Digital Egypt Pioneers

وزارة الاتصـــــــالات
وتكنولوجيا المعلومات

Breast Cancer Detection Using Deep Learning:

Model Development and Validation Using Kaggle's RSNA Dataset

Project Team Members:

Mohamed ElSayed AbdelAziz AbdelAty
Ahmed Jaber Abdelaziz
Omnia Mohamed Kamal
Tasneem Hussien Aboelsoud
Gamal Ehab Gamal Gerges

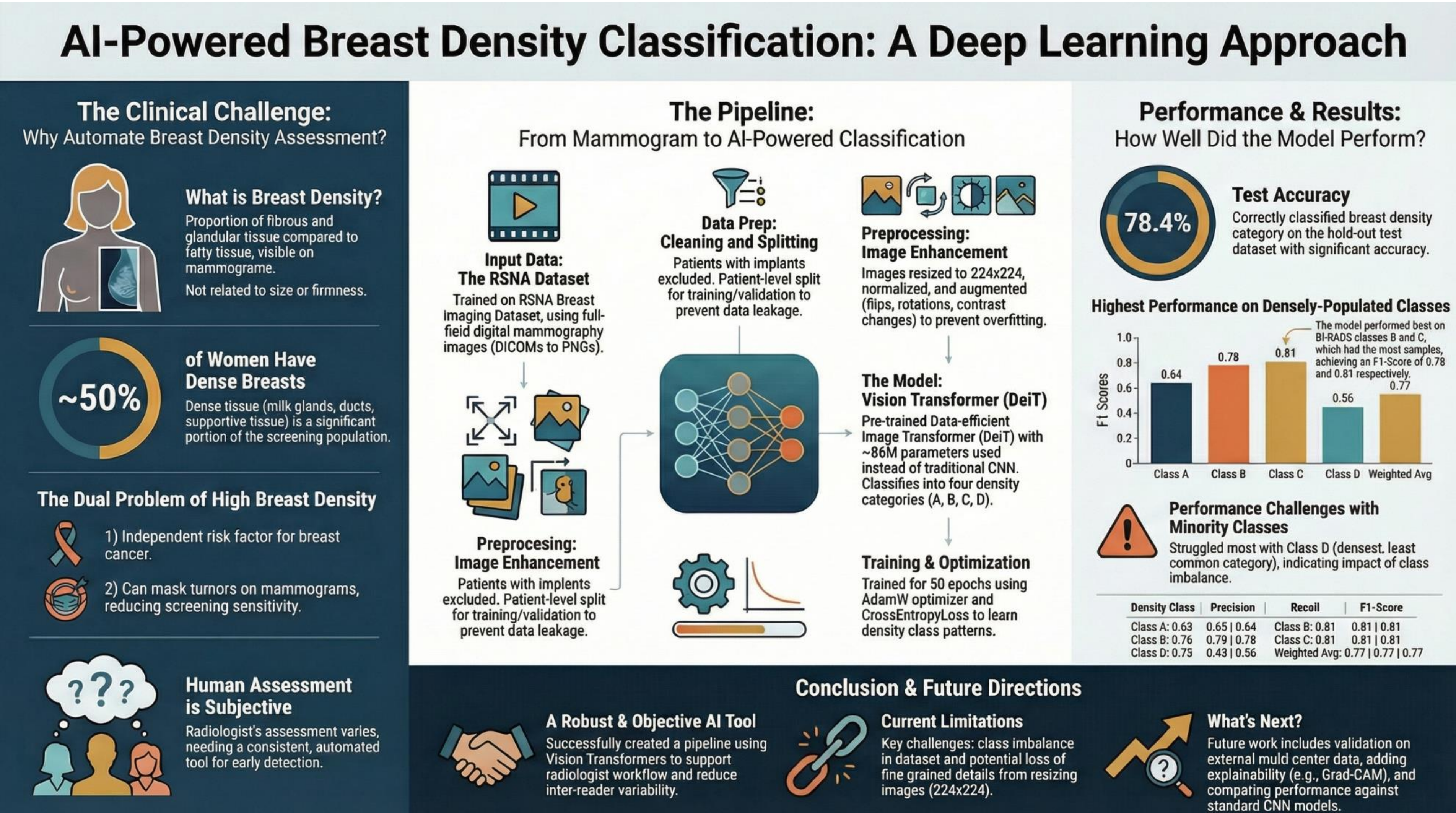
Project Advisor:

Eng. Sherif Salem

2025

Contents

Graphical Executive Summary

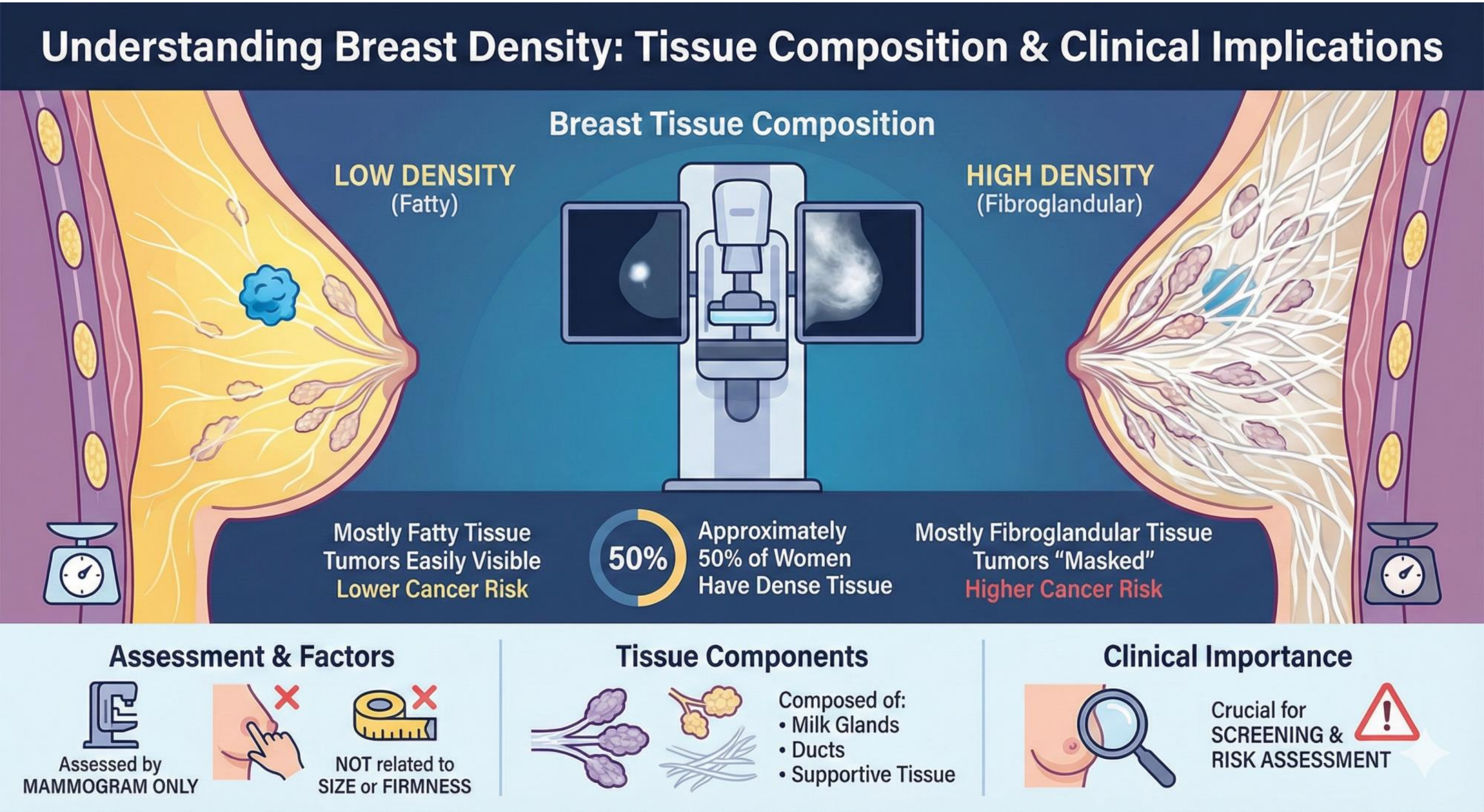


Digital Egypt Pioneers

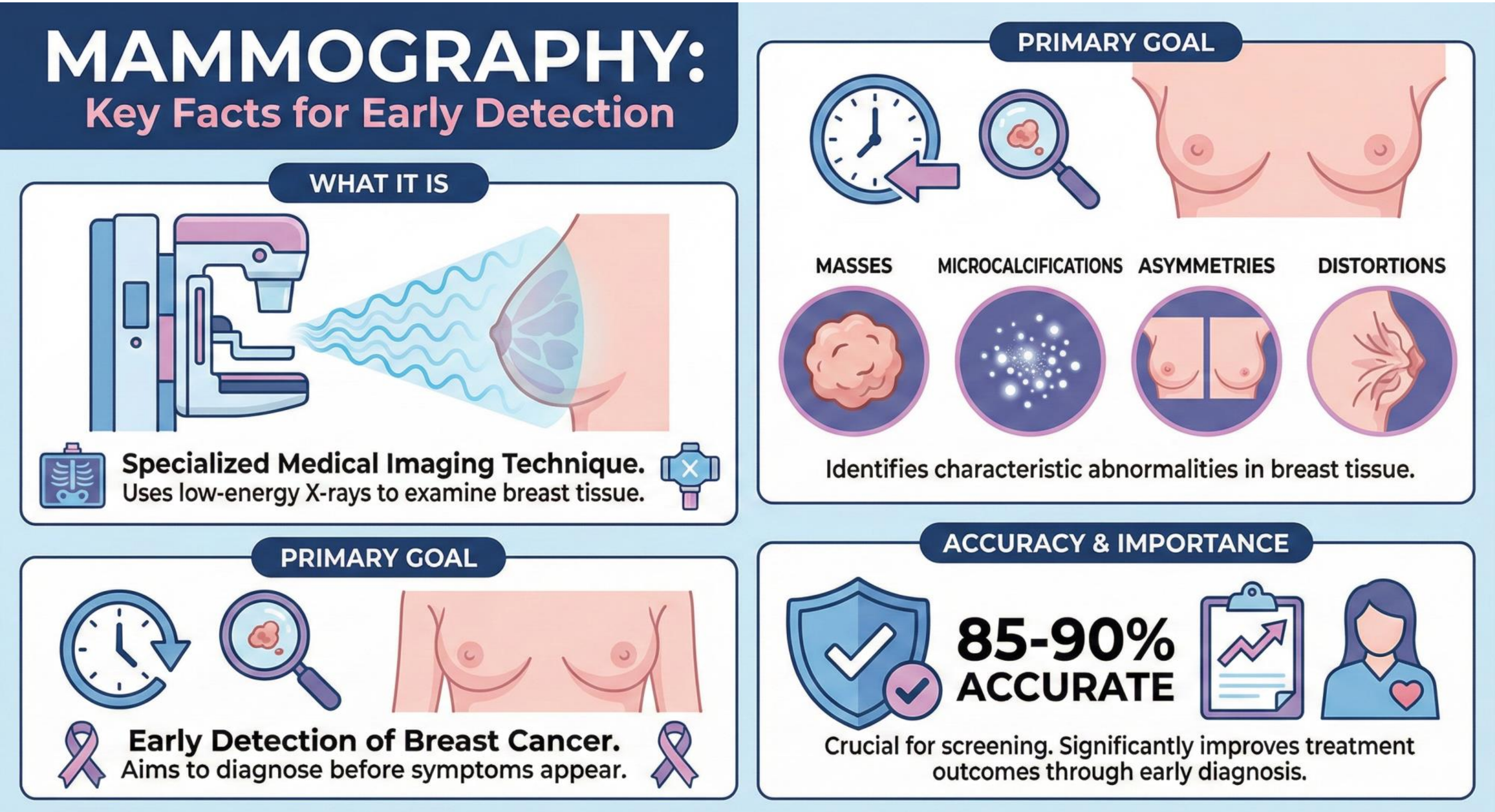
وزارة الاتصالات
وتكنولوجيا المعلومات

Background & Clinical Importance

Background & Clinical Importance



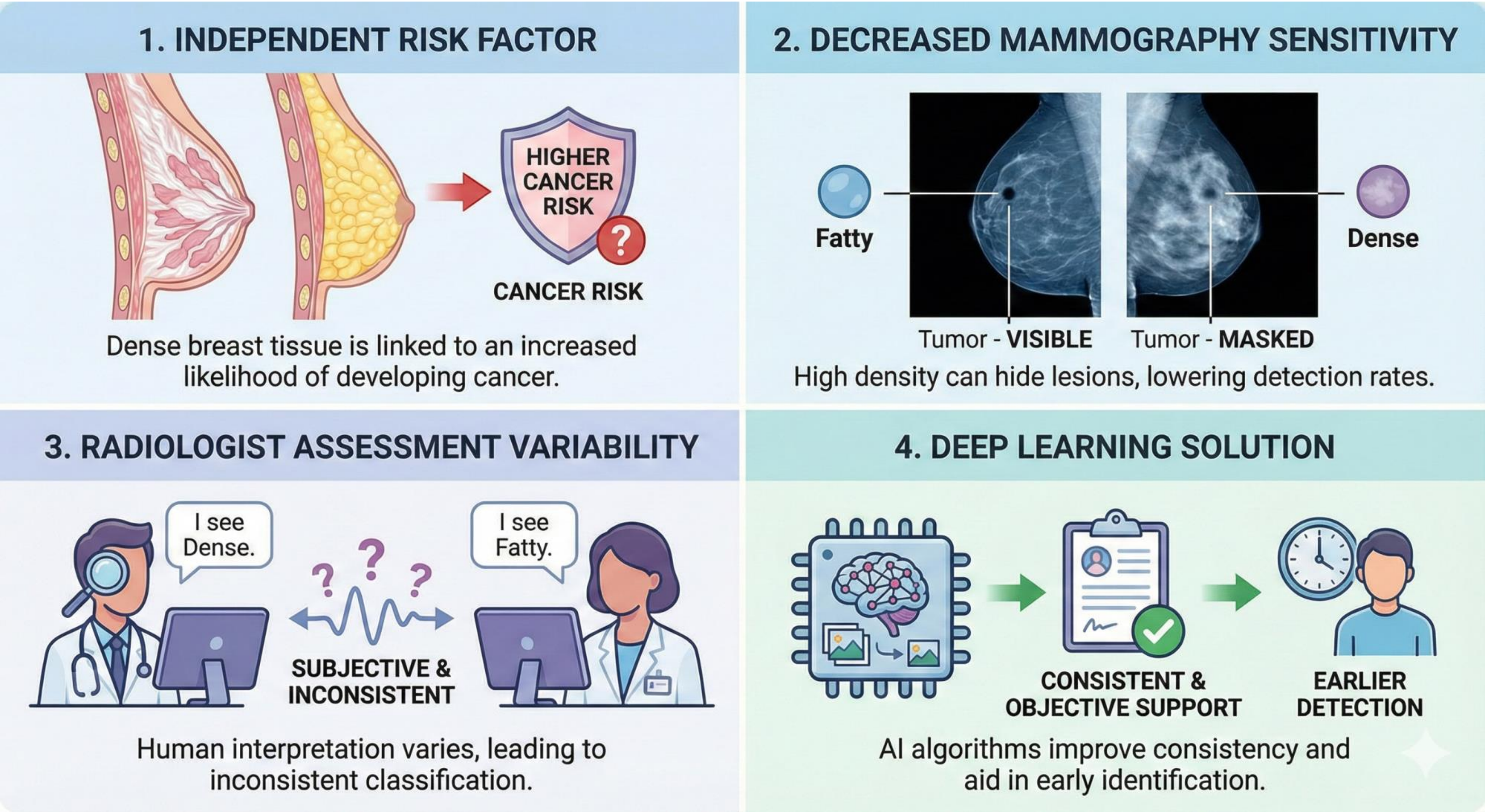
Background & Clinical Importance



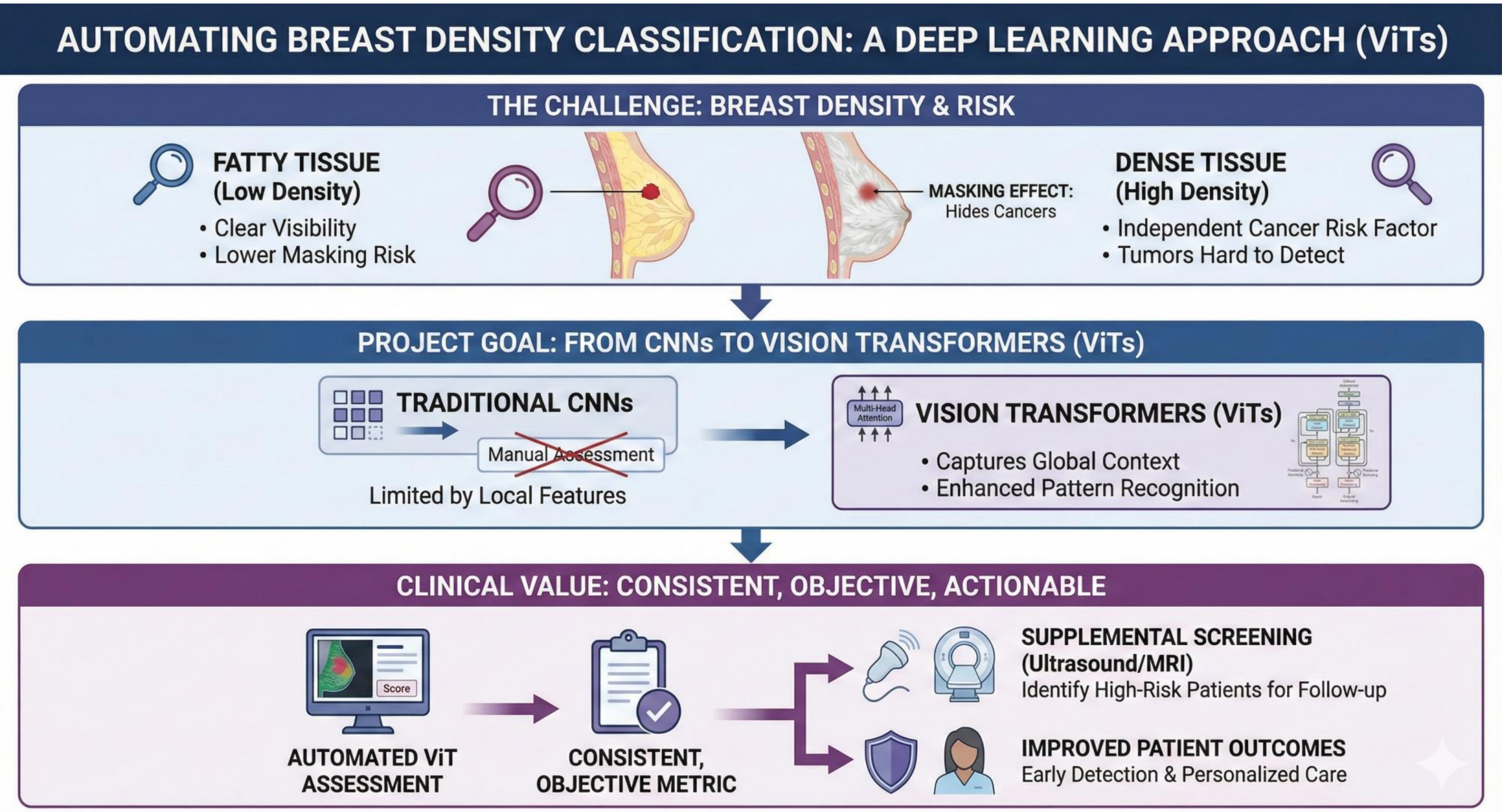
Digital Egypt Pioneers

وزارة الاتصالات وتكنولوجيا المعلومات

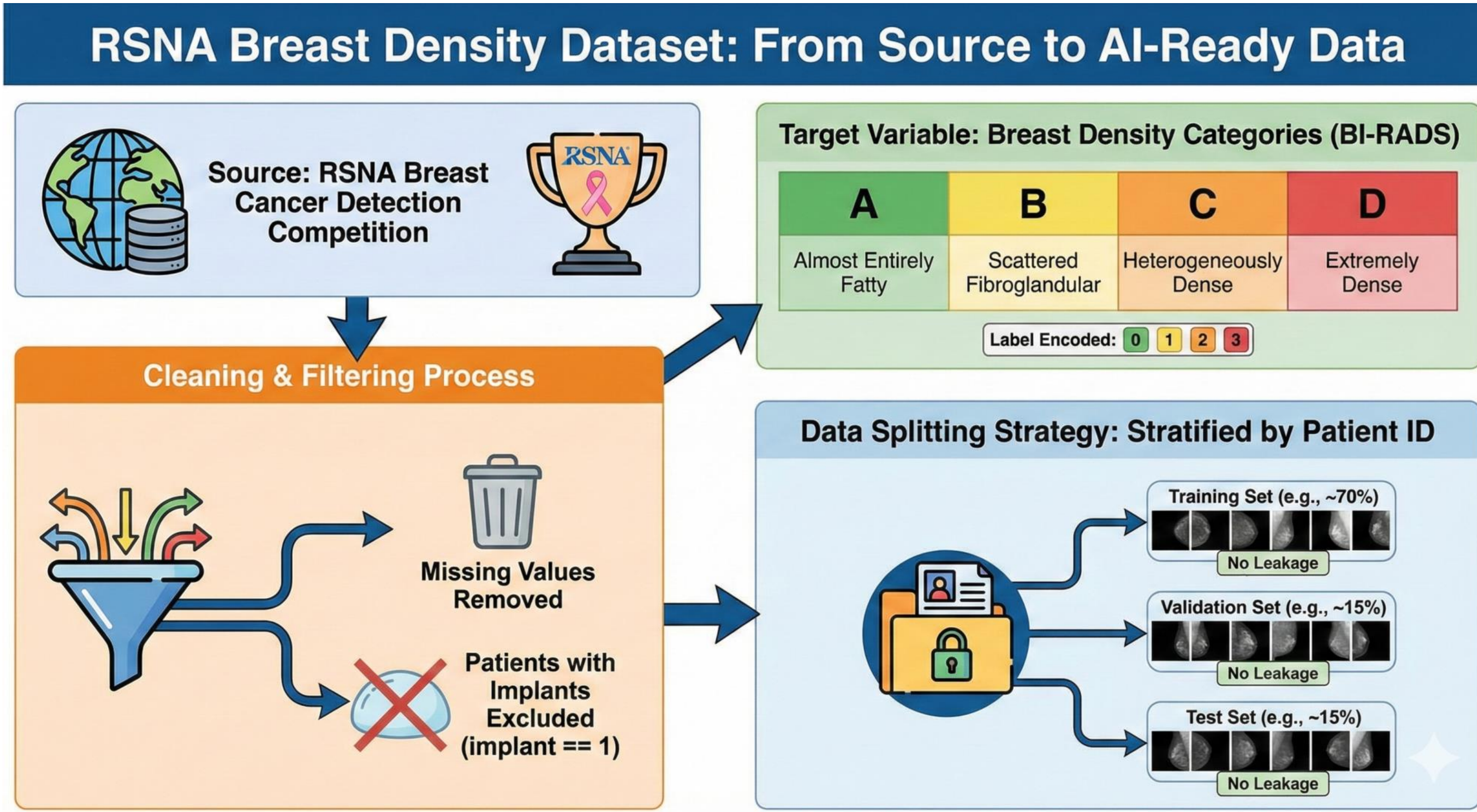
Background & Clinical Importance



Background & Clinical Importance

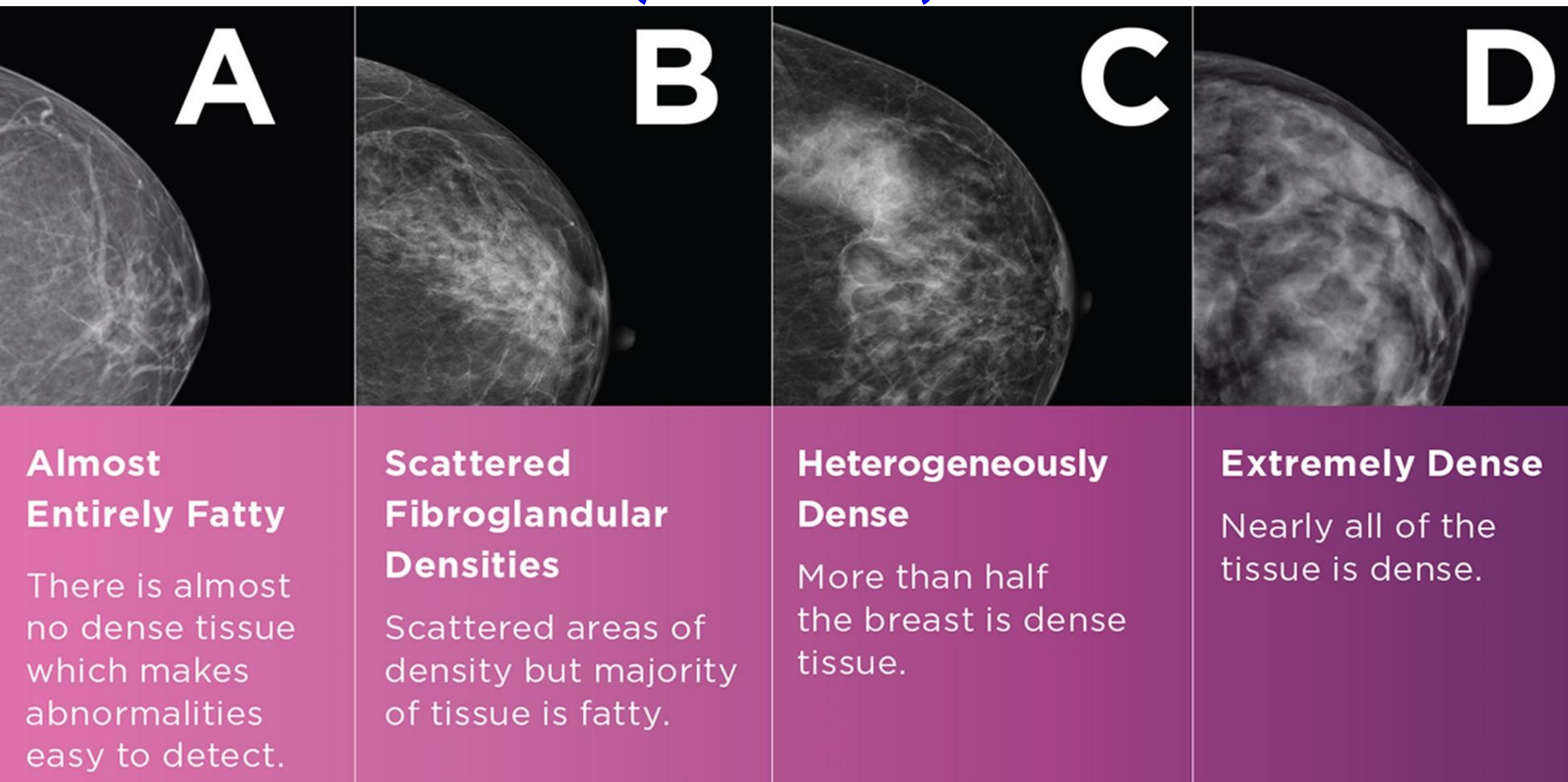


Dataset Description (RSNA)

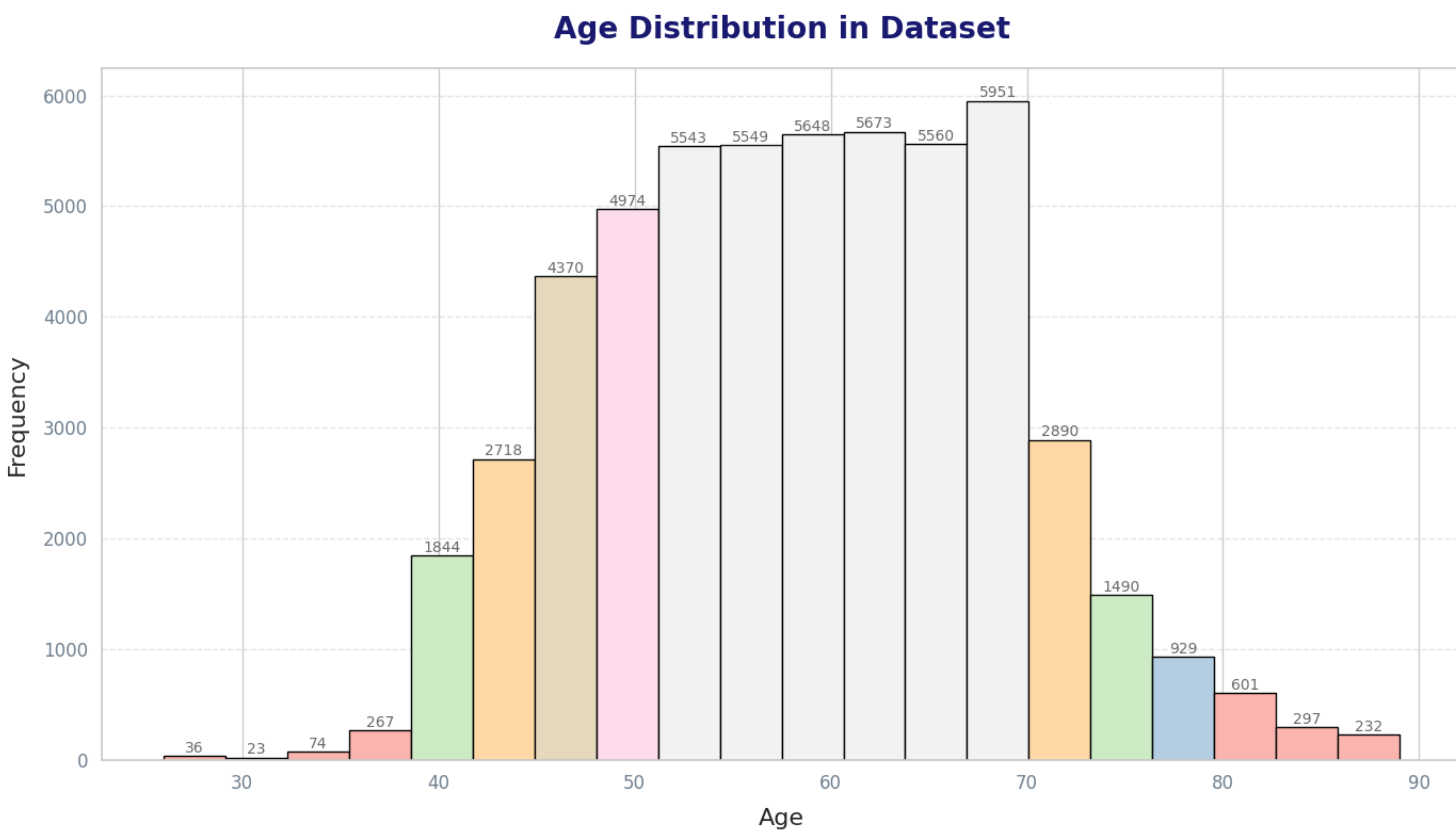


Dataset Description (RSNA)

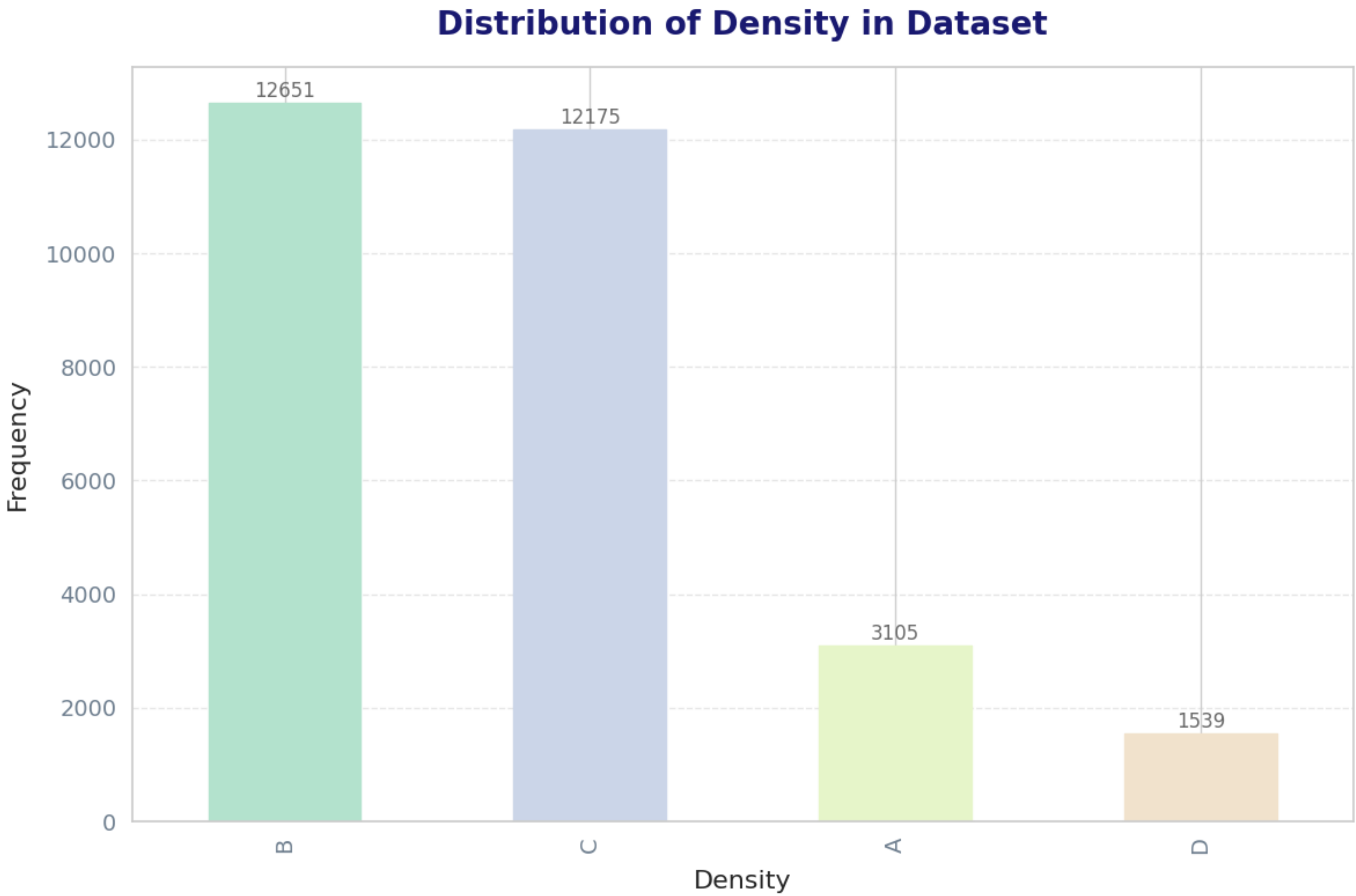
Breast Imaging Reporting and Data System (BI-RADS)



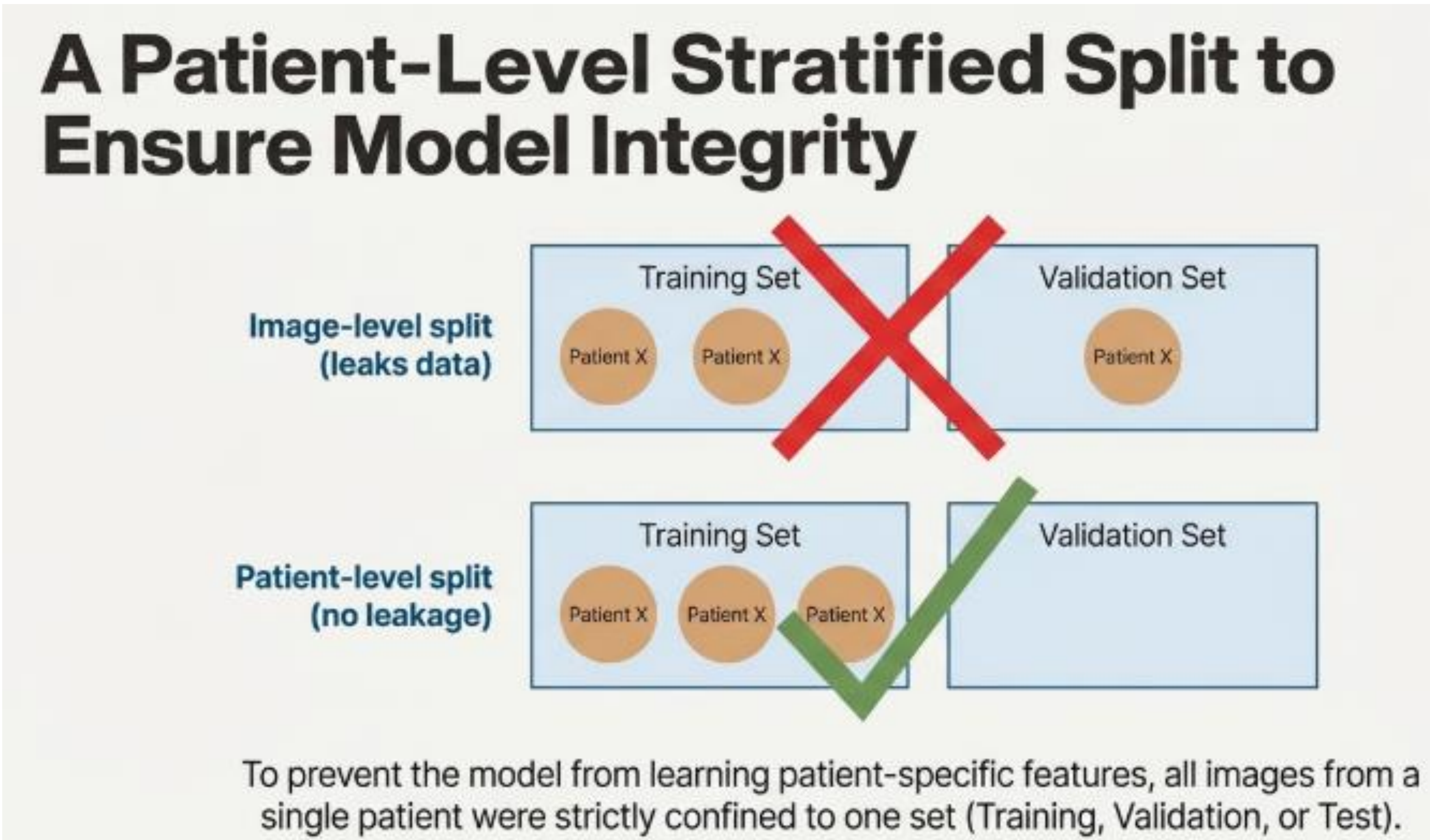
Dataset Description (RSNA)



Dataset Description (RSNA)

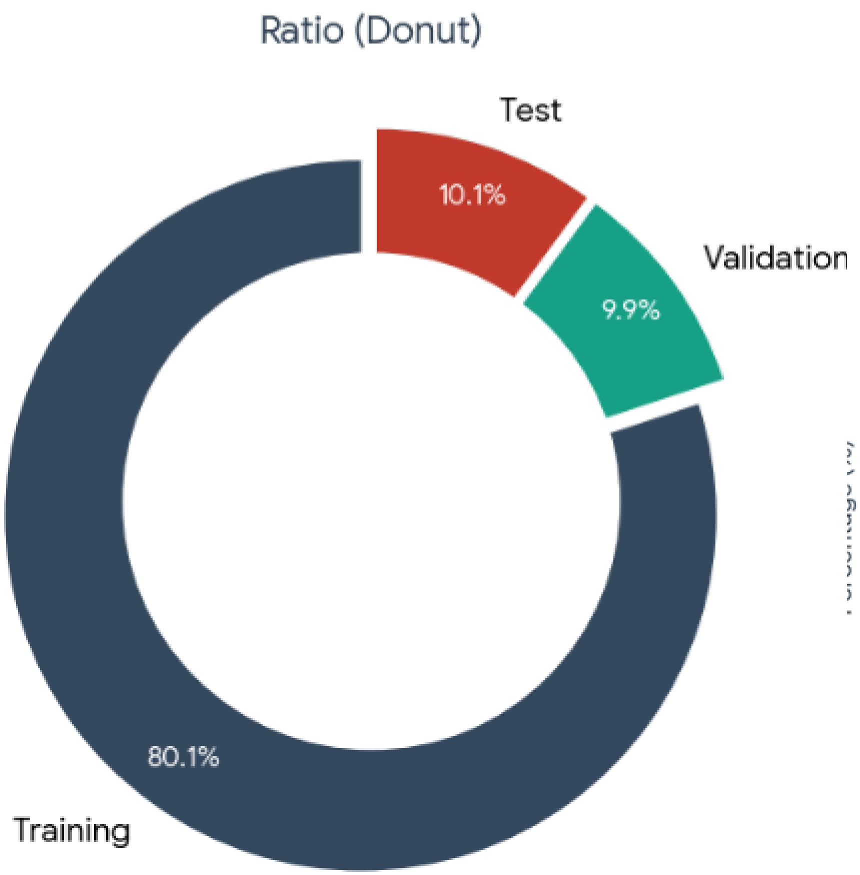


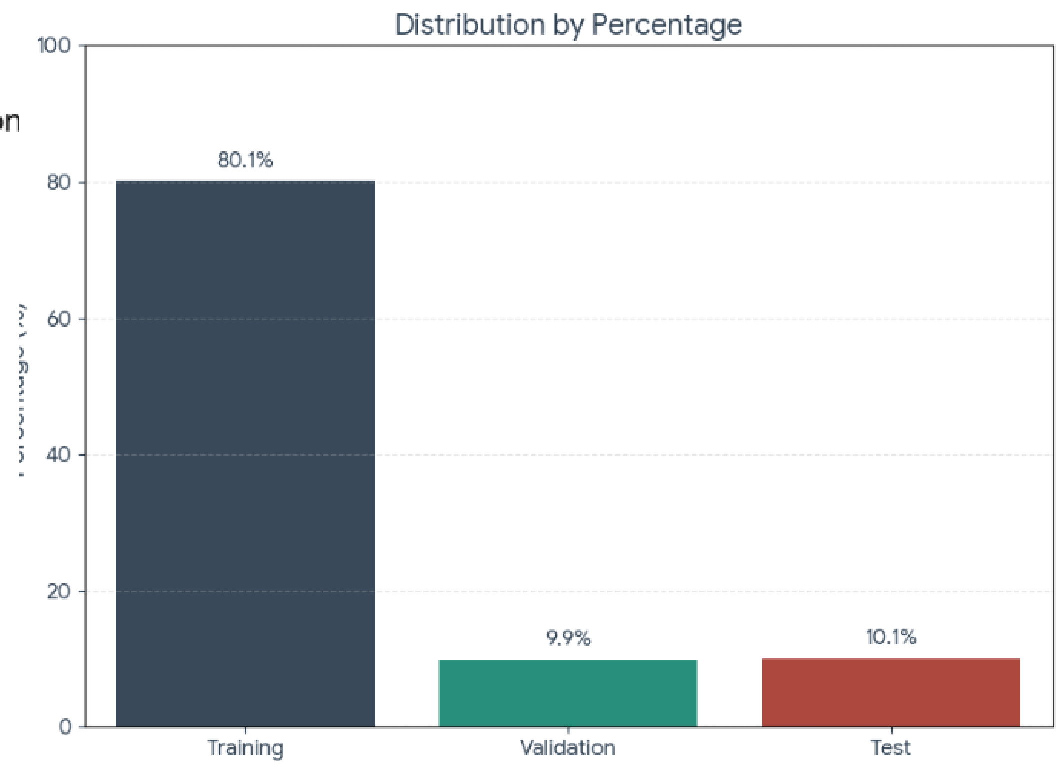
Dataset Description (RSNA)

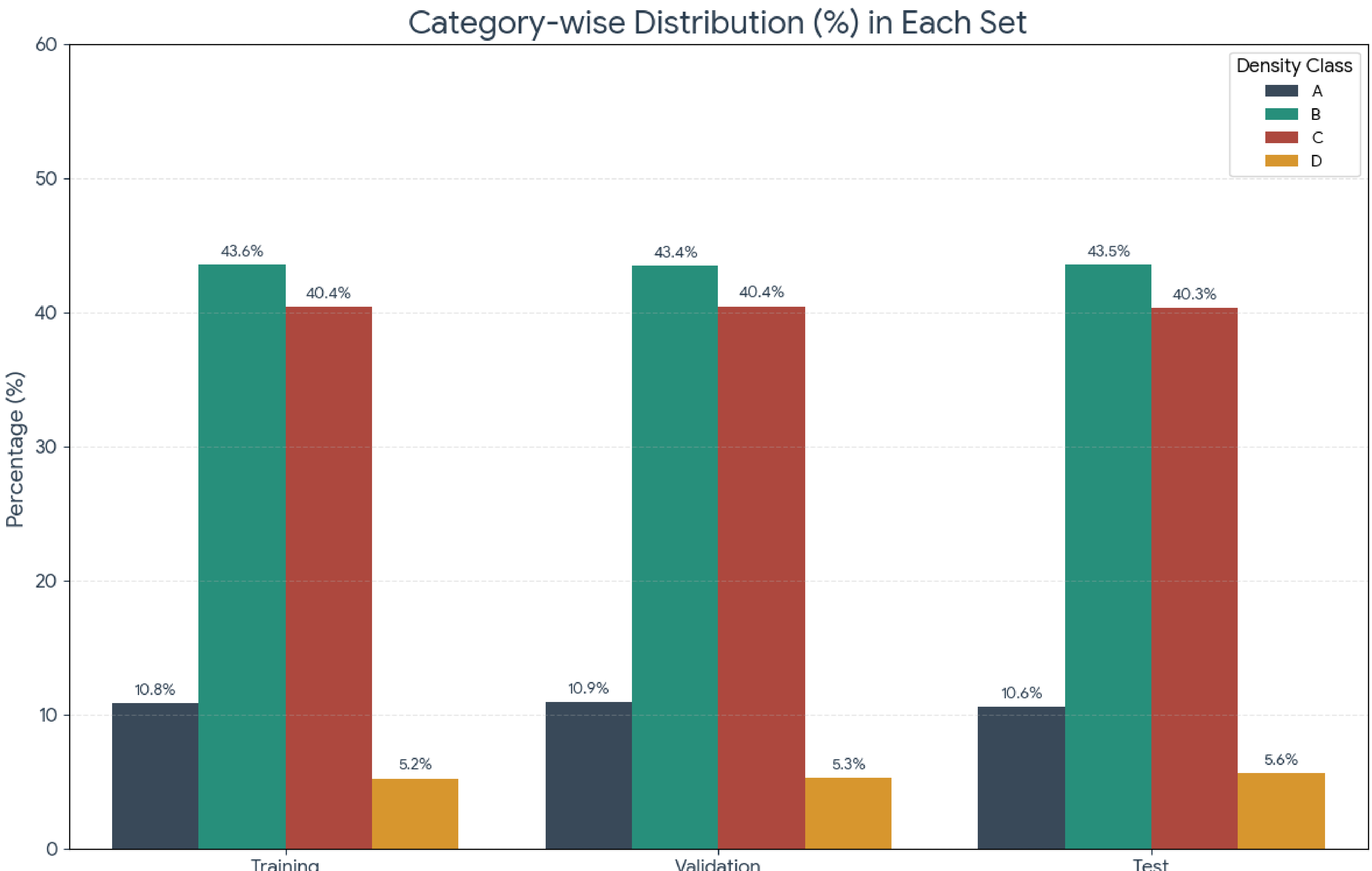


Dataset Description (RSNA)

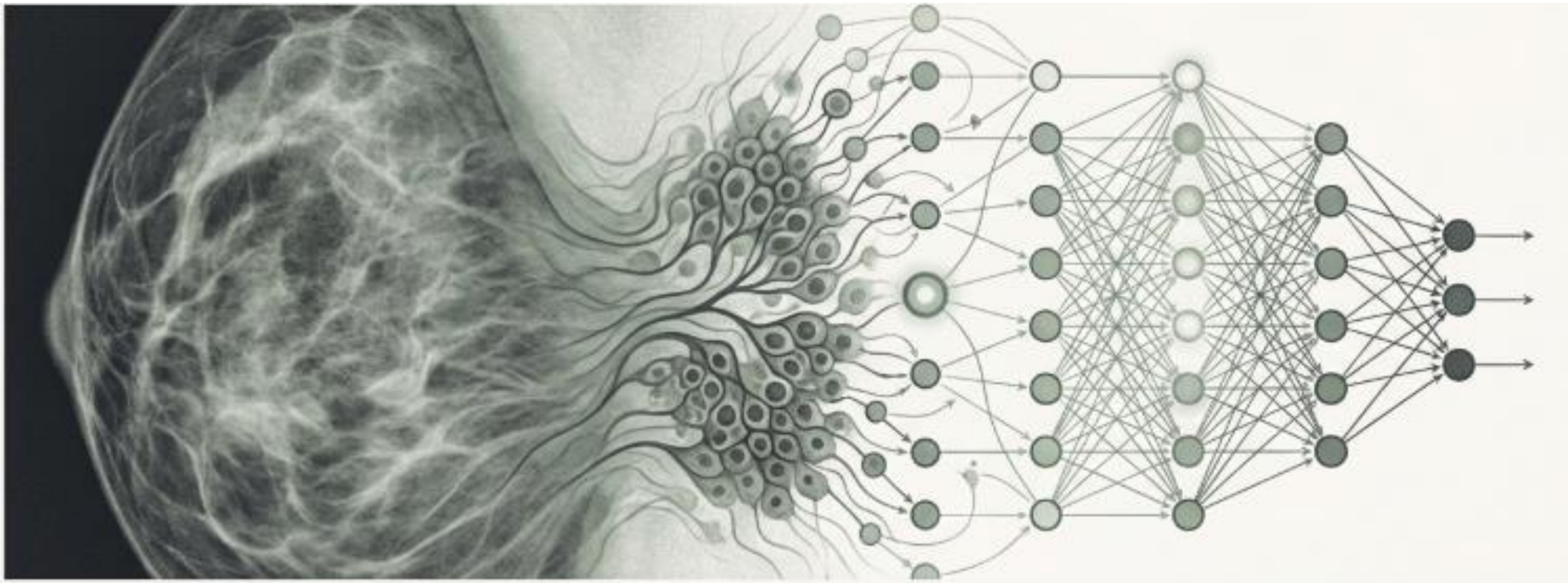
Dataset Split Distribution



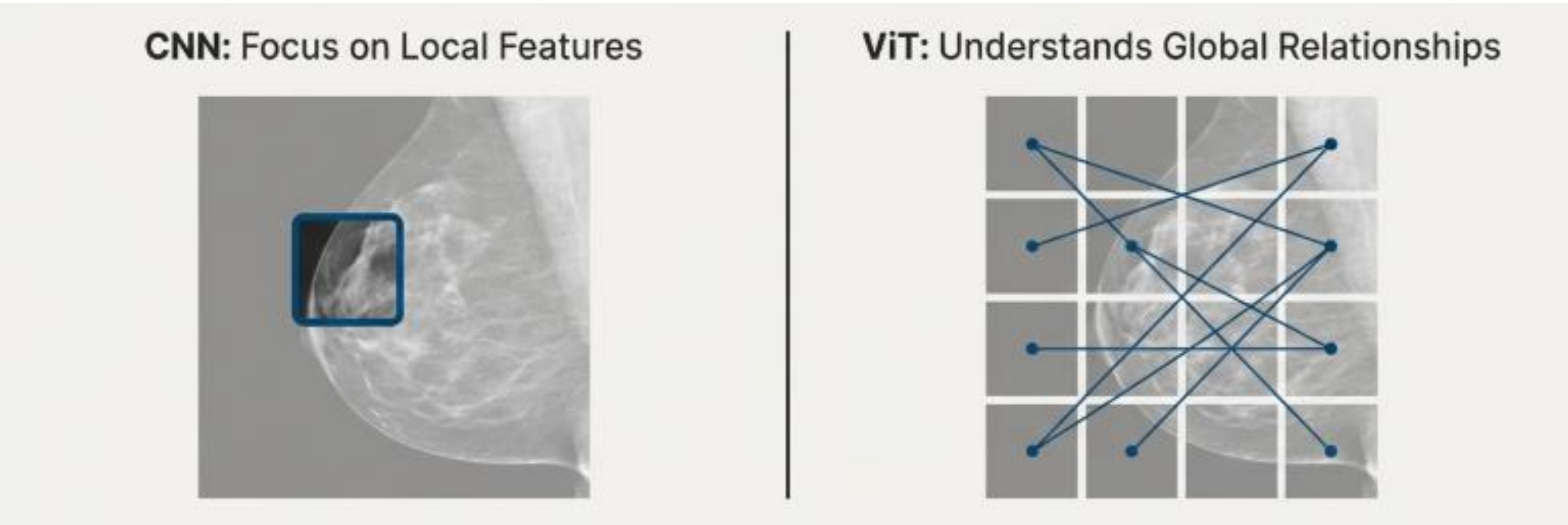



Category-wise Distribution (%) in Each Set
Density Class
A
B
C
D
Percentage (%)
0
10
20
30
40
50
60
10.8%
43.6%
40.4%
5.2%
10.9%
43.4%
40.4%
5.3%
10.6%
43.5%
40.3%
5.6%
Training
Validation
Test

Model Architecture



Model Architecture

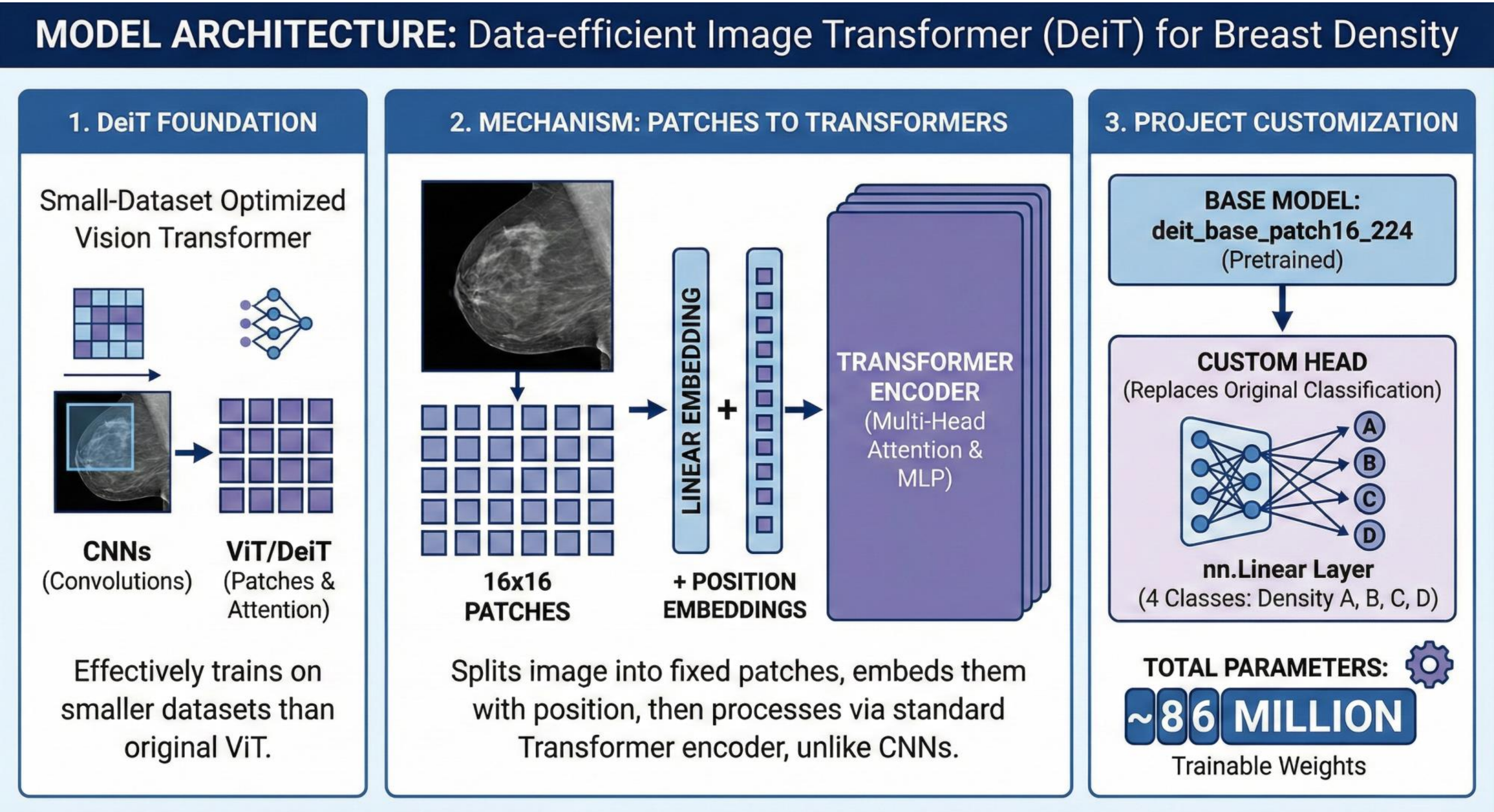


Unlike CNNs, Vision Transformers process an image's patches in parallel, allowing the model to weigh the importance of all regions simultaneously to understand the overall tissue structure.

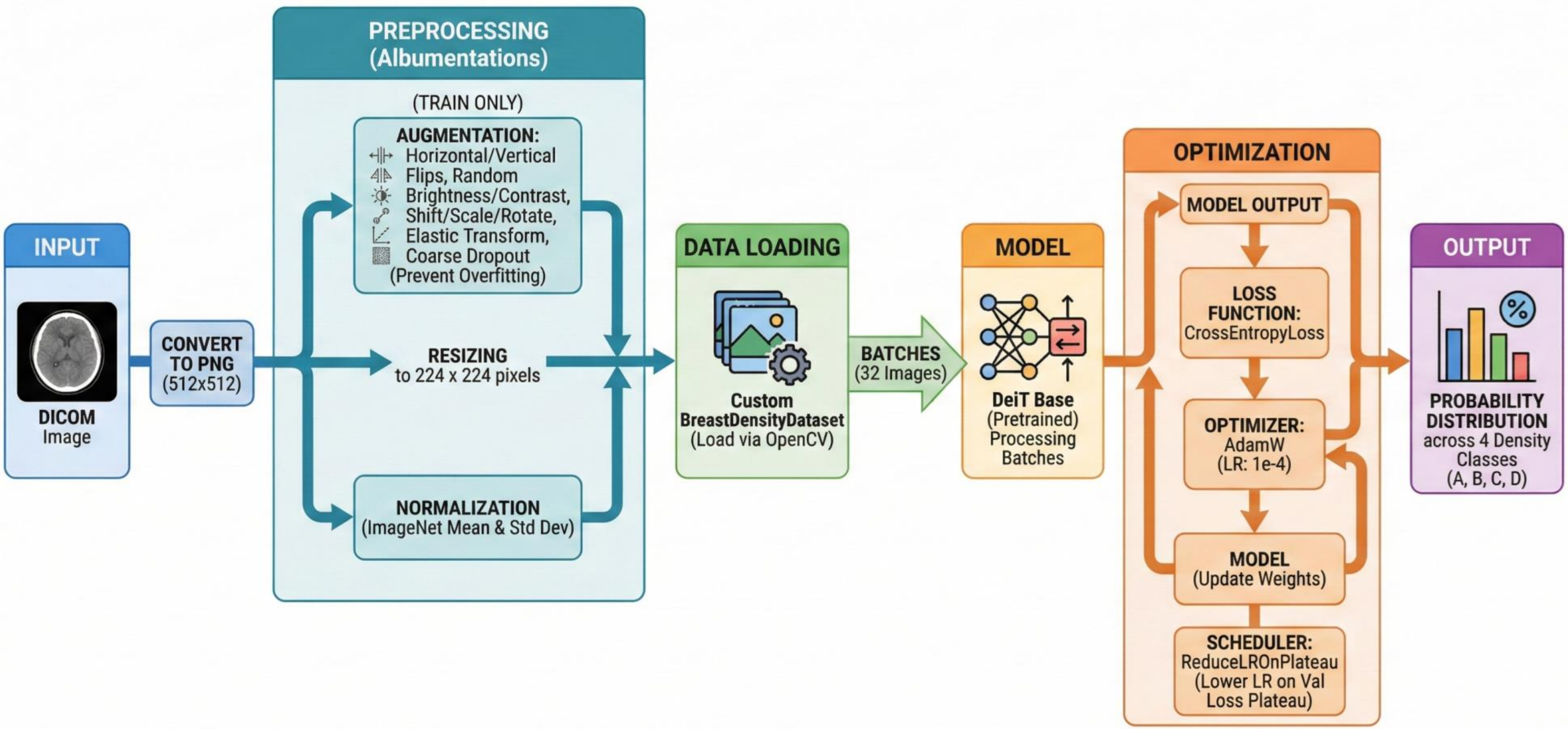
Digital Egypt Pioneers

وزارة الاتصالات وتكنولوجيا المعلومات

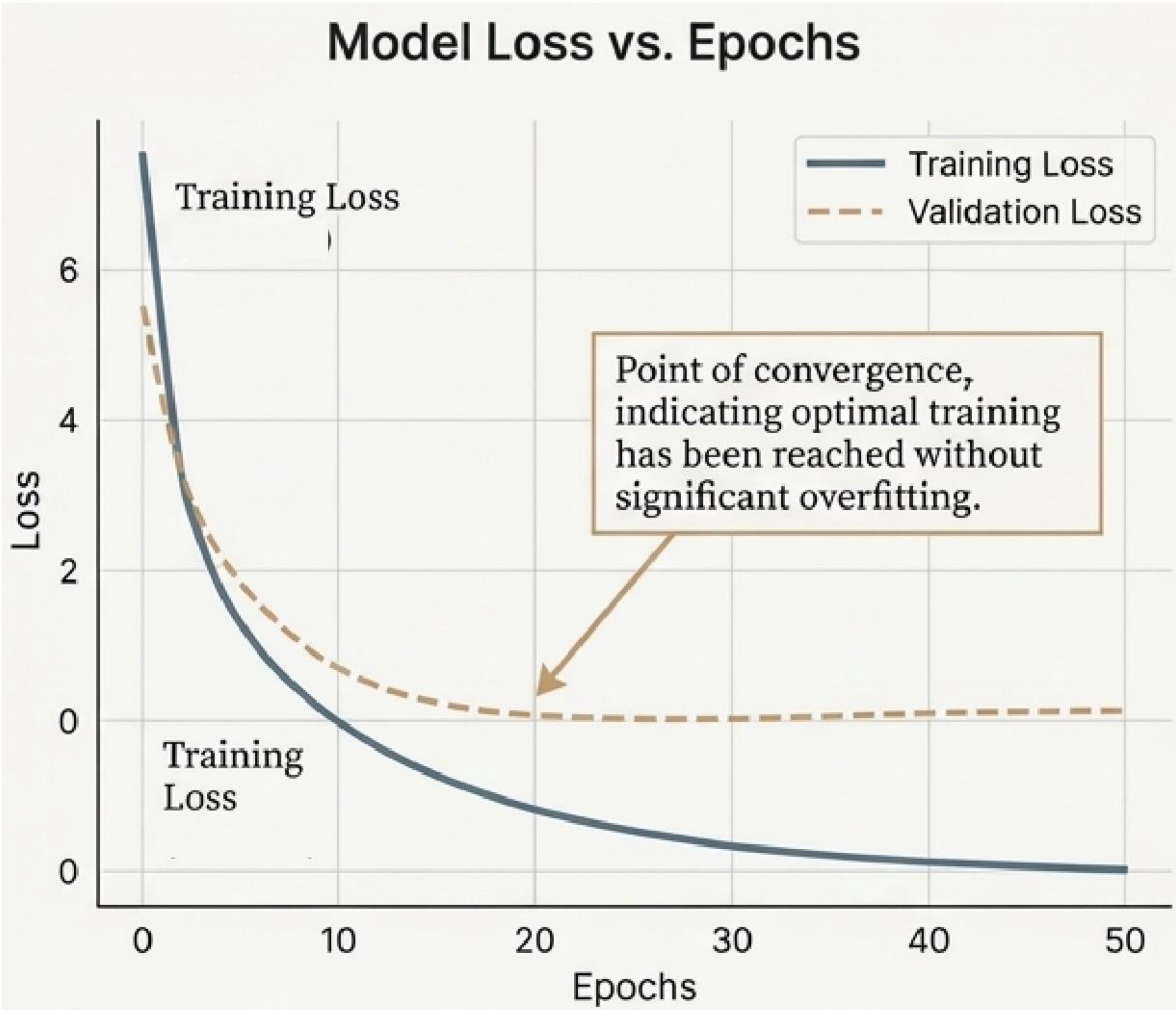
Model Architecture



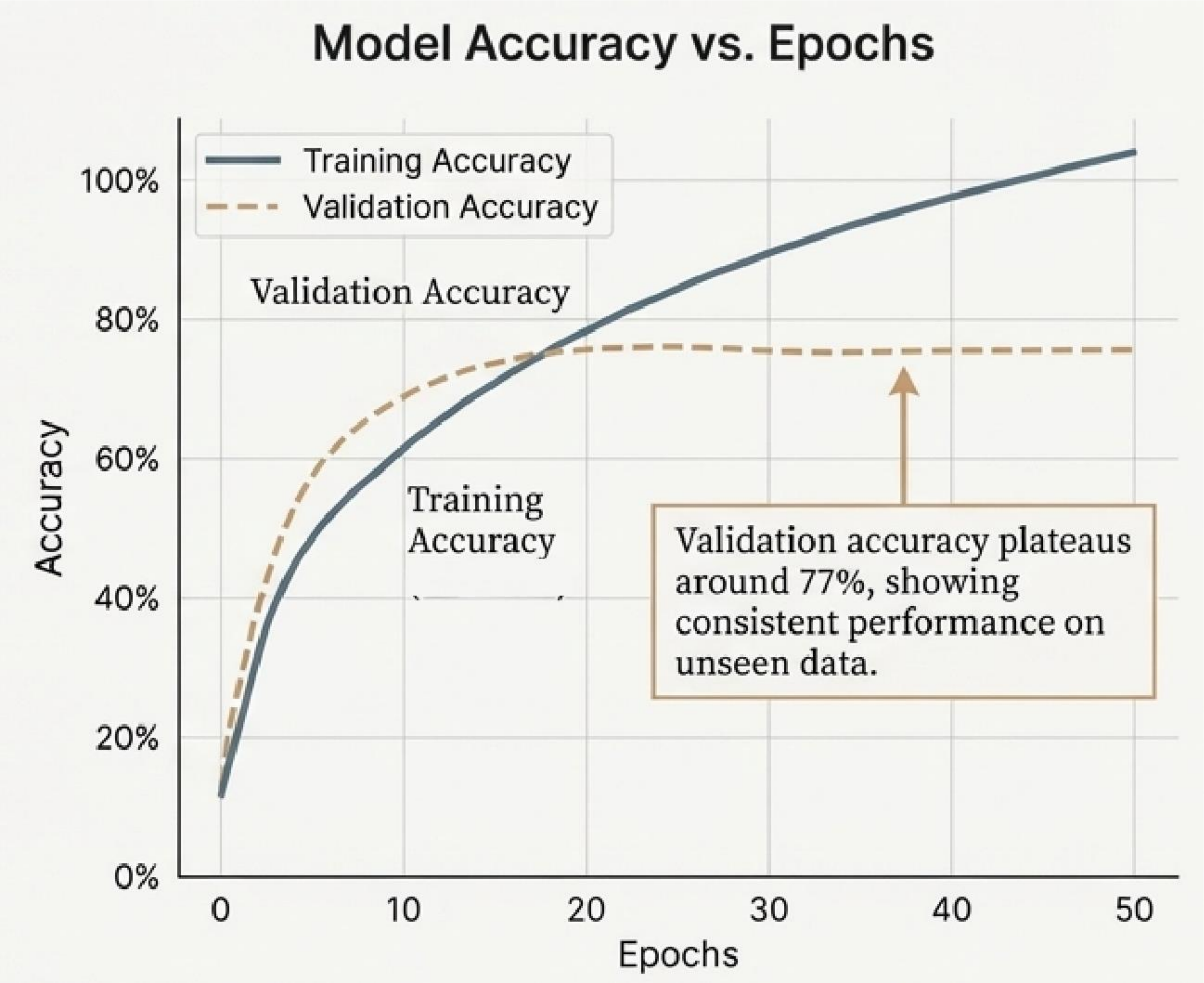
Deep Learning Pipeline Flowchart



Results


Model Loss vs. Epochs
Training Loss
Validation Loss
Training Loss
Point of convergence, indicating optimal training has been reached without significant overfitting.
Training Loss
6
4
2
0
0
Loss
0
10
20
30
40
50
Epochs

Results


Model Accuracy vs. Epochs
Training Accuracy
Validation Accuracy
Validation Accuracy
Training Accuracy
Validation accuracy plateaus around 77%, showing consistent performance on unseen data.
100%
80%
60%
40%
20%
0%
Accuracy
0
10
20
30
40
50
Epochs

Results

Density Class	Precision	Recall	F1-Score	Support
Class A	0.63	0.65	0.64	301
Class B	0.76	0.79	0.78	1,198
Class C	0.81	0.81	0.81	1,116
Class D	0.75	0.45	0.56	145
Accuracy			**0.77**	**2,760**
Macro Avg	0.74	0.68	0.70	2,760
Weighted Avg	0.77	0.77	0.77	2,760

Support refers to the **number of actual occurrences** of each class in the validation dataset

$$\text{Macro Avg} = \frac{\text{Score(A)} + \text{Score(B)} + \text{Score(C)} + \text{Score(D)}}{}$$

$$\text{Weighted Avg} = \frac{(\text{Score}_A \times 301) + (\text{Score}_B \times 1198) + (\text{Score}_C \times 1116) + (\text{Score}_D \times 145)}{2760}$$

Results

The Model Achieves 77% Weighted Average Accuracy on the Validation Set

Density Class	Precision	Recall	F1-Score	Support (N)
Class A	0.63	0.65	0.64	301
Class B	0.76	0.79	0.78	1,198
Class C	0.81	0.81	0.81	1,116
Class D	0.75	0.45	0.56	145
Weighted Avg	**0.77**	**0.77**	**0.77**	**2,760**

Key Insight

Excellent performance is observed on the most common classes (B and C). C). The lower recall (0.45) for Class D is a key area for improvement, likely due to the low number of samples (145) and its inherent visual similarity to Class C.

Digital Egypt Pioneers

وزارة الاتصالات
وتكنولوجيا المعلومات

Results

Density Class	Precision	Recall	F1-Score	Support
Class A	0.63	0.65	0.64	301
Class B	0.76	0.79	0.78	1,198
Class C	0.81	0.81	0.81	1,116
Class D	0.75	0.45	0.56	145
Weighted Avg	**0.77**	**0.77**	**0.77**	**2,760**

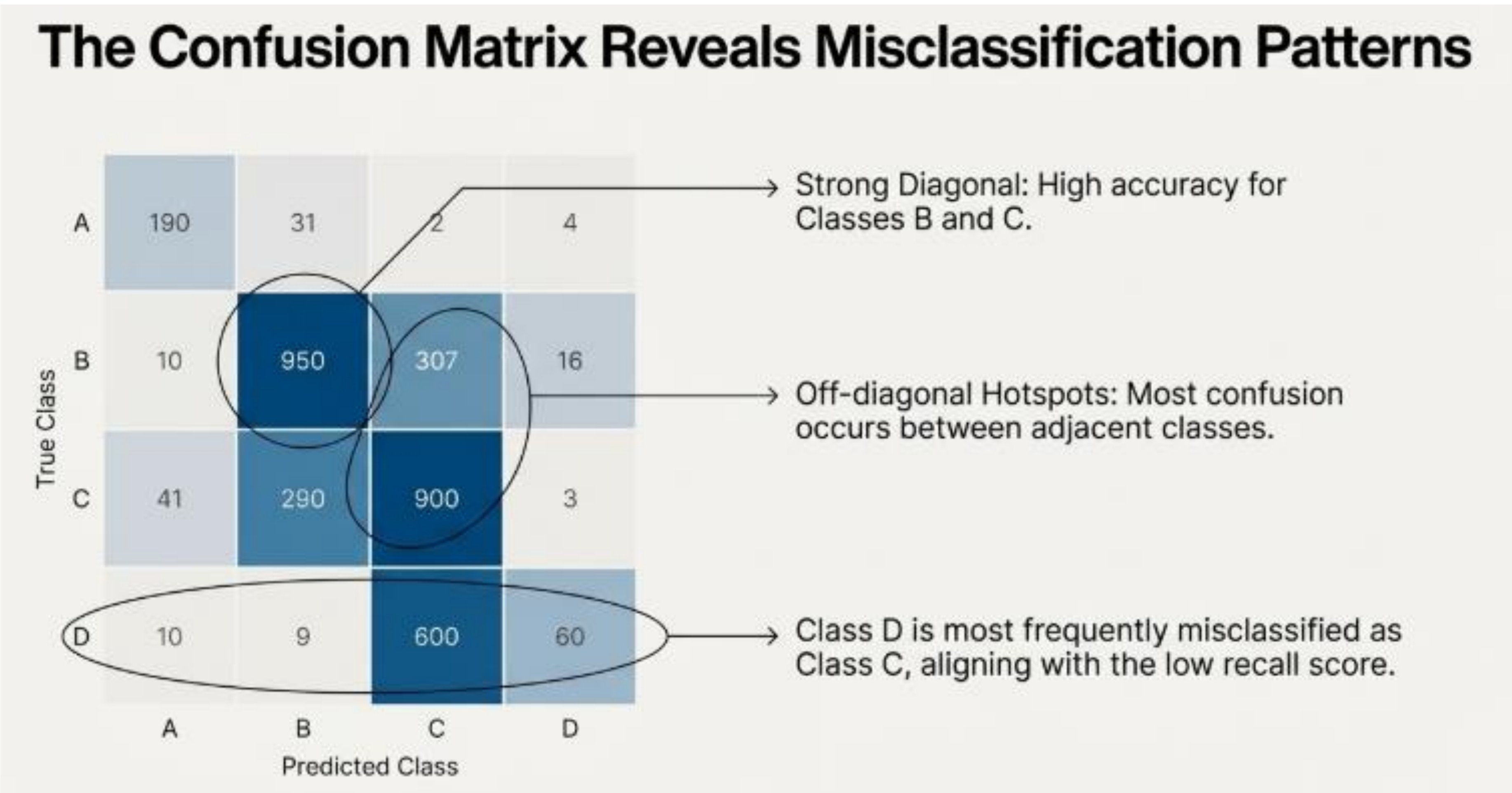
High Performance

Excellent F1-scores for the most common classes, B and C, demonstrating robustness where it matters most.

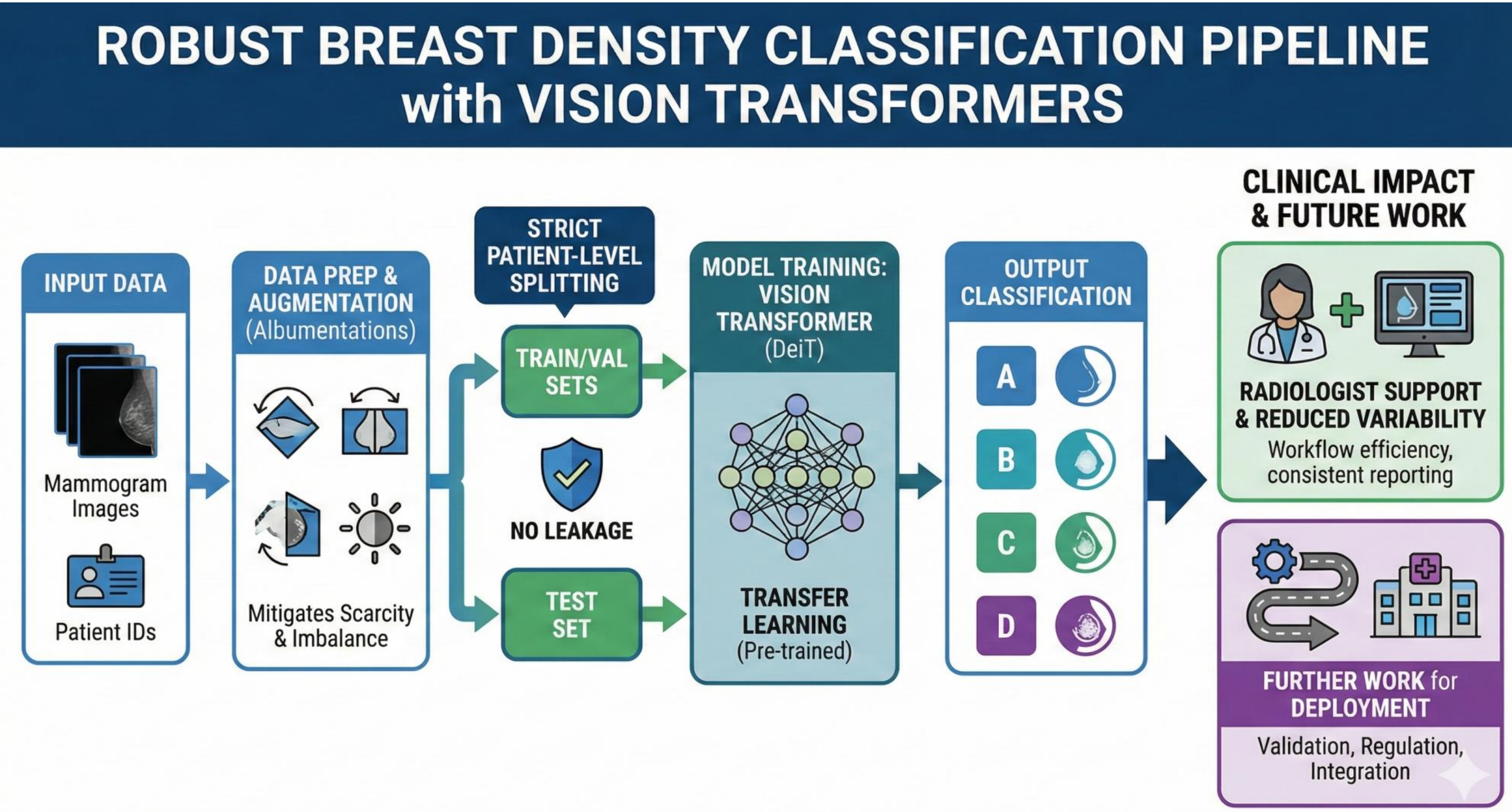
Imbalance Challenge

Lower recall (0.45) for Class D highlights the challenge of predicting the least frequent class, a known issue in imbalanced datasets.

Results

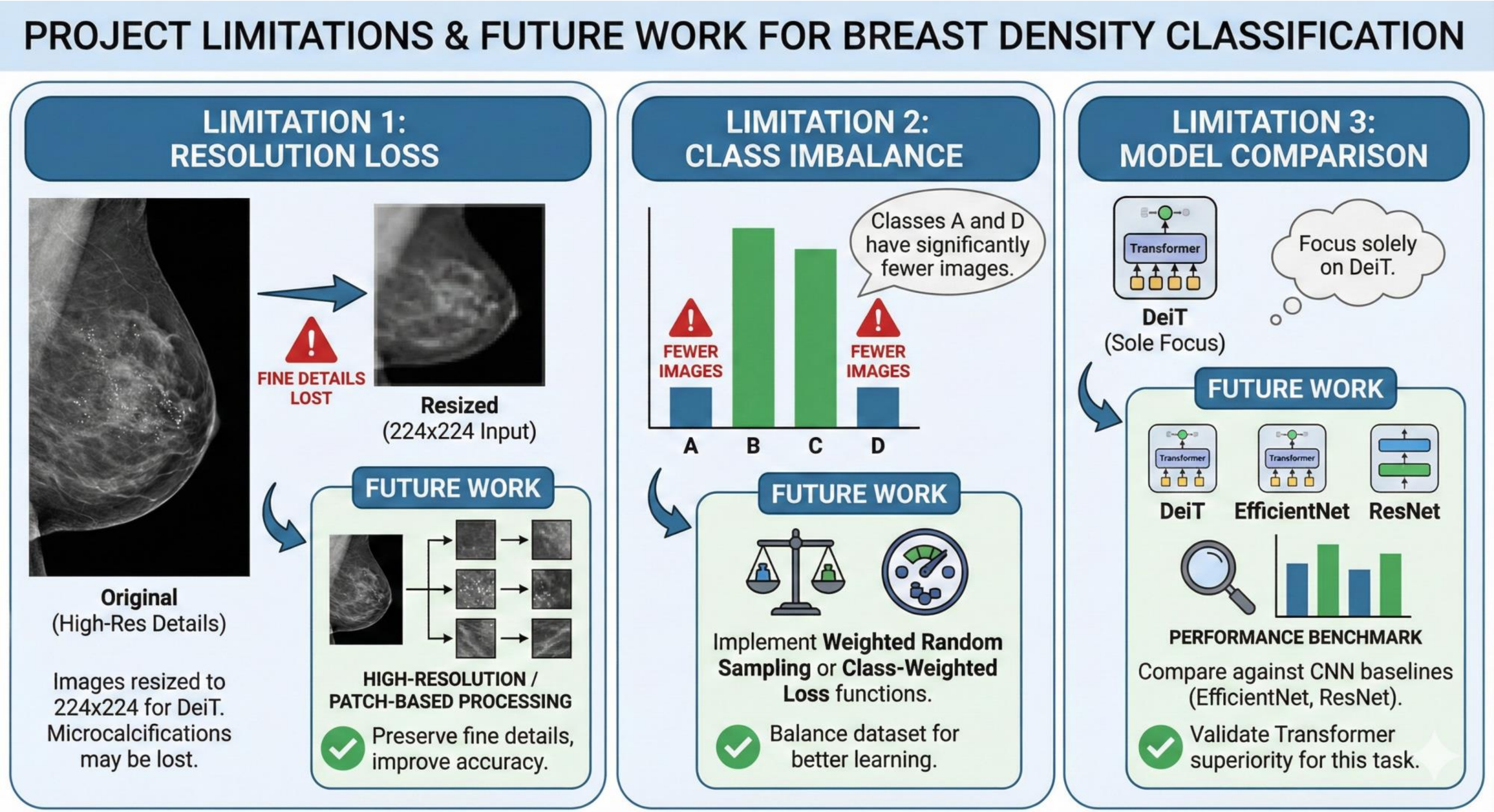

The Confusion Matrix Reveals Misclassification Patterns
True Class
A
B
C
D
190
31
2
4
10
950
307
16
41
290
900
3
10
9
600
60
A
B
C
D
Predicted Class
Strong Diagonal: High accuracy for Classes B and C.
Off-diagonal Hotspots: Most confusion occurs between adjacent classes.
Class D is most frequently misclassified as Class C, aligning with the low recall score.

Conclusion



ROBUST BREAST DENSITY CLASSIFICATION PIPELINE with VISION TRANSFORMERS
INPUT DATA
Mammogram Images
Patient IDs
DATA PREP & AUGMENTATION (Albumentations)
Mitigates Scarcity & Imbalance
STRICT PATIENT-LEVEL SPLITTING
TRAIN/VAL SETS
NO LEAKAGE
TEST SET
MODEL TRAINING: VISION TRANSFORMER (DeiT)
TRANSFER LEARNING (Pre-trained)
OUTPUT CLASSIFICATION
A
B
C
D
CLINICAL IMPACT & FUTURE WORK
RADIOLOGIST SUPPORT & REDUCED VARIABILITY
Workflow efficiency, consistent reporting
FURTHER WORK for DEPLOYMENT
Validation, Regulation, Integration

Limitations & Future Work



References

- RSNA Breast Imaging Dataset.
- Touvron, H., et al. "Training data-efficient image transformers & distillation through attention." (2020).