

# BREAST CANCER DETECTION USING MACHINE LEARNING

UG Group-16

Aadi, Andre, Gaurav, Kevin, Rhea





## **CBIS-DDSM BREAST CANCER DATASET**

### **What it includes**

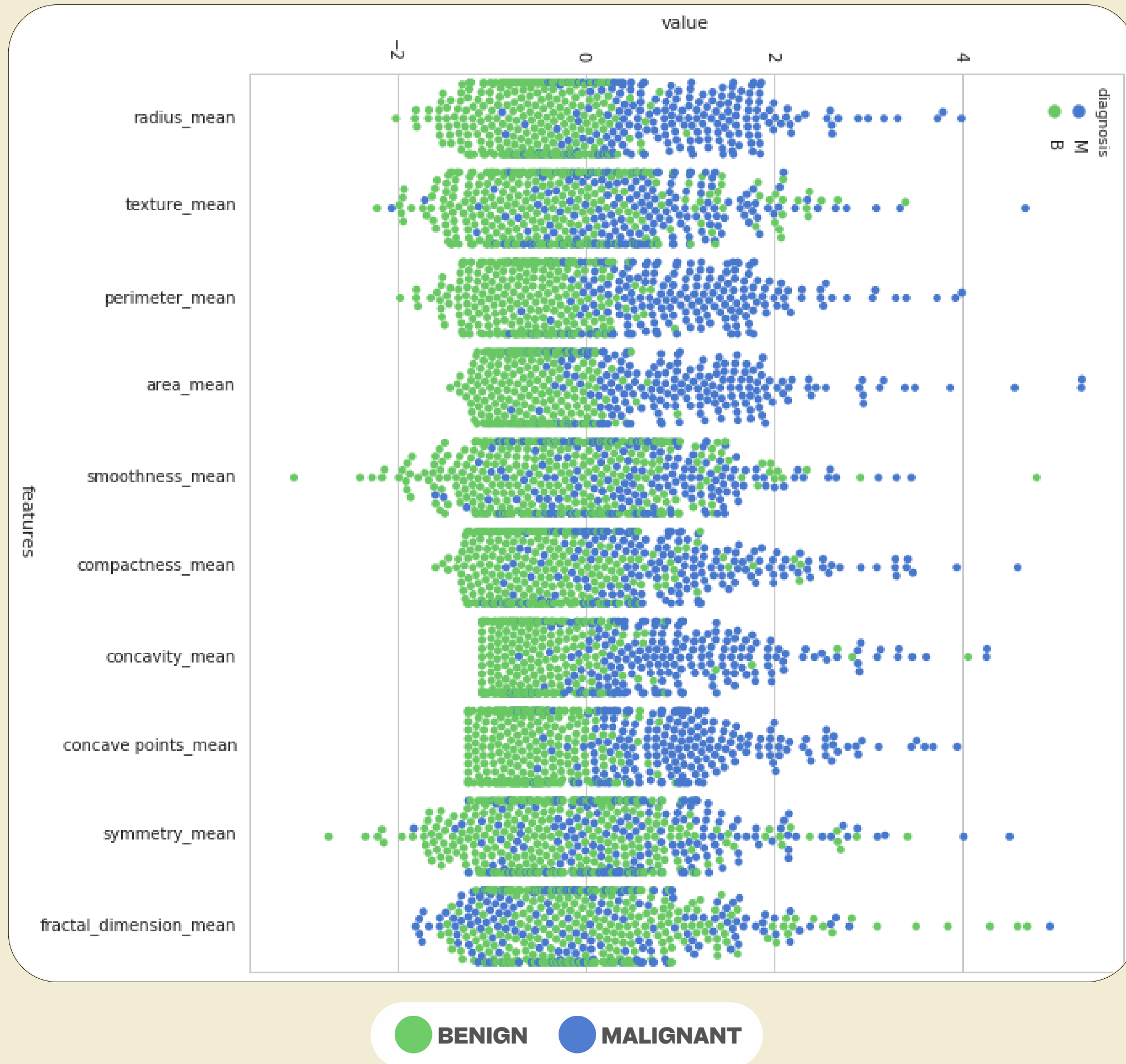
- It contains normal, benign, and malignant cases with verified pathology information.
- 2,620 scanned film mammography studies

## **BREAST CANCER WISCONSIN (DIAGNOSTIC) DATA SET**

### **What it includes**

- It contains benign and malignant cases
- 569 patient samples
- It has 30 numeric features such as radius, texture, smoothness, symmetry (with mean, standard error, and worst values computed for each)

## VISUALIZATION



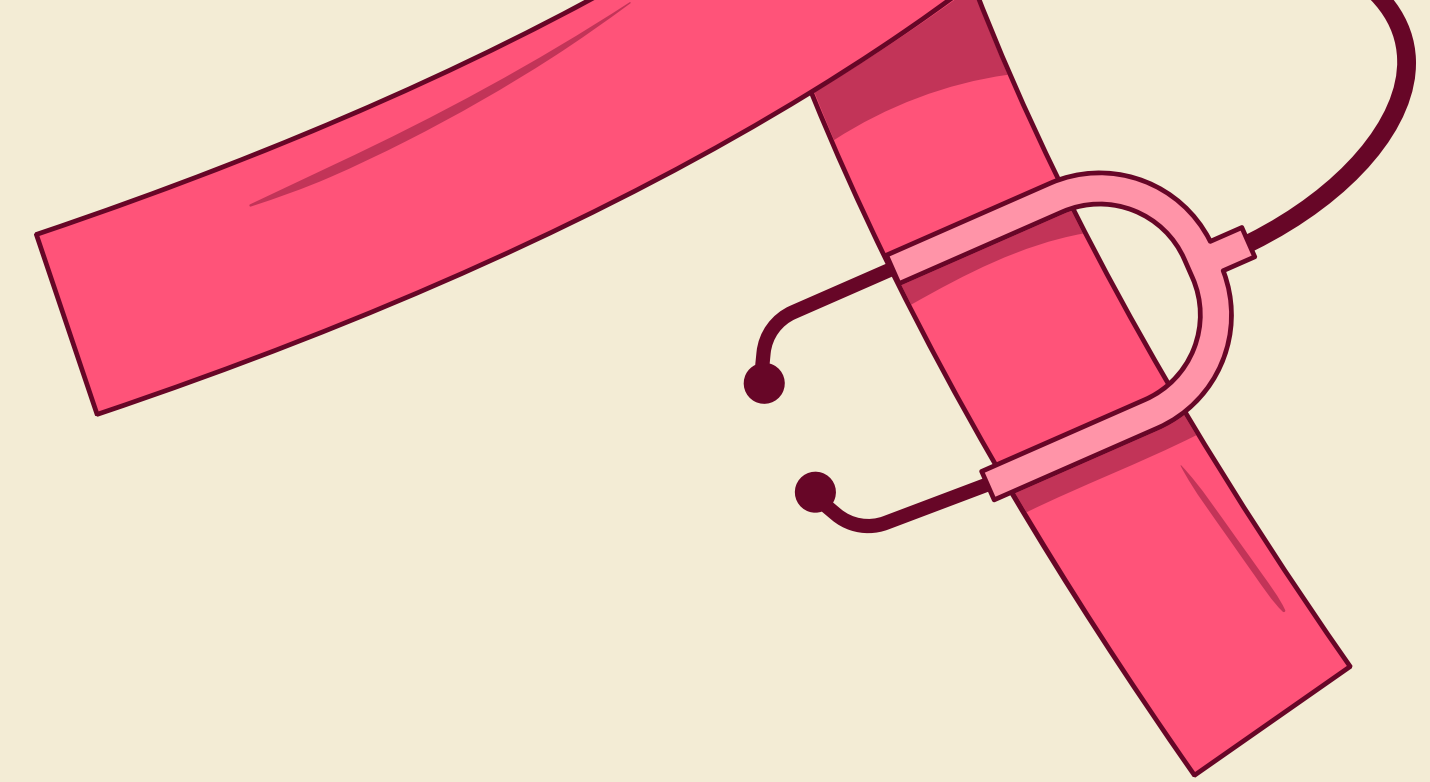
## WHY DID WE CHOOSE THESE DATASETS ?

**BREAST CANCER WISCONSIN-TABULAR** &  
**CBIS-DDSM: BREAST CANCER IMAGE DATASET**

- Structured dataset with 30+ features
- ideal for statistical analysis
- Suitable for deep learning & CNN-based approaches.
- Image dataset Includes multiple classes (benign, malignant, and normal).



# ML CONTRIBUTION



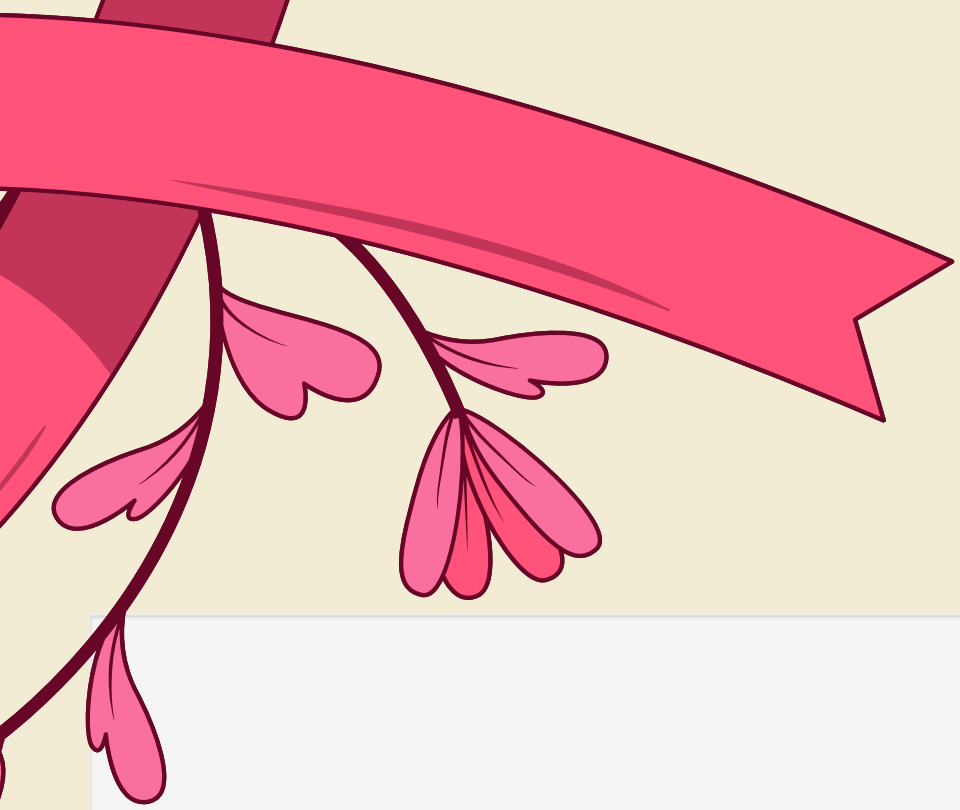
- **Wisconsin Breast Cancer (Diagnostic)**

- Benchmark dataset in ML — binary classification
- Used for algorithm comparison, feature selection,
- Accuracy often >95%

- **CBIS-DDSM (Curated Breast Imaging Subset of DDSM)**

- Cleaned & standardized version of DDSM
- Used for CNN classification, lesion detection & segmentation
- Benchmark for medical AI research



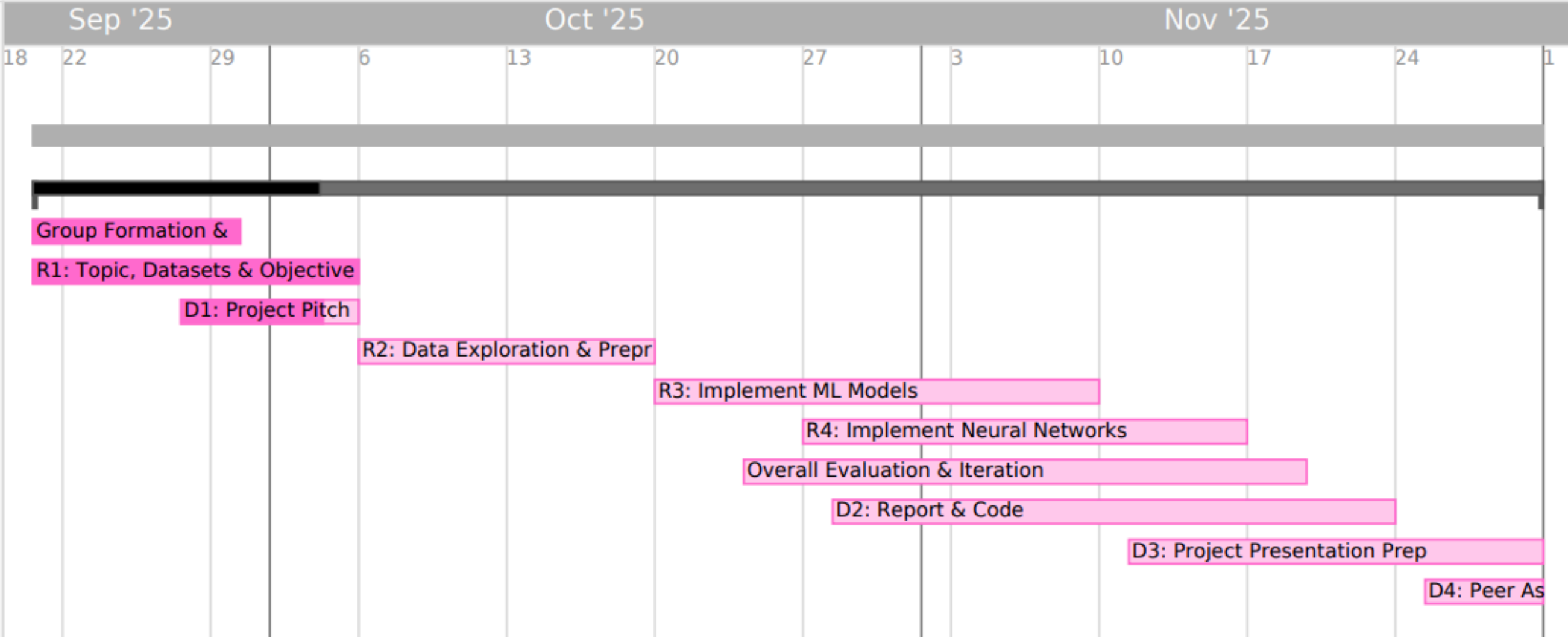


# PROJECT TIMELINE

## DMML Group Coursework

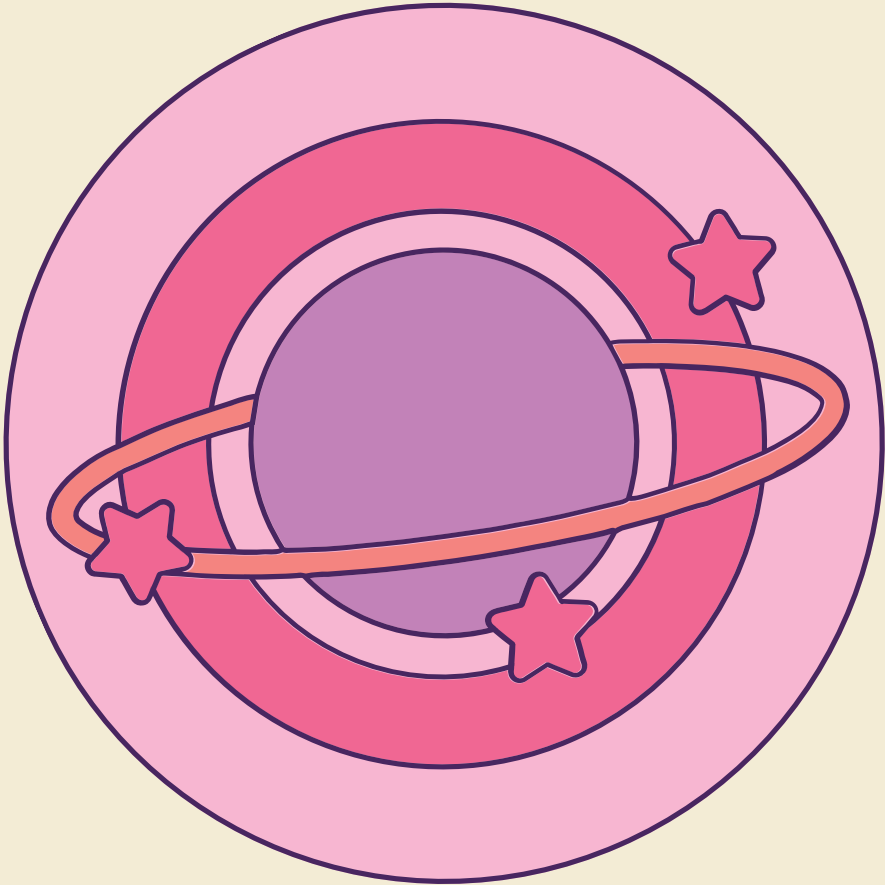
### Coursework Progress

- Group Formation & Setup
- R1: Topic, Datasets & Objectives
- D1: Project Pitch Prep
- R2: Data Exploration & Preprocessing
- R3: Implement ML Models
- R4: Implement Neural Networks
- Overall Evaluation & Iteration
- D2: Report & Code
- D3: Project Presentation Prep
- D4: Peer Assessment



ID	REQUIREMENTS	PRIORITY
R-1	Define a set of project objectives based on the selected topic and datasets.	M
R-2	Perform comprehensive data preprocessing, cleaning, and exploratory data analysis (EDA) with visualizations.	M
R-3	Apply machine learning algorithms for predictive modeling, evaluate them with appropriate metrics.	M
R-3	Implement, train, and fine-tune neural network models and assess their performance.	M
D-1	Deliver a project pitch in Week 4 (title, datasets, objectives).	M
D-2	Submit a 6-page PDF report and a zipped folder of the project code in Week 11.	M
D-3	Participate in a 15-minute mini-viva/presentation in Week 12.	M
D-4	Complete the peer-assessment form in Week 12.	M

# MoSCoW ANALYSIS



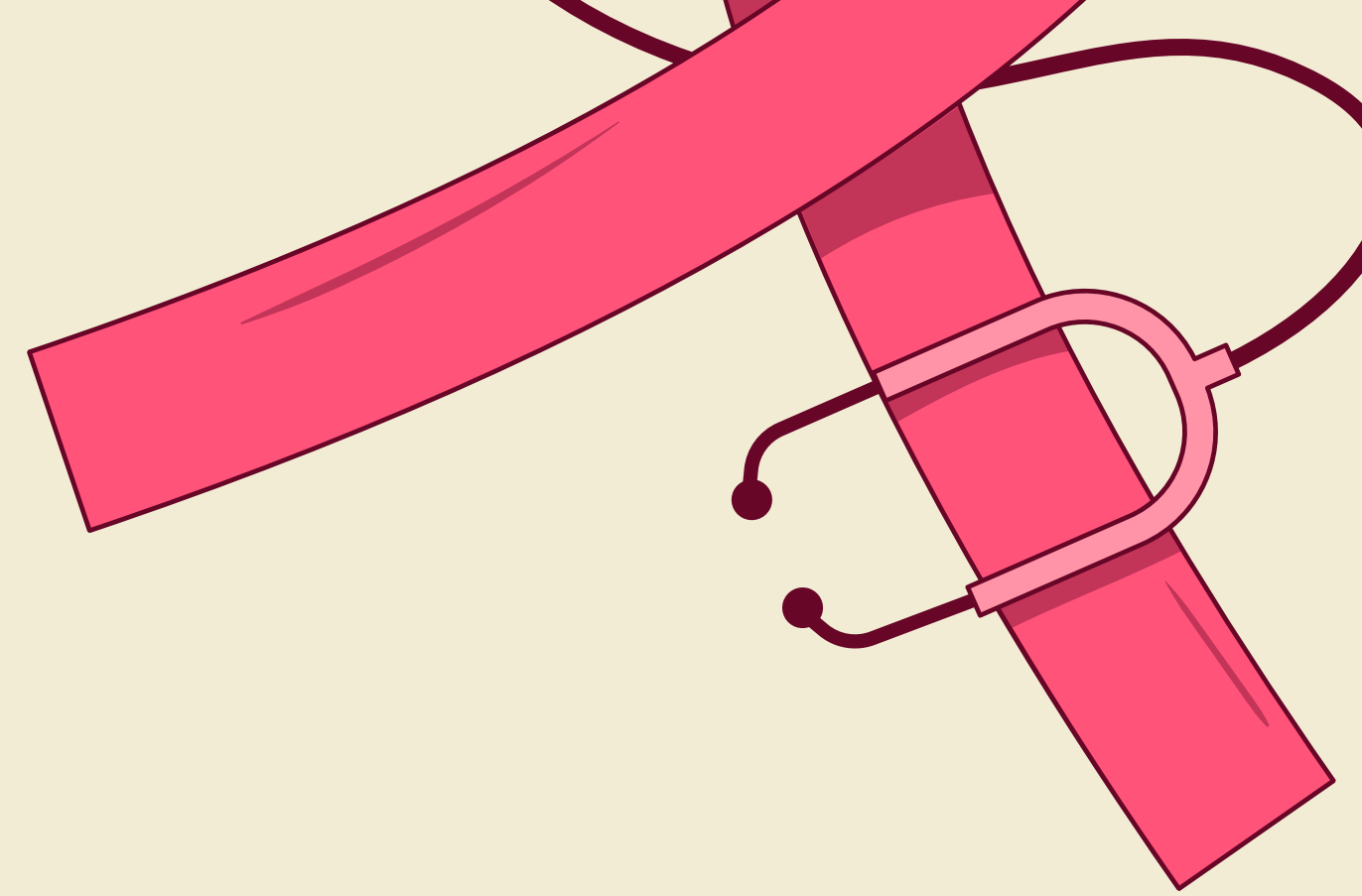
# RESEARCH QUESTIONS



- Can multimodal learning improve breast cancer detection accuracy?
- Which tabular features and image patterns are most important for early diagnosis?



# D1: FEEDBACK RECEIVED



- Ensure the dataset is properly split into training and testing sets before model training to avoid data leakage.
- Be cautious of overfitting by using techniques like regularization or cross-validation to improve model generalization.
- Ensure the dataset contains an equal or nearly equal number of malignant and benign samples to avoid class imbalance and biased model performance.



# GROUP CONTRIBUTION



## **Aadi-**

- Introduction and MOSCOW Analysis.

## **Gaurav-**

- Data Visualization and Dataset Justification.

## **Rhea-**

- Describing the Dataset.

## **Andre-**

- Dataset Contribution in ML Community.

## **Kevin-**

- Gantt Chart and Project planning.