Figure 1

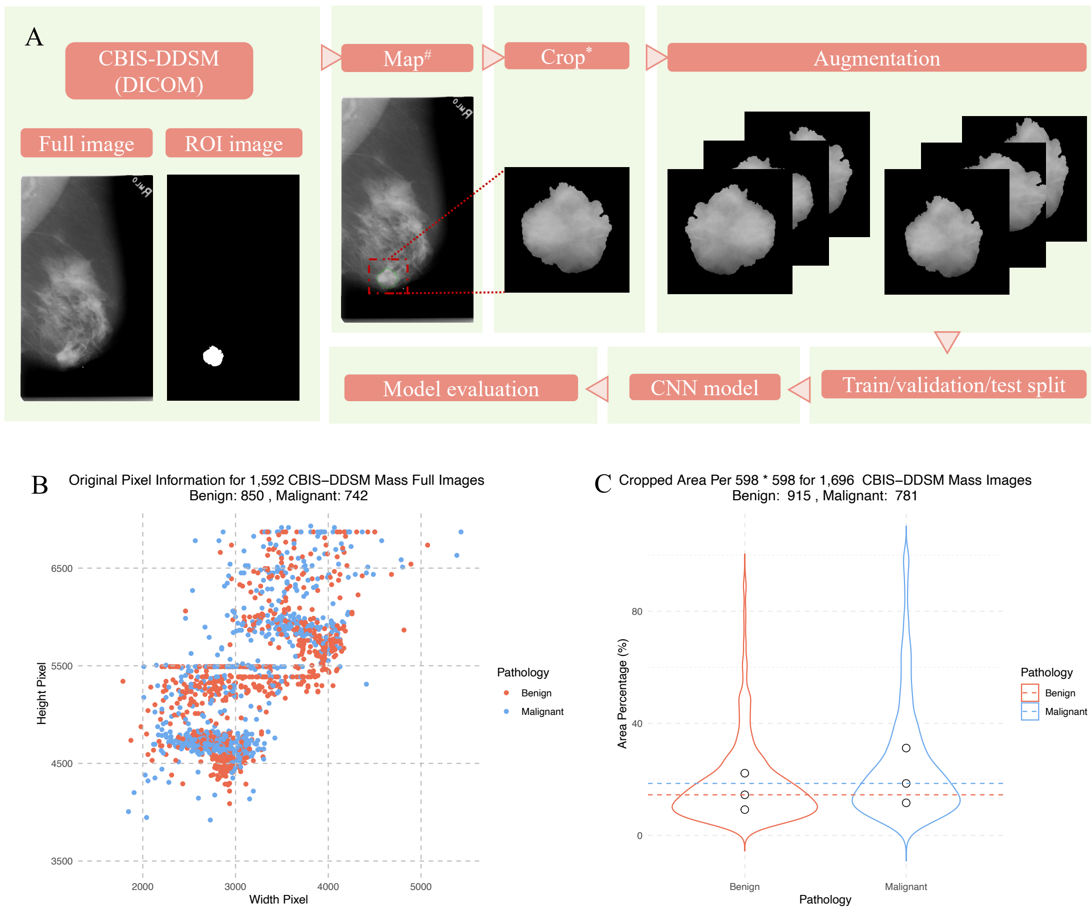

Figure 1: Overview of CBIS-DDSM Mass Dataset Application for Breast Cancer Diagnosis. A). A brief summary of our model training process. After converting the DICOM image to PNG format, we leverage ROIs as masks and map them to the corresponding full images to extract and crop the abnormal regions from both benign and malignant full images. Then, image augmentation is performed by flipping, rotation, zoom, shear, and shift. The cropped and augmented data are then split into 70% for training, 20% for validation, and 10% for testing. We selected CNN as our model structure, and the performance is further evaluated by various metrics, including AUROC, accuracy, and others. DICOM: Digital Imaging and Communications in Medicine. PNG: portable network graphics. CNN: convolutional neural network. AUROC: Area under the receiver operating characteristic curve. # 65 mismatched pixels between the full and ROI images are identified, and adjustments are implemented to the pixels of the ROIs before matching them with the corresponding full images; further details are provided in the method section. * There are three subcategories of cropped images in total, as detailed in Figure 2. B). Pixel information of 1,592 full CBIS-DDSM mass images. C). Cropped abnormal area per 598 × 598 region from 1,696 mapped CBIS-DDSM mass images. Some full images contain multiple abnormal regions, leading to a greater number of mapped images compared to the original full CBIS-DDSM mass images.

Figure 2

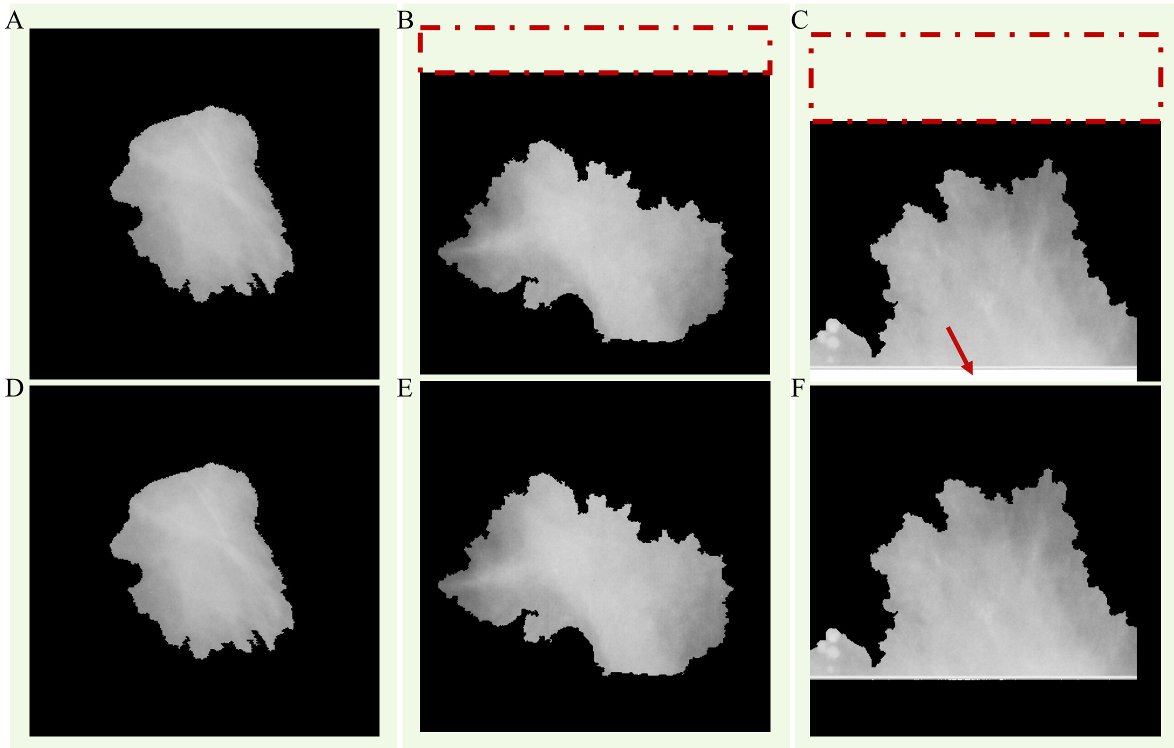

Figure 2. Illustration of Cropped Abnormal Regions: Original vs. Processed Versions. Panels 2A, 2B, and 2C showcase the original cropped images, while panels 2D, 2E, and 2F depict the corresponding processed versions. The red dashed box indicates areas with missing pixels, and the red arrow highlights an example of unwanted white areas.

Figure 3

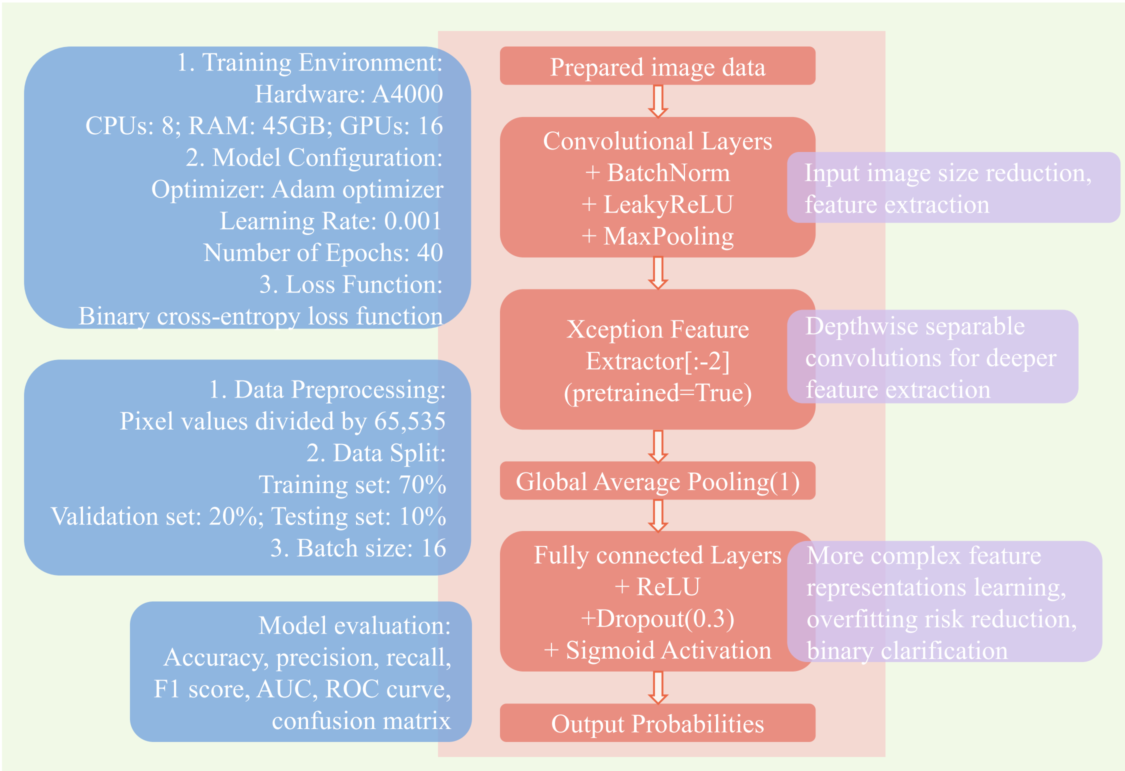

Figure 3. Key Aspects in Model Development, Training, and Evaluation, as well as the Workflow of Model Structure. For model development environment, computational tasks are executed using an A4000 GPU alongside 16 GPUs, 8 CPUs, and 45GB of RAM. The model is optimized with Adam optimizer and trained for 40 epochs with a learning rate of 0.001, assessing performance with a binary cross-entropy loss function. Data preprocessing involves normalizing pixel values by 65,535. The dataset is partitioned into training (70%), validation (20%), and testing (10%) sets, employing a batch size of 16 per training iteration. Model evaluation encompasses metrics including accuracy, precision, recall, F1 score, AUROC, ROC curve, and the confusion matrix for assessment of effectiveness and performance. The workflow of model structure, depicted in orange, includes processing prepared image data through convolutional layers, Batch Normalization, LeakyReLU activation, and MaxPooling for size reduction and feature extraction. Leveraging pretrained weights from an Xception feature extractor captures high-level representations, followed by Global Average Pooling to retain essential information. Fully connected layers with ReLU activation, Dropout at a rate of 0.3, and a Sigmoid activation function generate output probabilities. The purple highlights in the figure emphasize the rationale behind this structured approach.

Figure 4

A

Evaluation Metrics					
	AUROC	Accuracy	Precision	Recall	F1 Score
Value(%)	86.3	79.5	78.3	75.4	76.9
95% CI	[0.842, 0.885]	[0.771, 0.820]	[0.748, 0.819]	[0.714, 0.795]	[0.738, 0.799]

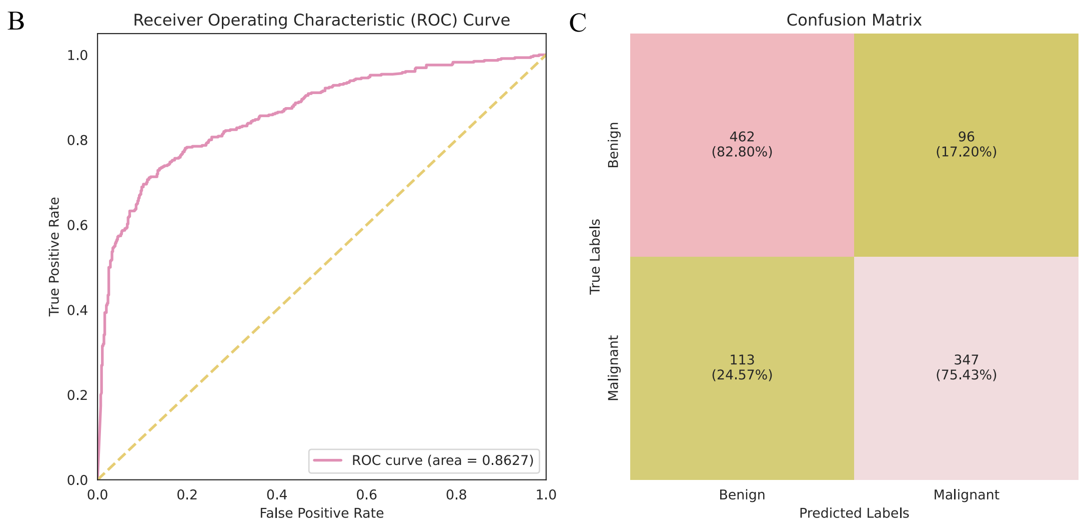

Figure 4. Evaluation matrix of model's performance on processed CBIS-DDSM mass image data. A): Evaluation matrix on AUROC, accuracy, precision, recall and F1 score with 95% CI. CI: confidential interval. B): ROC curve. C): Confusion matrix. CI: confidential interval.