Geometric alongside NetworkX**.
  + Specific Bioinformatics Tools: **Bioconductor (R) and Scanpy (Python)** (report uses Pandas, NumPy, Scikit-learn).
  + RL for **optimizing image model ensemble predictions** (The implemented RL is for *hyperparameter tuning* of *individual* models, a shift from the abstract's description).
* **Limitations & Future Work from Research Report:**
  + Rigorous Clinical Validation with large, real-world datasets.
  + Refinement of Gene Expression Model with clinically relevant labels and validated signatures.
  + Ensuring Generalizability to different populations/equipment.
  + Integration of Explainable AI (XAI) techniques.
  + Enhancements for Security and Privacy (for clinical deployment).
  + Scalability improvements (migrating database, backend optimization).
  + Full integration of RL-tuned models into the main inference pipeline.
  + Expanding RL tuning scope and exploring advanced strategies.
  + Addressing LLM challenges (accuracy, bias, privacy).
  + Longitudinal Patient Monitoring.
  + EHR Integration.
  + Comparative Performance Analysis against existing tools.
  + Prospective Validation.
  + Exploration of Federated Learning.

**Slide 5: Main Features of the Prototype**

* **User Interface:** Web-based interface for patient data input and report viewing.
* **Multi-modal Data Input:** Accepts patient questionnaire data, mammogram images, and gene expression files.
* **Automated Analysis Pipelines:** Processes image and gene expression data using trained ML models.
* **Image Analysis Ensemble:** Combines predictions from multiple image models for a potentially more robust result.
* **RL-Enhanced Models (Potential):** Capability to use models tuned via Reinforcement Learning for improved individual performance.
* **Automated Medical Report Generation:** Generates comprehensive narrative reports summarizing findings using an LLM.
* **Patient Record Management:** Stores and allows retrieval of patient data, predictions, and reports.
* **Modular Design:** Code organized into distinct modules for maintainability and future expansion.

**Slide 6: Prototype Visuals**

* *(This is where you will insert screenshots of your working prototype)*
* **Screenshot 1:** [Description: e.g., Homepage/Questionnaire Input]
* **Screenshot 2:** [Description: e.g., Image Upload Interface]
* **Screenshot 3:** [Description: e.g., Gene Data Upload Interface]
* **Screenshot 4:** [Description: e.g., Example Generated Medical Report]
* **Screenshot 5:** [Description: e.g., Patient Records View]

**Slide 7: Conclusion**

* Successfully developed a multi-modal Breast Cancer Risk Prediction Tool prototype.
* Integrated diverse data types and advanced ML techniques.
* Explored novel RL-based hyperparameter tuning.
* Implemented automated LLM-powered reporting.
* System provides a solid foundation for a comprehensive decision support tool.
* Significant progress made towards the project goals outlined in the abstract, with a clear roadmap for future enhancements and validation.

**Slide 8: Q&A**

* Open for questions and discussion.