MEDIZINISCHE UNIVERSITÄT WIEN

acmit AUSTRIAN CENTER FOR MEDICAL INNOVATION AND TECHNOLOGY

DANUBE PRIVATE UNIVERSITY Fakultät Medizin/Zahnmedizin

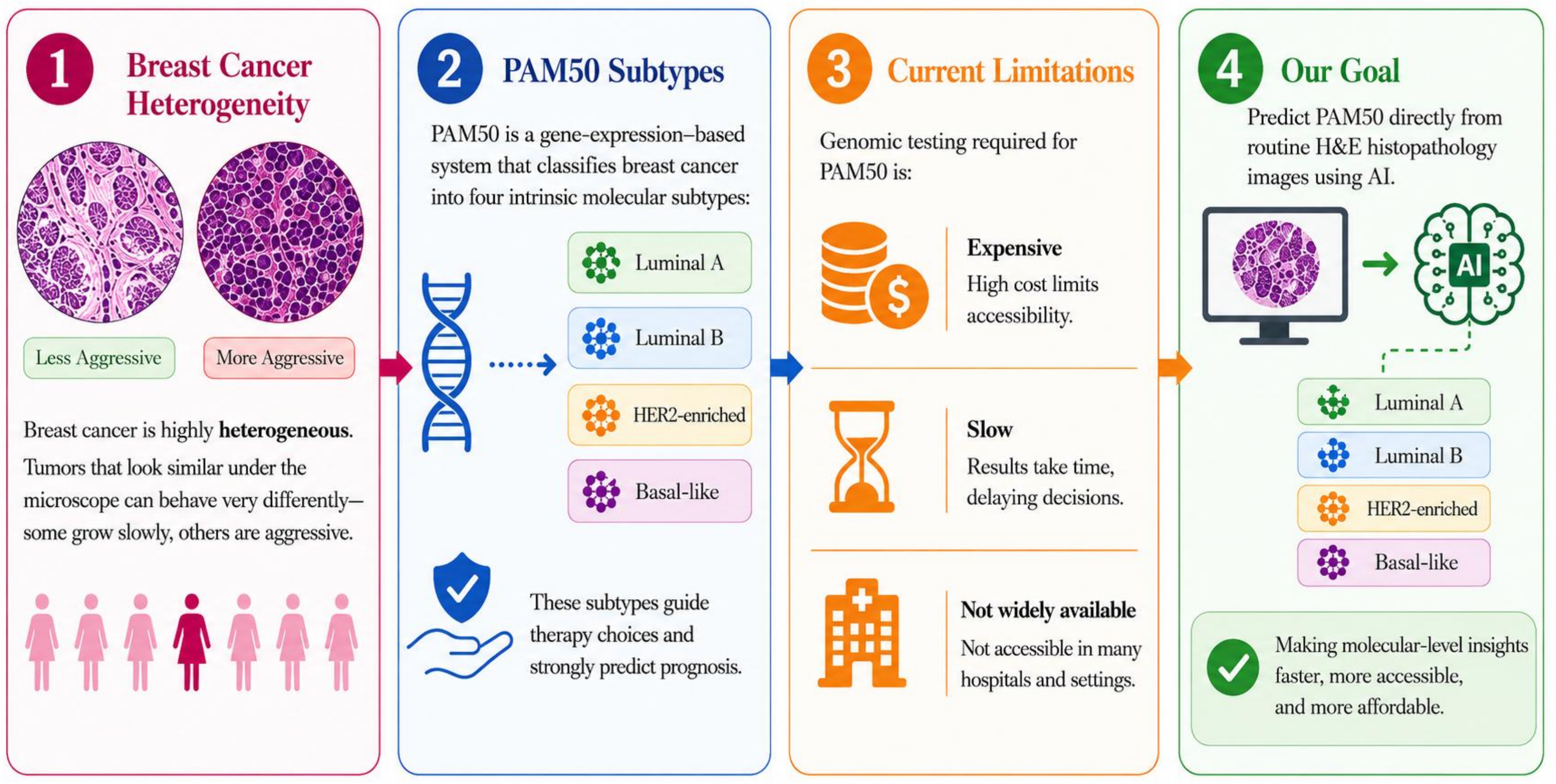
A deep learning pipeline for PAM50 subtype classification using histopathology images and multi-objective patch selection

Arezoo Borji [1, 2, 3], Gernot Kronreif [1], Bernhard Angermayr[2, 4], Francisco Mario Calisto[5,6], Ali Abbasian Ardakani [2], Wolfgang Birkfellner [3], Inna Servetnyk[2], Yinyin Yuan[7], Sepideh Hatamikia [2, 1, 3*]

1) Austrian Center for Medical Innovation and Technology, Wiener Neustadt, Austria. 2) Danube Private University, Krems an der Donau, Austria. 3) Department of Medical Physics and Biomedical Engineering, Medical University of Vienna, Vienna, Austria. 4) Patho im Zentrum, Saint Pölten, Austria. 5) Institute for Systems and Robotics (ISR-Lisboa/LARSyS), Instituto Superior Técnico, University of Lisbon, 1049-001, Lisbon, Portugal. 6) BreastScreening-AI, SensiPerception, Coimbra, Portugal. 7)The University of Texas MD Anderson Cancer Center, Houston, TX, United States

Objective, what and why

- ✓ Breast cancer is one of the leading causes of cancer-related mortality among women worldwide.
- ✓ PAM50 subtype classification is clinically valuable, but current assessment usually relies on costly gene-expression profiling.
- ✓ This study aimed to classify PAM50 breast cancer subtypes directly from H&E histopathology slides as a cost-effective and clinically accessible alternative to genomic assays.
- ✓ We present a novel multi-objective optimization framework for the automatic selection of informative and representative tissue patches. By reducing redundancy, the framework improves classification efficiency and outperforms state-of-the-art methods.



1 Breast Cancer Heterogeneity

Breast cancer is highly **heterogeneous**. Tumors that look similar under the microscope can behave very differently—some grow slowly, others are aggressive.

2 PAM50 Subtypes

PAM50 is a gene-expression–based system that classifies breast cancer into four intrinsic molecular subtypes:

These subtypes guide therapy choices and strongly predict prognosis.

3 Current Limitations

Genomic testing required for PAM50 is:

Expensive High cost limits accessibility.

Slow Results take time, delaying decisions.

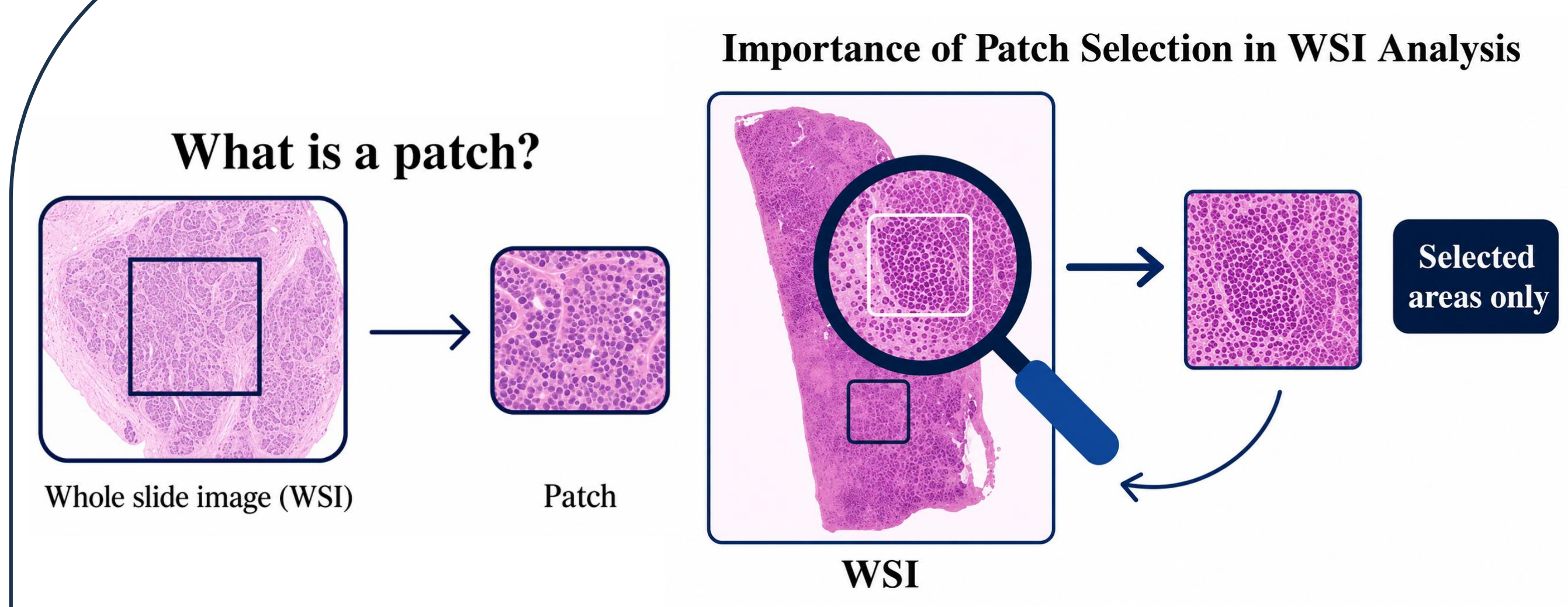
Not widely available Not accessible in many hospitals and settings.

4 Our Goal

Predict PAM50 directly from routine H&E histopathology images using AI.

Making molecular-level insights faster, more accessible, and more affordable.

Methods



- ➢ Division of WSI into candidate patches
- ➢ A multi-objective optimization strategy to select informative regions based on tissue quality followed by preprocessing.
- ➢ Feature extraction using ResNet were performed from the selected patches.
- ➢ Classification of four subtypes and evaluate on TCGA dataset as an internal dataset with 627 WSIs and CPTAC dataset as an external dataset 122 WSIs (stained-normalized patches).
- ➢ **NSGA-II-based Multi-Objective Patch Selection**

Candidate patch subsets are encoded as chromosomes and evaluated using four objectives: **Diversity, informativeness, compactness,** and **reliability**.

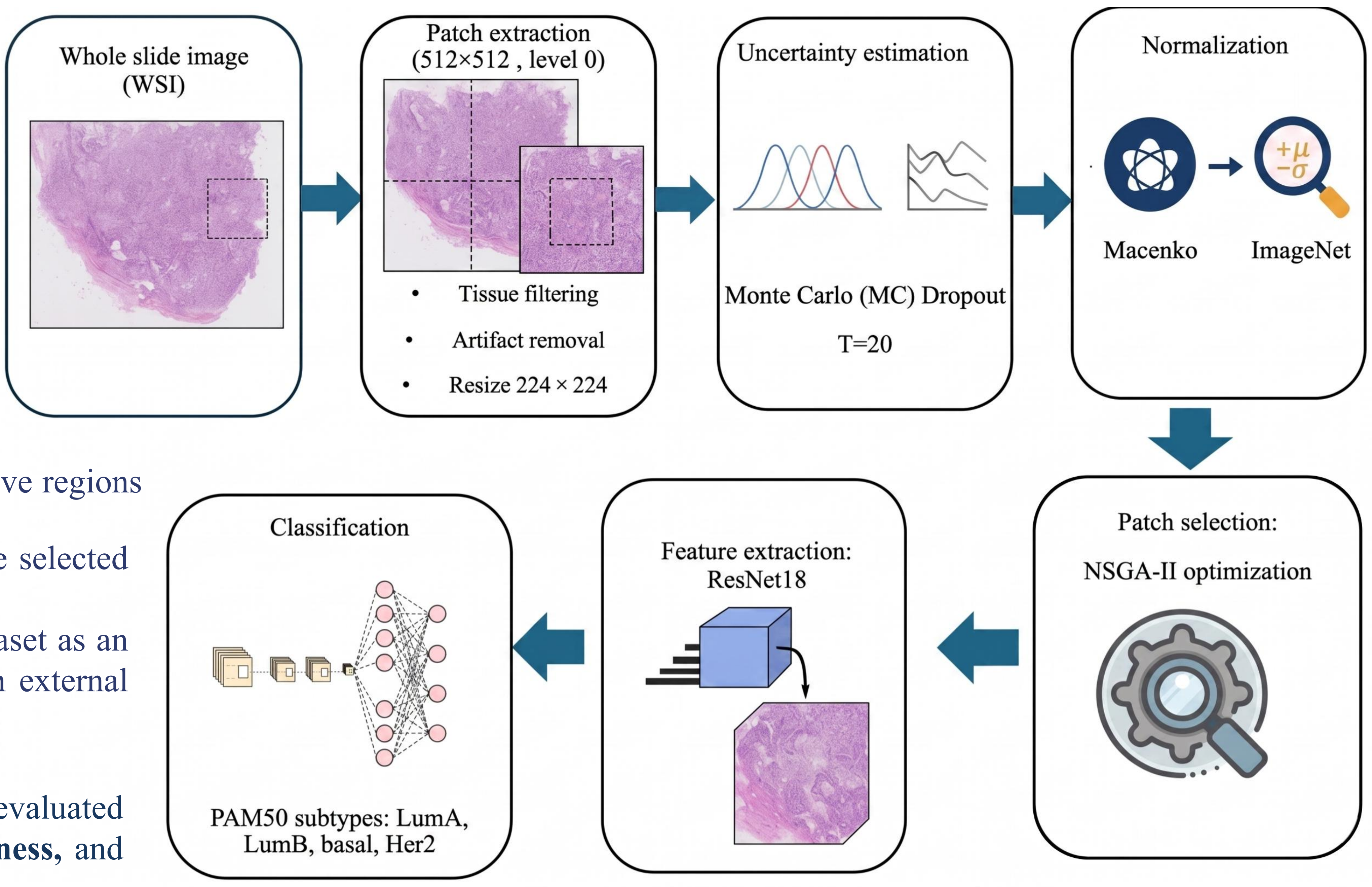


Figure 1: The overall procedure of the proposed method

Results

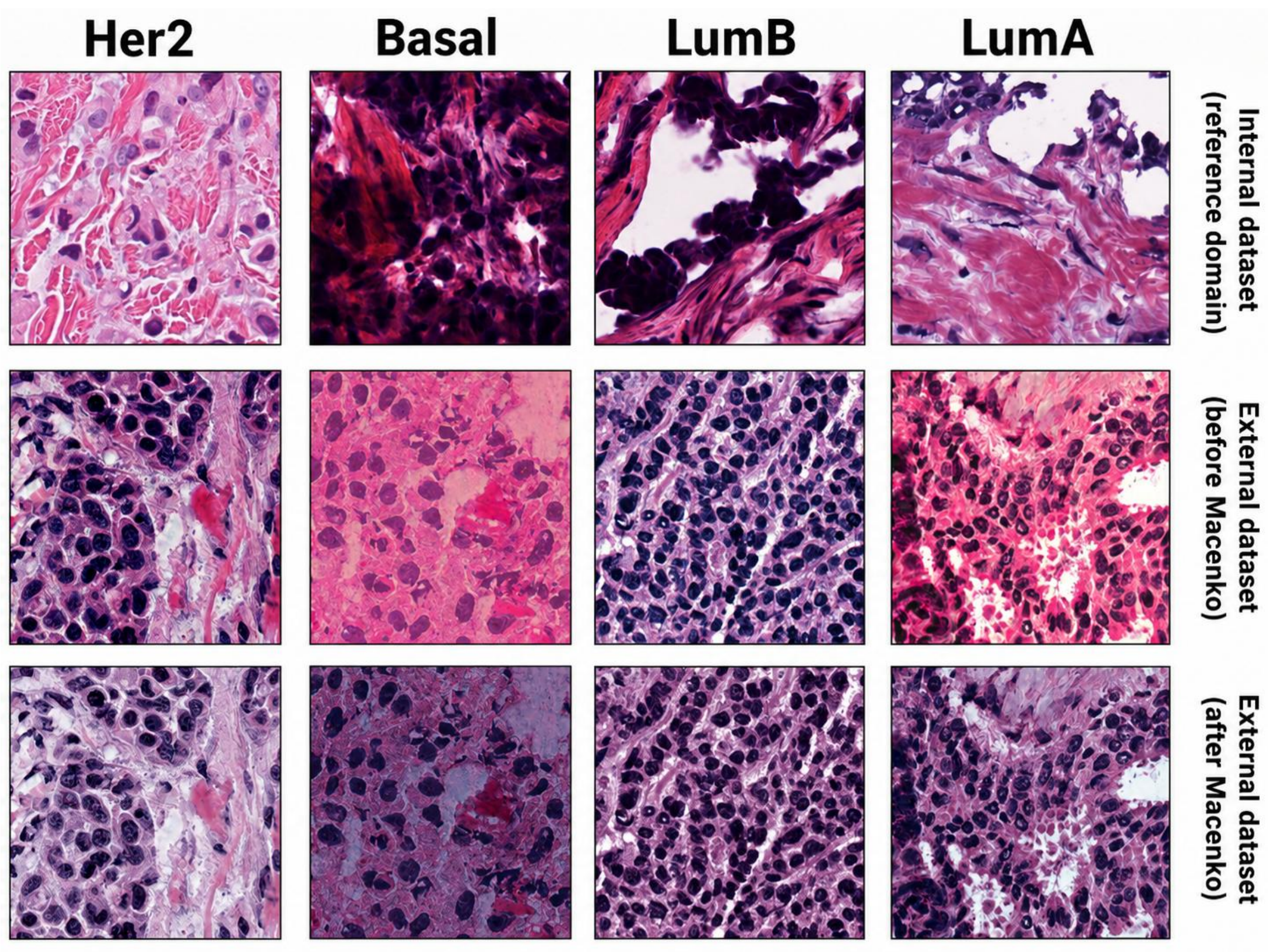


Figure 2: Representative H&E-stained tissue patches from the internal (TCGA) for staining CPTAC patches across four molecular subtypes

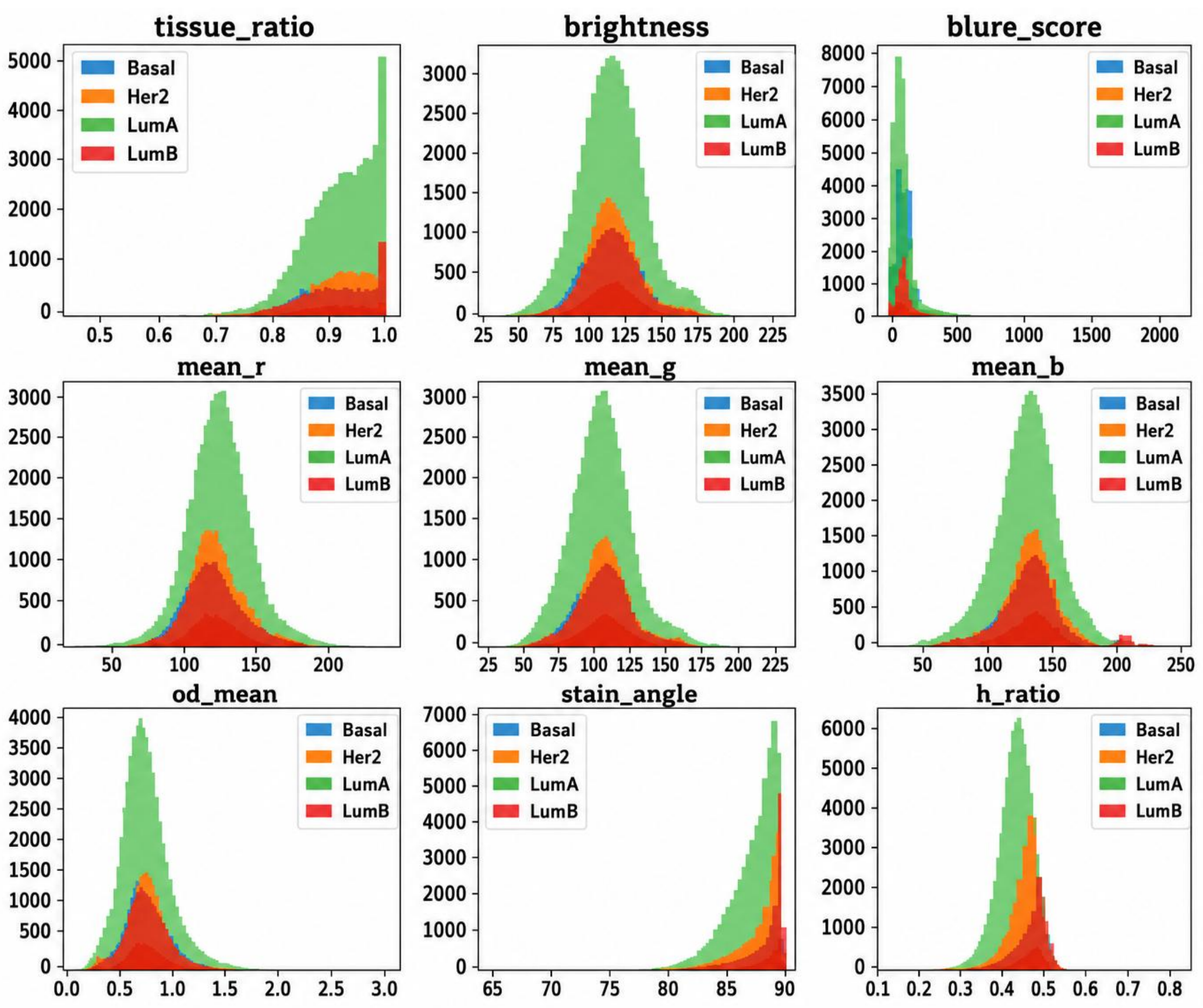


Figure 3: Distribution of image quality and stain characteristics across the four subtypes in the internal (TCGA) dataset

Figure 3 shows that the reference image for stain normalization was selected from the **top 50 ranked patches**, ranked using the statistical features shown in this figure.

Table 1: Model performance comparison on internal vs. external test dataset

Class	TCGA				CPTAC			
	Precision	Recall	F1-score	AUC	Precision	Recall	F1-score	AUC
Basal	0.8951	0.9599	0.9264	0.9944	0.8291	0.8867	0.8569	0.9757
Her2	0.8606	0.9462	0.9014	0.9945	0.8811	0.7930	0.8347	0.9707
LumA	0.9750	0.9243	0.9490	0.9907	0.8715	0.7968	0.8325	0.9658
LumB	0.8922	0.9263	0.9089	0.9896	0.8075	0.9013	0.8518	0.9744
Overall	0.9057	0.9392	0.9214	0.9923	0.8473	0.8444	0.8440	0.9716

On the internal dataset, the model achieved an accuracy of **0.9325**, precision **0.9057**, recall **0.9392**, F1-score **0.9214**, and AUC **0.9923**. For the external evaluation, the selected patches from CPTAC dataset has been stained (Figure 2), then model achieved an accuracy **0.8445**, precision **0.8473**, recall **0.8444**, F1-score **0.8440**, and AUC of 0.9716 (Table 1) on them.

Conclusion

These findings show that the proposed optimization-based patch selection could **achieve high performance** and improve the computational efficiency of histopathology-based PAM50 classification compared to existing methods, proposing an imaging-based alternative to aid clinical decision-making.