From Cropping Strategies to Multimodal Fusion: A Systematic Evaluation and Ablation Study on ISIC 2019 Skin Lesion Classification

Group 9 Project Presentation

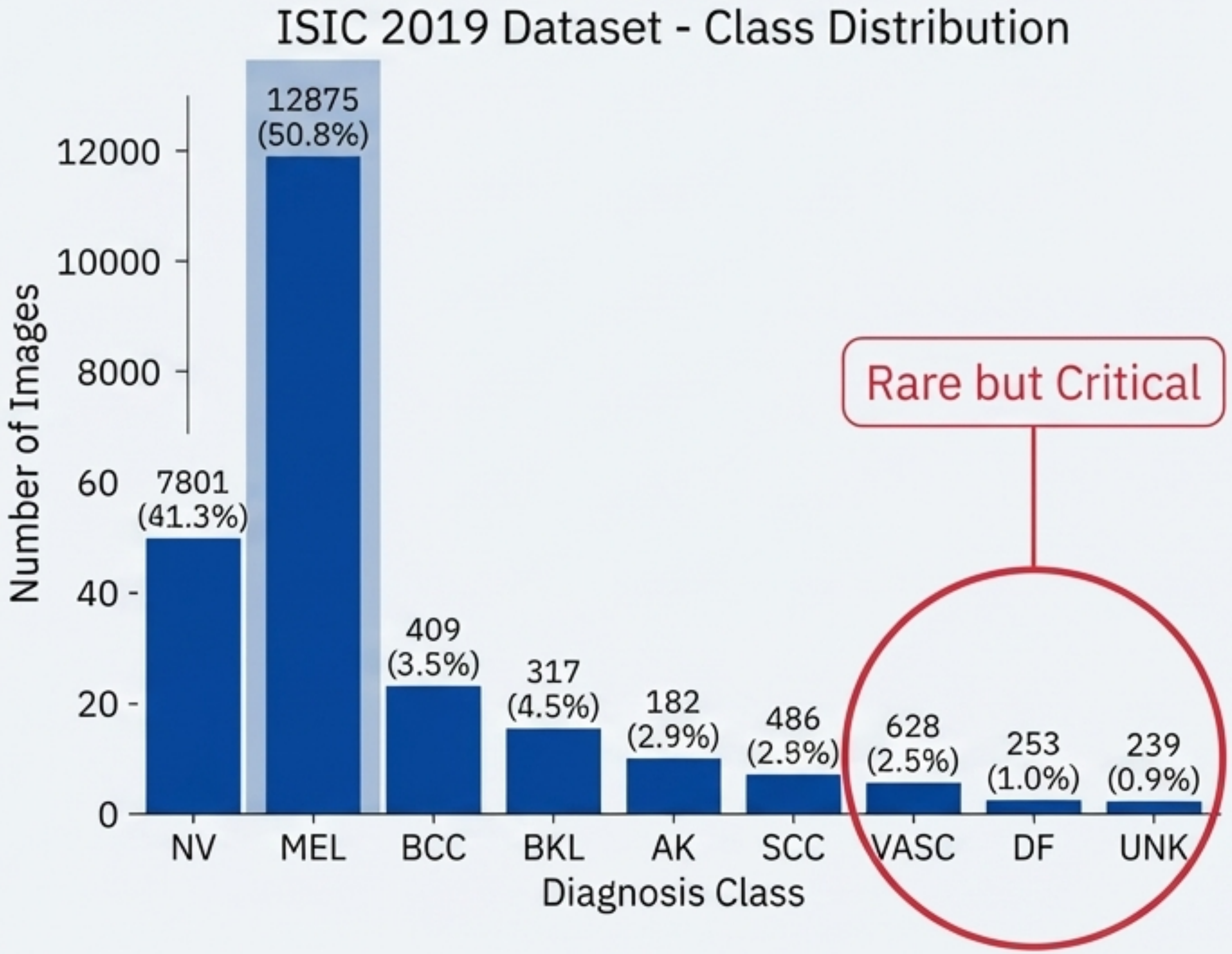
Du Yuxi (2330026036)
Hou Shuoran (2330026054)
Lu Yunxiao (2330026114)

Introduction

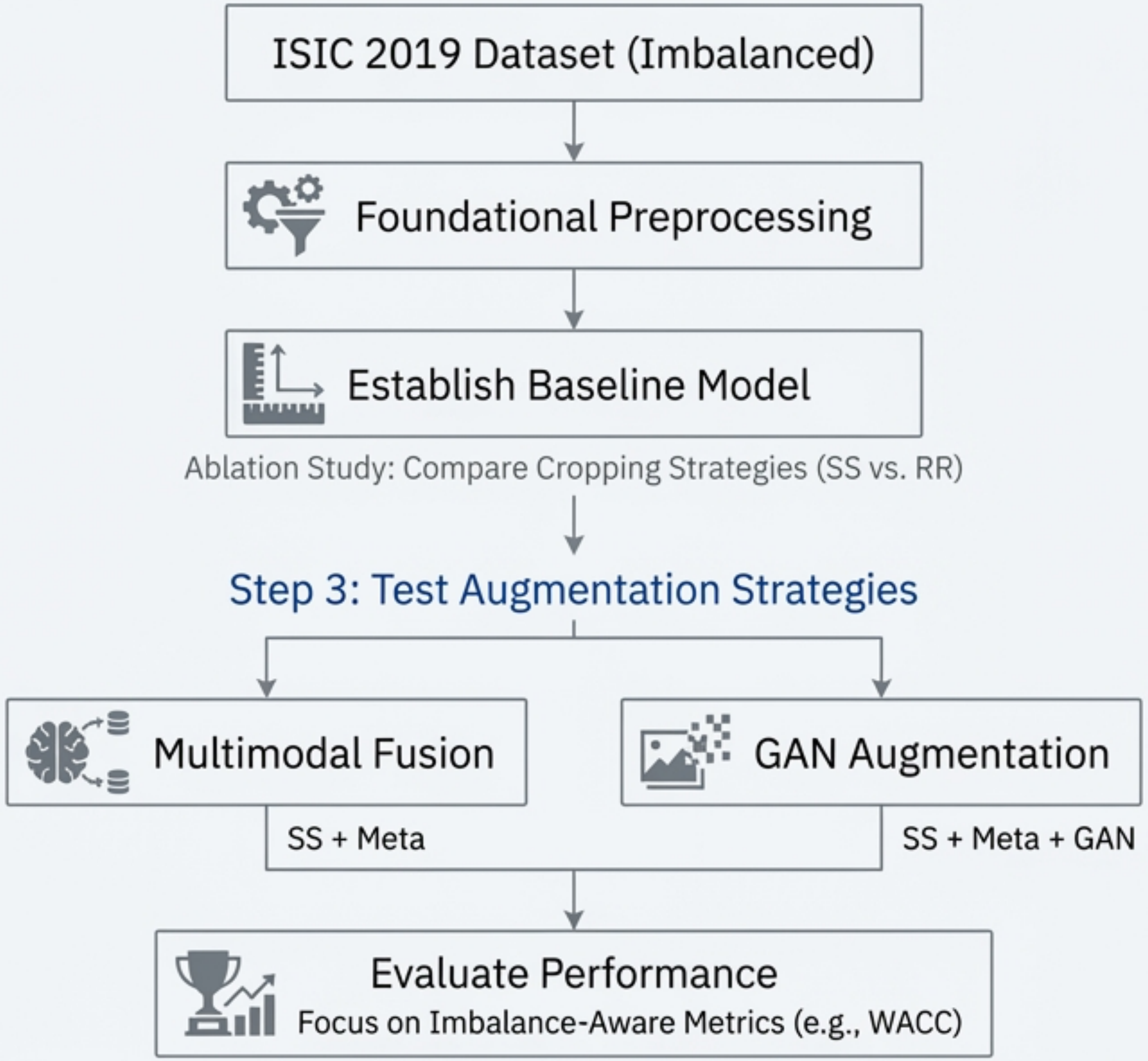
Du Yuxi 2330026036

The Challenge: A Crisis of Imbalance in Medical AI

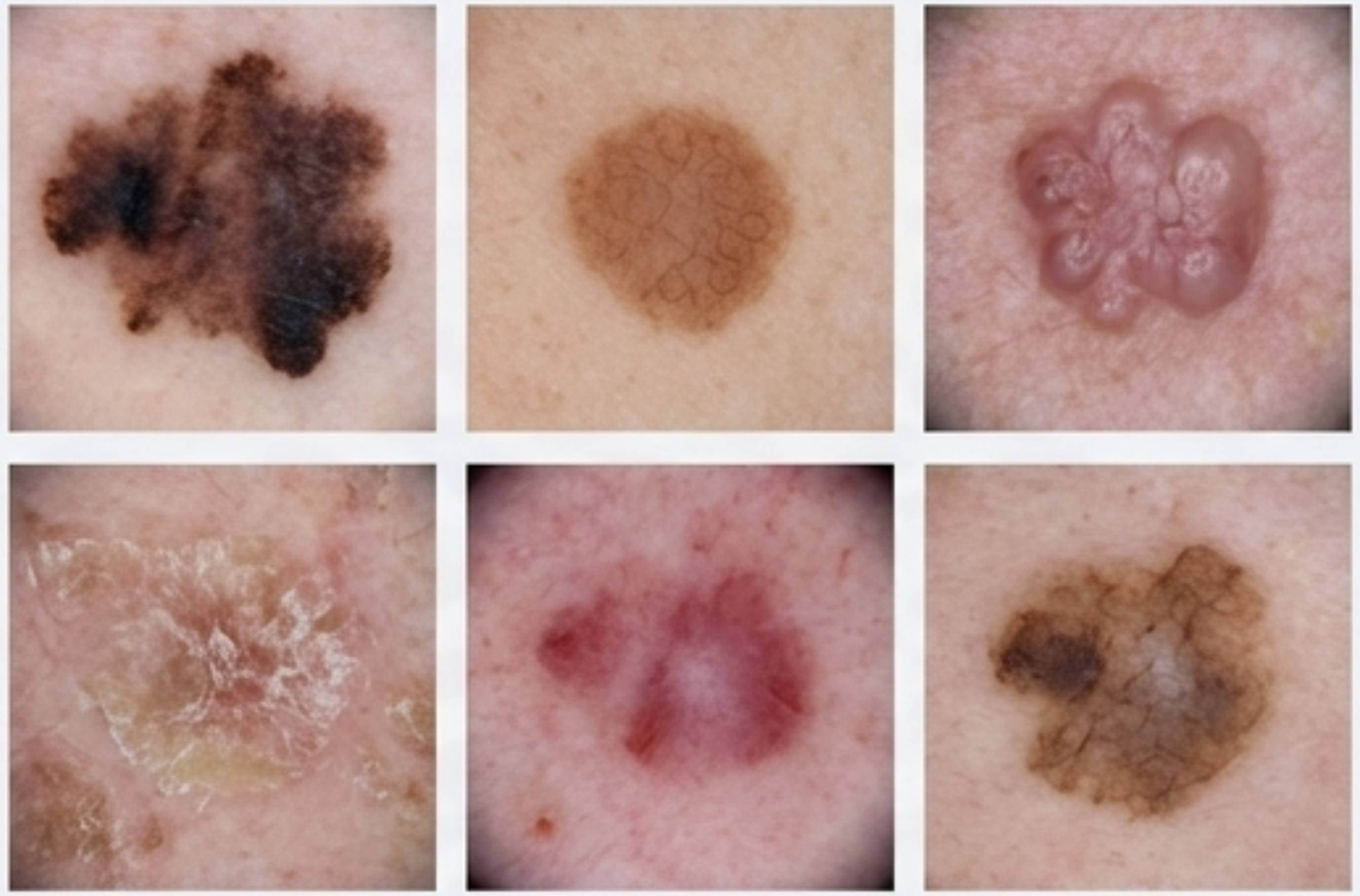
- **Problem:** In medical imaging, models are biased towards common diseases due to severe class imbalance.
- **Impact:** This leads to poor detection of rare but clinically significant conditions, undermining reliability.
- **Our Goal:** Systematically evaluate strategies to improve minority-class recognition for skin lesion diagnosis.



Our Approach: A Systematic Investigation



The Dataset: Images, Labels, and Clinical Context



25,331 Dermoscopy Images

- MEL (Melanoma)
- NV (Melanocytic nevus)
- BCC (Basal cell carcinoma)
- AK (Actinic keratosis)
- BKL (Benign keratosis)
- DF (Dermatofibroma)
- VASC (Vascular lesion)
- SCC (Squamous cell carcinoma)
- UNK (Unknown)

9 Diagnostic Classes

Age
(Numeric)

Sex
(Categorical)

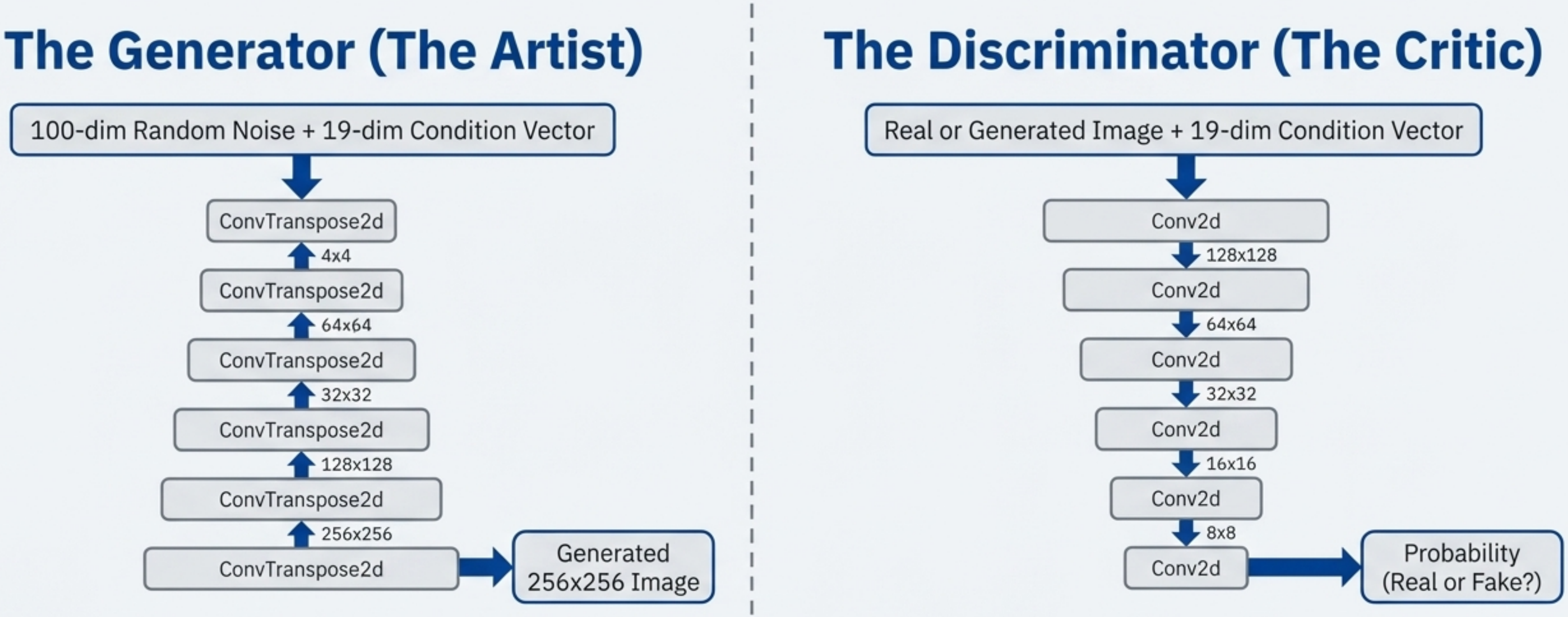
Anatomical Site
(Categorical)

Rich Clinical Metadata

GAN Improvement - cGAN

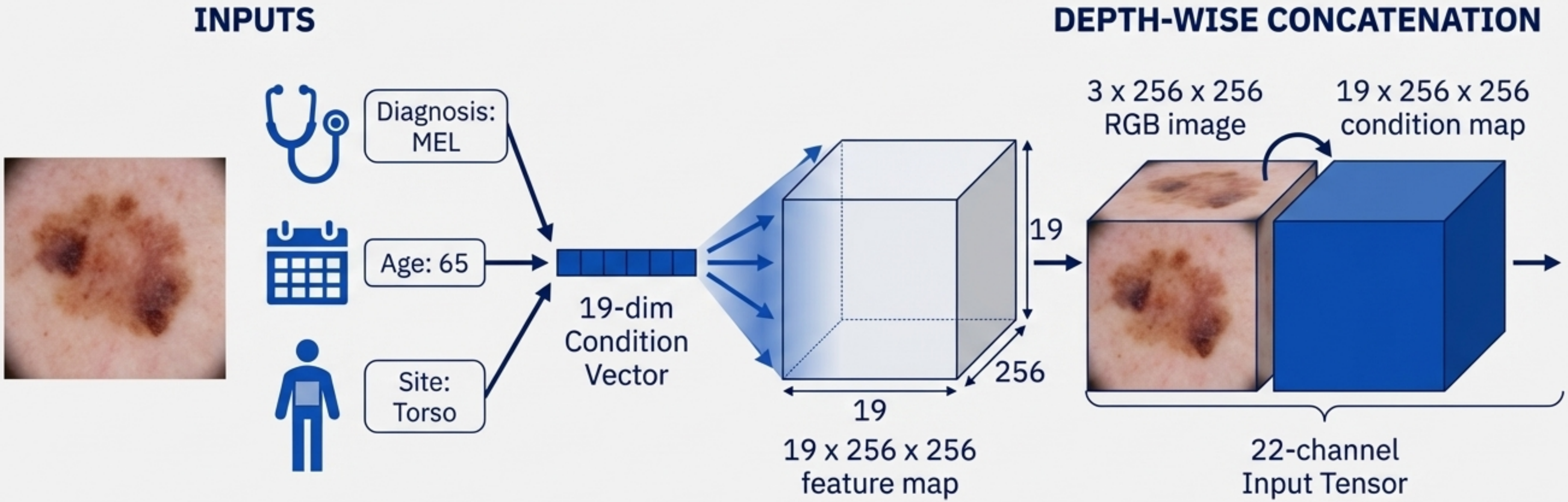
Hou Shuoran 2330026054

How the cGAN Learns



The Generator creates images for a specific condition, while the Discriminator learns to spot fakes, conditioned on the same clinical context.

Conditioning on Clinical Reality



This forces the Discriminator to evaluate image realism in the context of specified clinical conditions, not just pixel patterns.

Image classification model

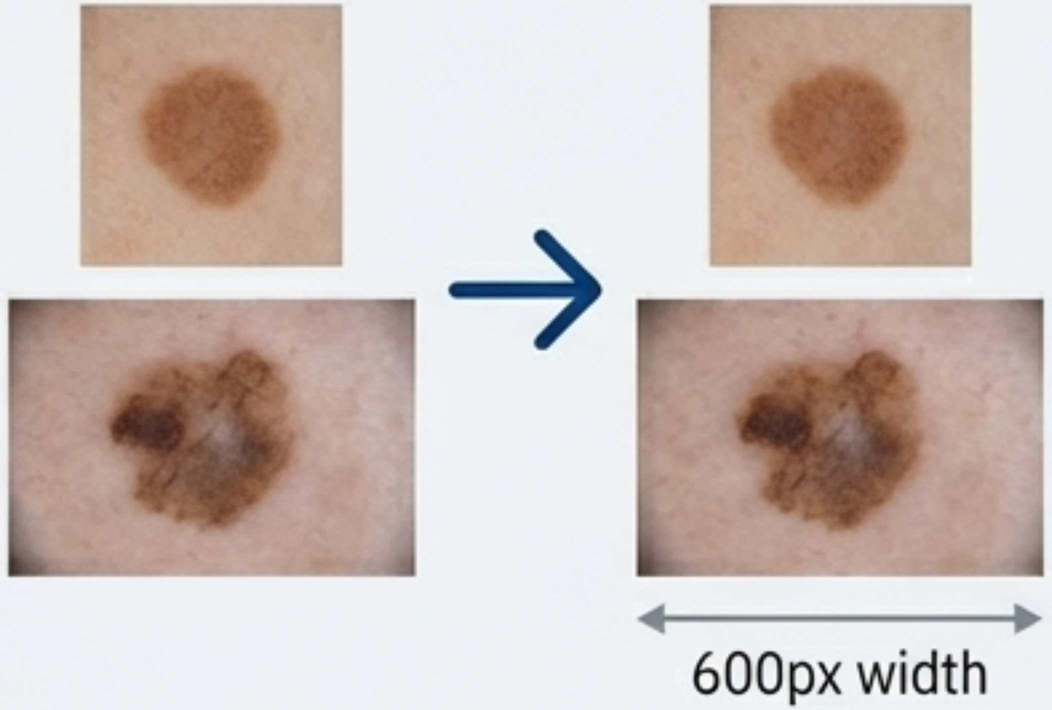
Lu Yunxiao 2330026114

Standardizing the Battlefield: Image Preprocessing

Black-Border Cropping

Suppress background cues.

Geometric Standardization



Normalize scale and orientation.

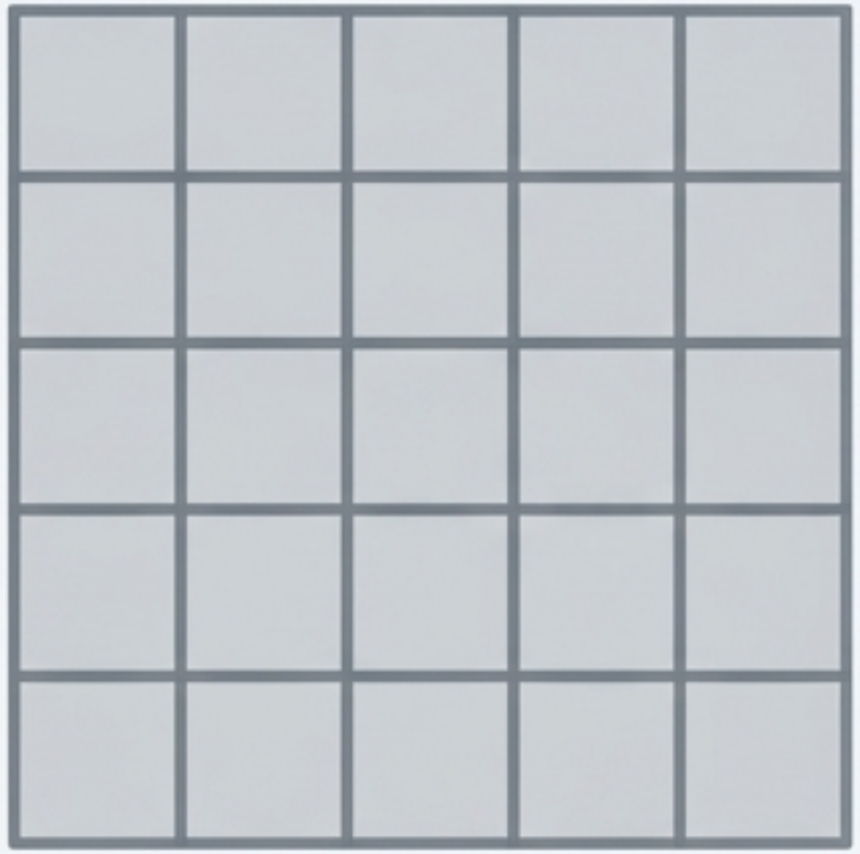
Color Constancy

Correct for device-induced color shifts.

Ensuring a fair comparison by reducing background noise and device-induced variations.

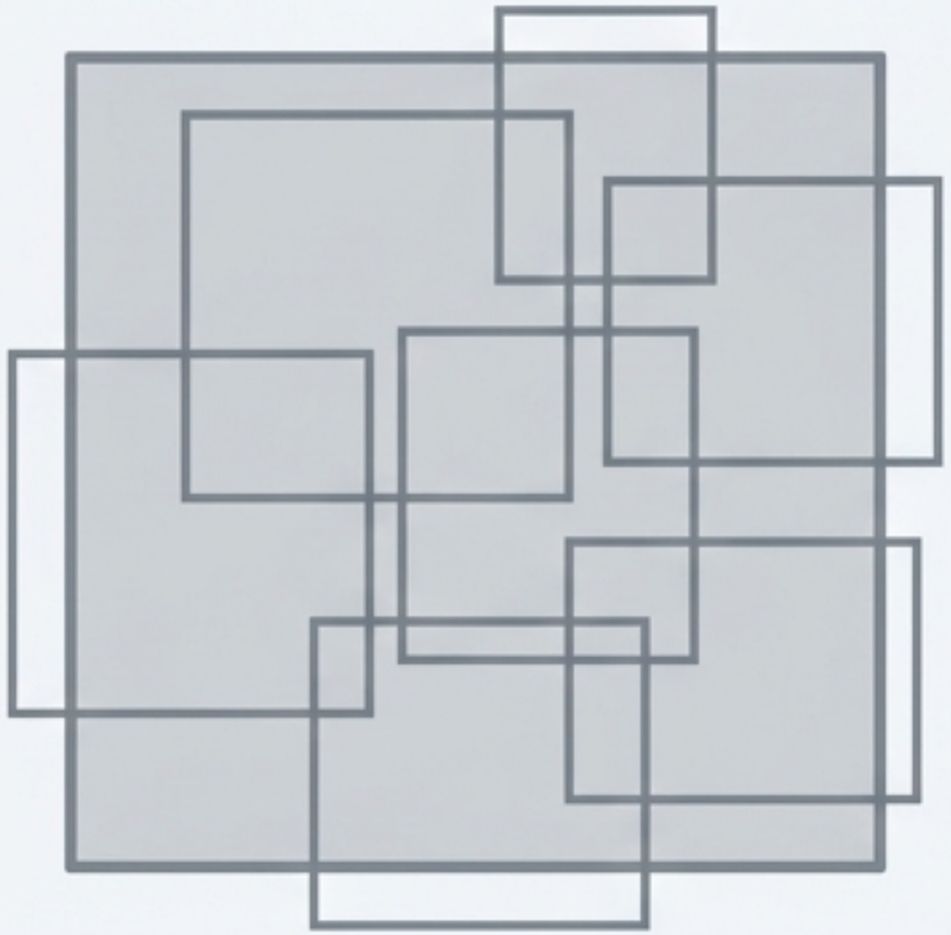
Establishing a Baseline: SS vs. RR Cropping

Same-Sized (SS) Cropping



Key Idea: Emphasizes **Spatial Coverage** to reduce sensitivity to lesion location.

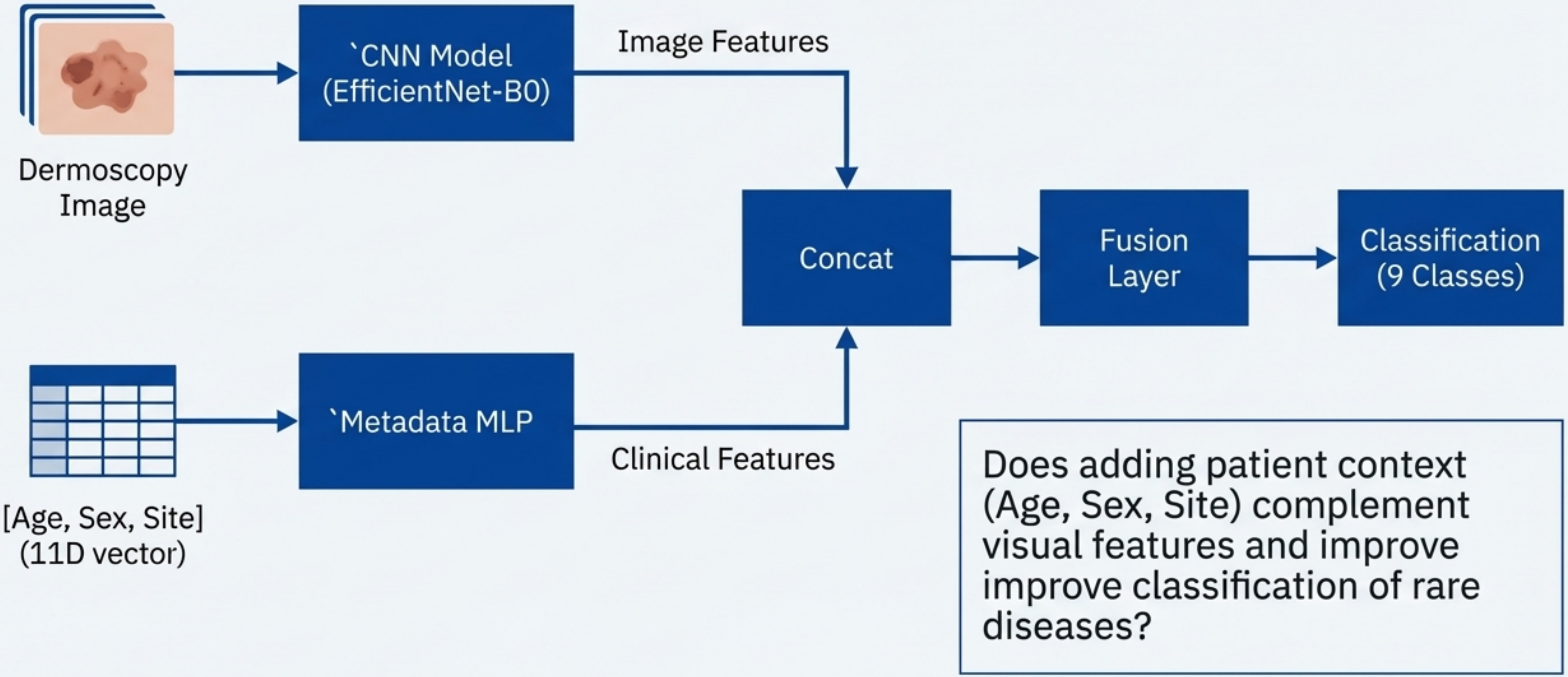
Random-Resize (RR) Cropping



Key Idea: Emphasizes **Scale Robustness** to handle lesion size variation.

Both strategies aligned to a 25-view inference budget. **Result: SS achieved a higher WACC (0.6050 vs. 0.5959) and was chosen as the baseline for all subsequent experiments.**

Hypothesis 1: Fusing Clinical Metadata



Hypothesis 2: Synthesizing Minority Class Data

Generation Flow

Methodological Control

Training vs. Validation

Strictly avoiding evaluation bias by keeping the validation set real-only.

The Moment of Truth: Results

Config	WACC	Macro-F1	Mean AUC
SS (Baseline)	0.6050	0.6414	0.9067
RR	0.5959	[illegible]	0.9081
SS + Meta	**0.6810**	**0.7007**	**0.9326**
SS + Meta + cGAN	**0.6416**	**0.6585**	**0.9177**

+12.6% Relative Gain.
Hypothesis 1 Validated: Metadata provides a powerful boost.

-5.8% Relative Drop.
Hypothesis 2 Refuted: GAN augmentation hurt performance.

Discussion: The Surprising Failure of GAN Augmentation

Distribution Mismatch

The model may have learned “GAN artifacts” (unnatural textures, edges) instead of real pathology, which failed to generalize to the real validation set.

Imperfect Semantics

Despite conditioning, synthetic images might lack subtle, true-to-life pathological details, effectively adding “label noise” to the training process.

Distribution Shift

Injecting 5,000 synthetic samples may have shifted the training distribution too far from the real data, optimizing a decision boundary that is not effective for real-only test data.

Conclusion & Future Directions

Key Takeaways

1. ✅ **Context is King:** Multimodal fusion with clinical metadata is a highly effective and data-efficient strategy for this task.
2. ❌ **Quality over Quantity:** Naive GAN augmentation can be harmful. The synthetic-real domain gap is a major hurdle that degrades generalization.
3. ✅ **Rigor Matters:** A systematic, ablation-style approach is crucial for isolating the true impact of each component.

Future Work

- ➡️ **Ensemble Models:** Combine models across folds or initializations to reduce variance and improve stability.
- ➡️ **Synthetic Quality Control:** Filter cGAN samples using discriminator scores or an artifact detector before training to verify if image quality was the root cause of the performance drop.

Thank You

Questions?

Du Yuxi (2330026036)
Hou Shuoran (2330026054)
Lu Yunxiao (2330026114)