

# An Interpretable Hybrid SIFT-CNN Framework for Wound Image Classification

Namrata Surve<sup>1</sup> | Luyi Li<sup>1</sup> | Dr. Jordon Gilmore<sup>1</sup> | Dr. Ivan Dungan<sup>2</sup>  
<sup>1</sup>Clemson University, Clemson, SC | <sup>2</sup>Francis Marion University, Florence, SC

## BACKGROUND

Wound image datasets are small and highly variable, making reliable deep learning classification challenging.

Diabetic ulcers affect 15–25% of diabetic patients and are a leading cause of amputation when misdiagnosed.

Manual clinical assessment is subjective and inconsistent. Automated classification offers scalable, consistent wound type detection.

SHAP provides feature-level attribution that complements spatial attention maps for clinical interpretability

## METHODS

### Dataset

- 621 images, AZH Wound Dataset - Diabetic: 139 | Non-Diabetic: 482
- Task: Diabetic vs Non-Diabetic (binary)

### CNN Branch

- ResNet18, ImageNet pretrained -Weighted CrossEntropy, 5-fold CV

### SIFT Branch

- Keypoint descriptors + K-means clustering
- 100 visual clusters

### Interpretability

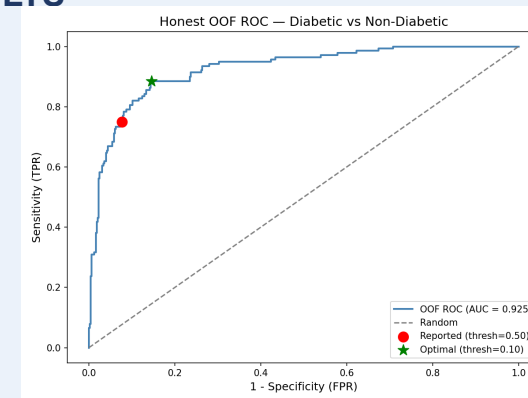
- Grad-CAM + SHAP
- Planned: SIFT-CNN feature fusion (future work)

## RESULTS

- Optimized Threshold (0.25)
- Sensitivity: 0.820 (+6.5%).
- Specificity: 0.898

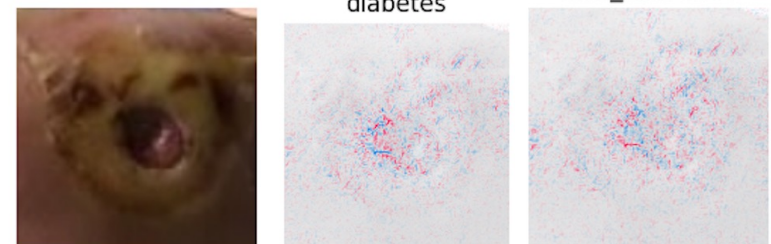
	CNN	SIFT
Accuracy	0.88	0.61
Sensitivity	0.75	0.42
Specificity	0.92	0.78

CNN significantly outperforms SIFT, particularly in diabetic detection (sensitivity).



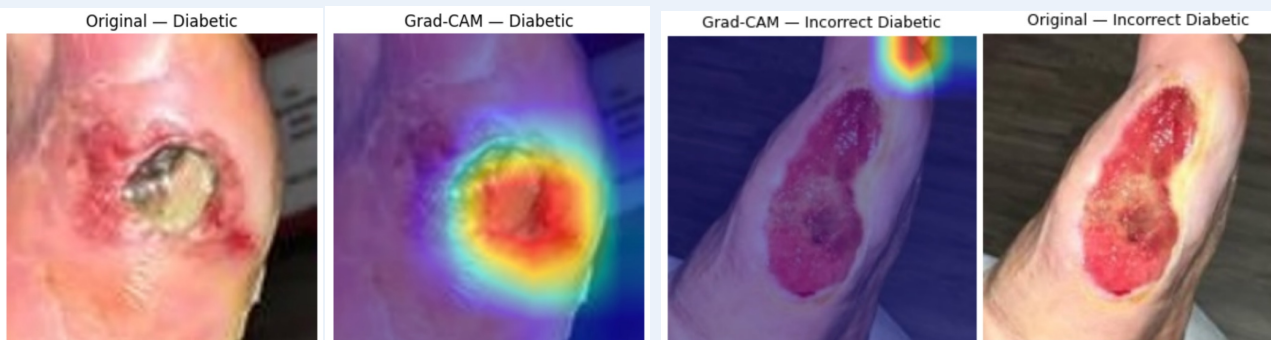
## SHAP ANALYSIS

diabetes non\_diabetes



Diffuse feature contributions indicate weak and inconsistent feature importance across images. SHAP operates at the pixel level for CNN models, making it difficult to derive structured, human-interpretable feature importance distributions (e.g., violin plots).

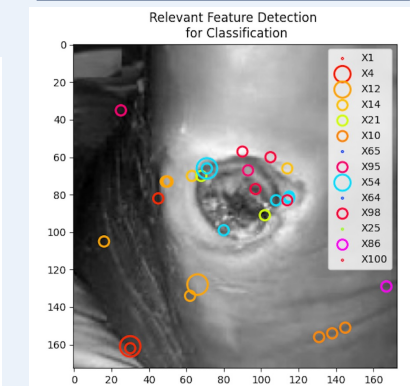
## GRAD-CAM ANALYSIS



Correct → focuses on wound

Misclassified → focuses on background

## SIFT ANALYSIS



SIFT features capture local texture but show inconsistent relevance to wound regions.

## CONCLUSIONS

- ResNet18 achieves AUC 0.925 on diabetic vs non-diabetic wound classification
- Threshold calibration (0.50→0.25) recovers 6.5% sensitivity at minimal specificity cost
- Both CNN and SIFT models exhibit inconsistent feature usage, which leads to reduced sensitivity in diabetic wound detection.
- This limitation motivates the use of SIFT-based features, which provide more structured and interpretable feature representations.

## FUTURE WORK

- Complete SIFT-CNN feature fusion
- ROI-guided attention mechanisms
- Expand the dataset beyond 621 images
- External dataset validation

## ACKNOWLEDGEMENTS

[1] This work was supported primarily by the National Science Foundation EPSCoR Program under NSF Award #OIA-2242812. Any Opinions, findings and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect those of the National Science Foundation