UMBC

Toward Trustworthy AI for Medical Imaging

ResNet-Based Diagnosis with Grad-CAM++ Explanations

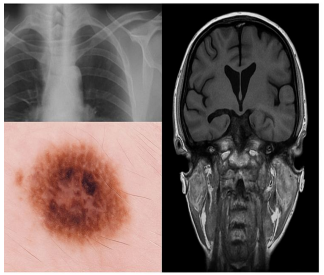
Presented by

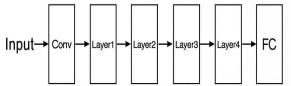
Aryan Jagani & Vishvakumar Patel

Motivation & Significance

Why Trustworthy AI in Healthcare?

- Skin cancer is one of the most common cancers worldwide.
- Early detection greatly improves outcomes.
- Deep learning models achieve high accuracy, but they act as black boxes.
- Clinicians need to know why a model made a prediction.
- Explainable AI (XAI) builds trust, improves understanding, and supports safer clinical decisions.





Key Significance:

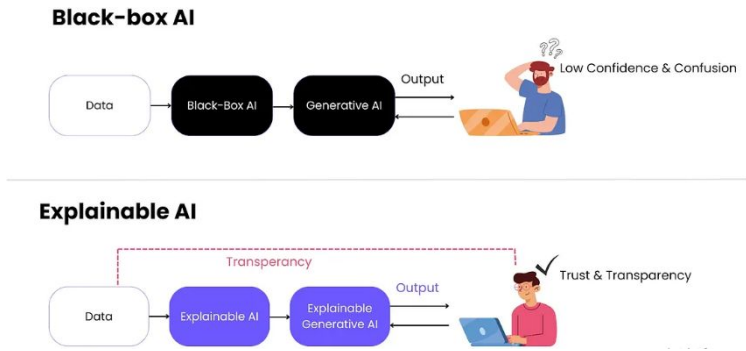
Interpretability = trust → adoption → safer clinical AI

Problem Statement & Hypothesis

Problem Statement

Medical imaging models such as ResNet-50 perform well, but:

- They offer **no intrinsic interpretability**.
- Explanations like Grad-CAM are often **coarse and qualitative**.
- Faithfulness of explanations is rarely **quantitatively evaluated**.
- Model confidence is often **poorly calibrated**, which can be dangerous in clinical settings.



Hypothesis

Grad-CAM++ will:

- Produce more **fine-grained** and **clinically aligned** heatmaps.
- Show **higher overlap** with ground-truth pathology masks.
- Improve clinician trust when paired with **calibrated confidence scores**.

Related Works & Contributions

Related Works

- **Grad-CAM (Selvaraju et al.)**: class-discriminative heatmaps, but low spatial precision.
- **Grad-CAM++ (Chattopadhay et al.)**: uses higher-order gradients → sharper localization.
- **Xie et al. (2023)**: emphasize quantitative explanation metrics (IoU, pointing-game).
- **Ihongbe et al. (2024)**: applied XAI to pneumonia detection but lacked quantitative validation.

Our Contributions

- Implement and compare **Grad-CAM**, **Grad-CAM++**, **Score-CAM**.
- Perform **quantitative evaluation** using IoU, pointing-game, deletion curves.

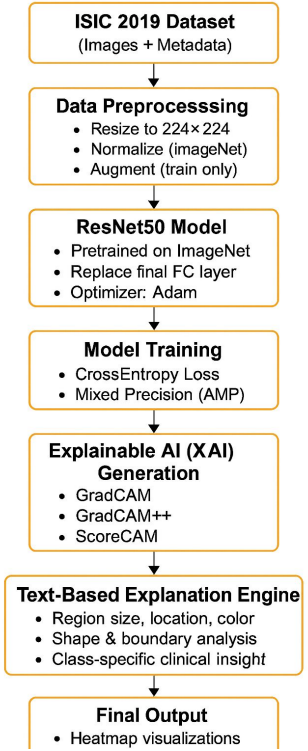
Methodology (Modeling & Training Pipeline)

Modeling Pipeline

- Load ISIC dataset
- Preprocess images (resize, normalize, augment)
- Train ResNet50 using Adam optimizer
- Predict lesion class
- Generate XAI heatmaps (GradCAM / GradCAM++ / ScoreCAM)
- Generate textual explanations

Explainability Modules

- **Grad-CAM**: gradient × activation method.
- **Grad-CAM++**: higher-order gradients for multi-instance localization.
- **Score-CAM** (planned): perturbation-based, gradient-free.



Dataset:ISIC 2019

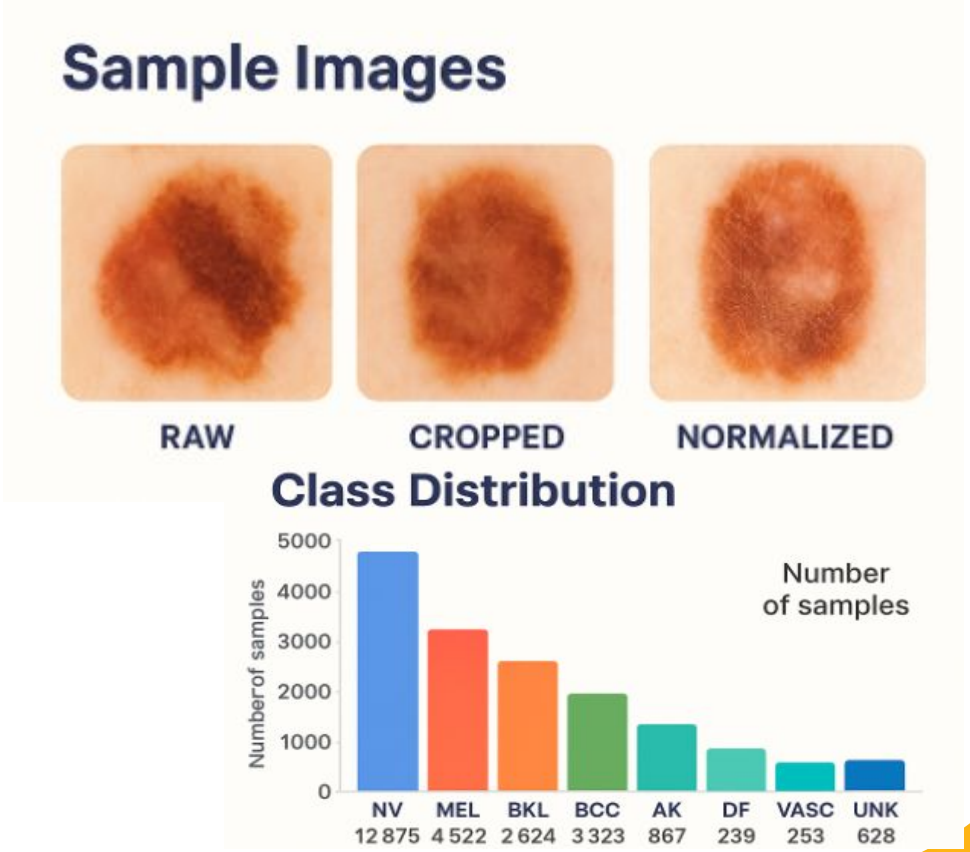
- Public dermatology dataset containing 25,331 dermoscopic images.
- 9 classes: MEL, NV, BCC, AK, BKL, DF, VASC, SCC, UNK.
- Includes metadata: age, sex, anatomical site.
- High variability in lesion appearance, lighting, and skin tone.
- Class imbalance: NV and BKL dominate; MEL and SCC are rare.

Merged DF: (25331, 14)

	image	age_approx	anatom_site_general	lesion_id	sex	MEL	NV	BCC	AK	BKL	DF	VASC	SCC	UNK	class_name	label
0	ISIC_0000000	55.0	anterior torso	NaN	female	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	NV	1
1	ISIC_0000001	30.0	anterior torso	NaN	female	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	NV	1
2	ISIC_0000002	60.0	upper extremity	NaN	female	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	MEL	0
3	ISIC_0000003	30.0	upper extremity	NaN	male	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	NV	1
4	ISIC_0000004	80.0	posterior torso	NaN	male	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	MEL	0

Experimental Setup & Datasets

- **Train/Validation split:** 80% / 20% (stratified).
- **Image size:** 224 × 224 pixels.
- **Optimizer:** Adam (lr = 1e-4).
- **Loss function:** CrossEntropyLoss.
- **Batch size:** 64.
- **Training environment:** Google Colab GPU.
- **Mixed precision (AMP)** used to speed up training.



CAM Method Comparison Across Layers

Shallow Layer (Layer1)

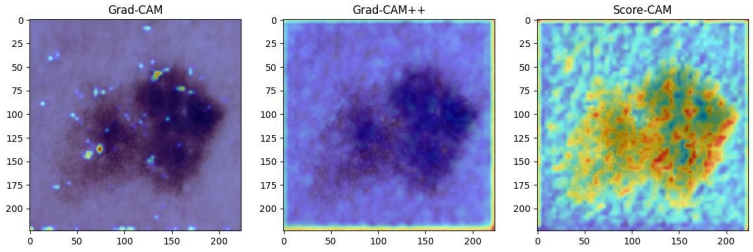
- Highlights low-level features such as edges and textures
- Explanations are broad and not lesion-specific

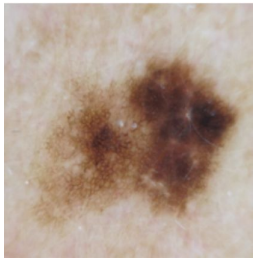
Middle Layer (Layer3)

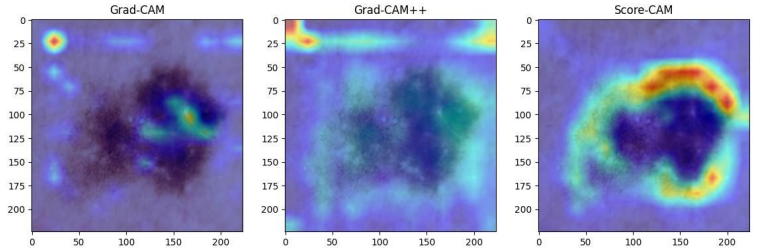
- Captures more meaningful lesion structure
- Heatmaps begin to align with lesion boundaries

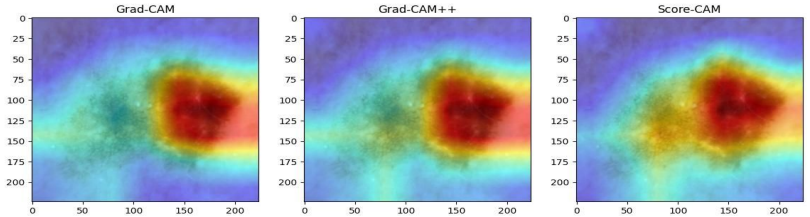
Deep Layer (Layer4)

- Produces the most clinically relevant explanations
- Strong localization around the lesion region









Results & Discussion (Qualitative XAI Output)

GradCAM: Highlights key regions that influence the prediction.

GradCAM++: Produces sharper, more precise attention maps.

ScoreCAM: Gradient-free method that generates smooth attention heatmaps.

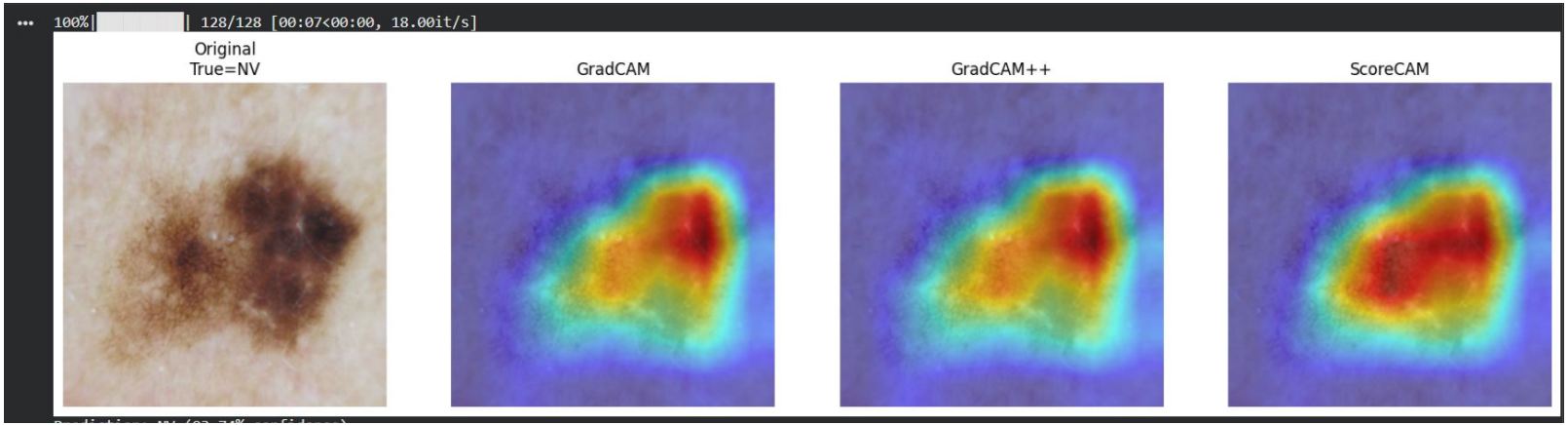
Text explanations describe:

- Region size
- Location (upper/lower, left/right)
- Shape (round/elongated)
- Color intensity
- Boundary (smooth/irregular)

```
Prediction: NV (93.74% confidence)
Focus: medium, elongated region in the lower-right part of the image.
Clinical Insight: Nevus is benign and symmetrical.
```

Results & Discussion (Quantitative)

- Training accuracy: **81%**
- Validation accuracy: **82%**
- Performs best on common classes (NV, BKL).
- Struggles with rare classes (MEL, SCC) due to imbalance.
- XAI visualizations mostly align with lesion locations.
- Occasional background activation suggests room for improvement.



Conclusions, Future Work

Conclusions

- ResNet50 accurately classifies 9 skin lesion types from the ISIC 2019 dataset.
- GradCAM, GradCAM++, and ScoreCAM provide meaningful visual explanations of model focus, with GradCAM++ offering sharper and more localized heatmaps.
- The text-based explanation engine enhances interpretability by describing lesion region size, location, color, and shape.
- Explore more advanced architectures like EfficientNet or Vision Transformer

Future Work

- Improve performance on minority classes using class balancing techniques.
- Build an interactive interface (e.g., Streamlit) for model and XAI visualization.

Team Contributions

Team Contributions

Aryan:

- Model architecture, Grad-CAM++, calibration
- Literature review & method comparisons

Vishva:

- Preprocessing, segmentation alignment, explainability metrics
- Robustness analysis

Shared:

- Documentation, proposal, analysis, presentation