Autonomous Multi-modality Burn Wound Characterization using Artificial Intelligence

Maxwell Jacobson[1], Daniela Arrubla[3], Maria Romeo Tricas[1], Mohammed El Masry[2], Surya Gnyawali[2], Gayle Gordillo[2], Yexiang Xue[1], Chandan Sen[2], Juan Wachs[1]

1. Purdue University, West Lafayette, Indiana 2. Indiana University, Bloomington, Indiana 3. Emory University, Atlanta, Georgia

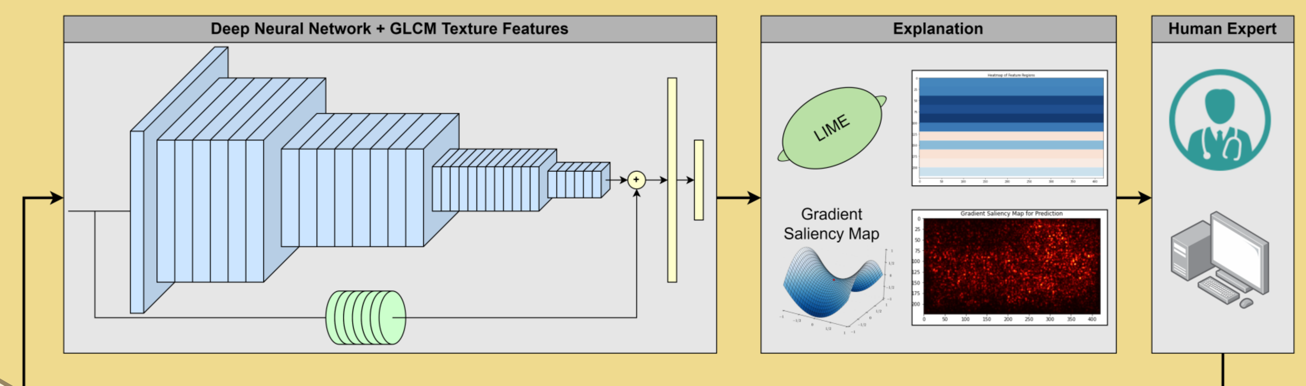
Project Website: https://www.cs.purdue.edu/homes/jacobs57/ambush/

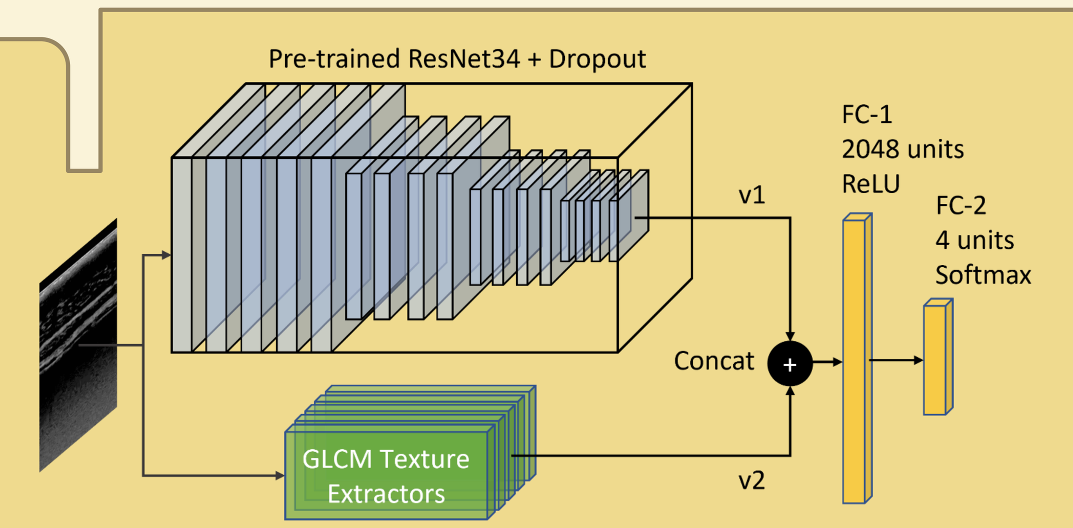
Summary

- We describe an autonomous Artificial Intelligence (AI) system to analyze burns using multiple modalities including ultrasound and RGB images.
- We assess the system's ability to predict burn depth and relative surface area.
- Classification is implemented as a deep convolutional network that makes use of GLCM texture features[4].
- Segmentation for burn area prediction is accomplished using a modified U-net convolutional autoencoder.
- Classification results are further visualized and explained via a LIME-based[2] Explainable AI (XAI) subsystem. This information is used to improve the system's accuracy and reliability.

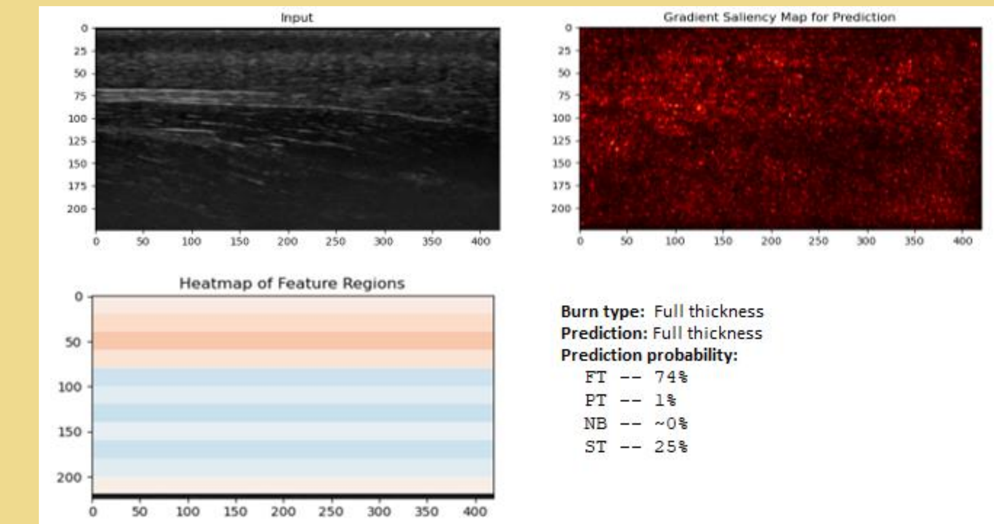
Burn Classification & Explanation

- We use convolutional neural networks (CNNs) to learn progressively richer features from the data and then use these features to train task-specific networks.
- Pre-trained models are used as a base, and then finetuned to our task. This reduces the number of training samples needed.
- We further enhance our classifier using traditional computer vision features (GLCM texture) that have been shown to be effective on ultrasound data.
- We make use of a human-in-the-loop system that utilizes explainable AI to improve our prediction models and verify our results.



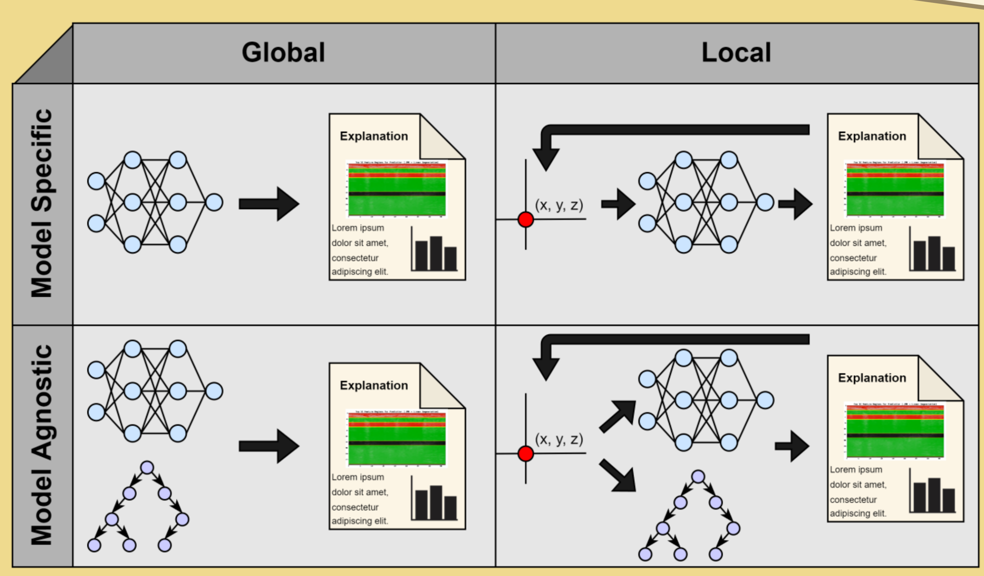


Our classification model. This makes use of a pre-trained ResNet34 component modified with dropout to reduce overfitting to our small dataset. GLCM texture features are extracted and included as features in the final stage of the classifier.



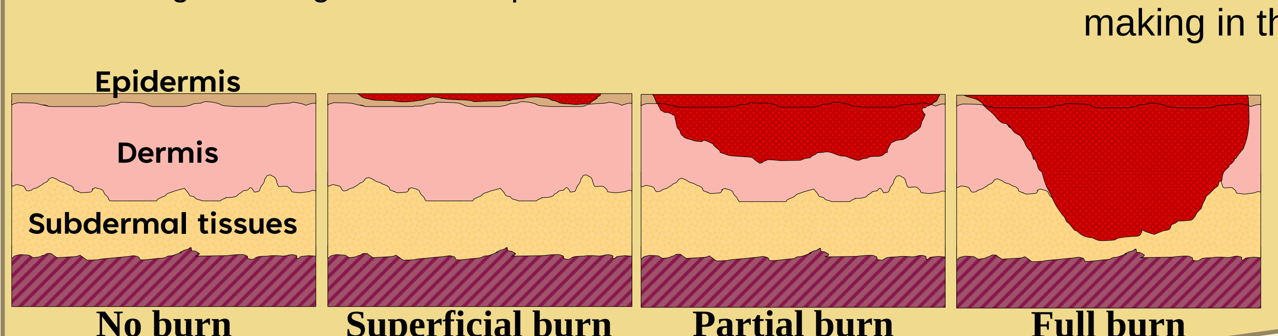
Output information after classification and explanation. Blue features in heatmap indicate support for prediction. Red features indicate contradiction of prediction.

Background



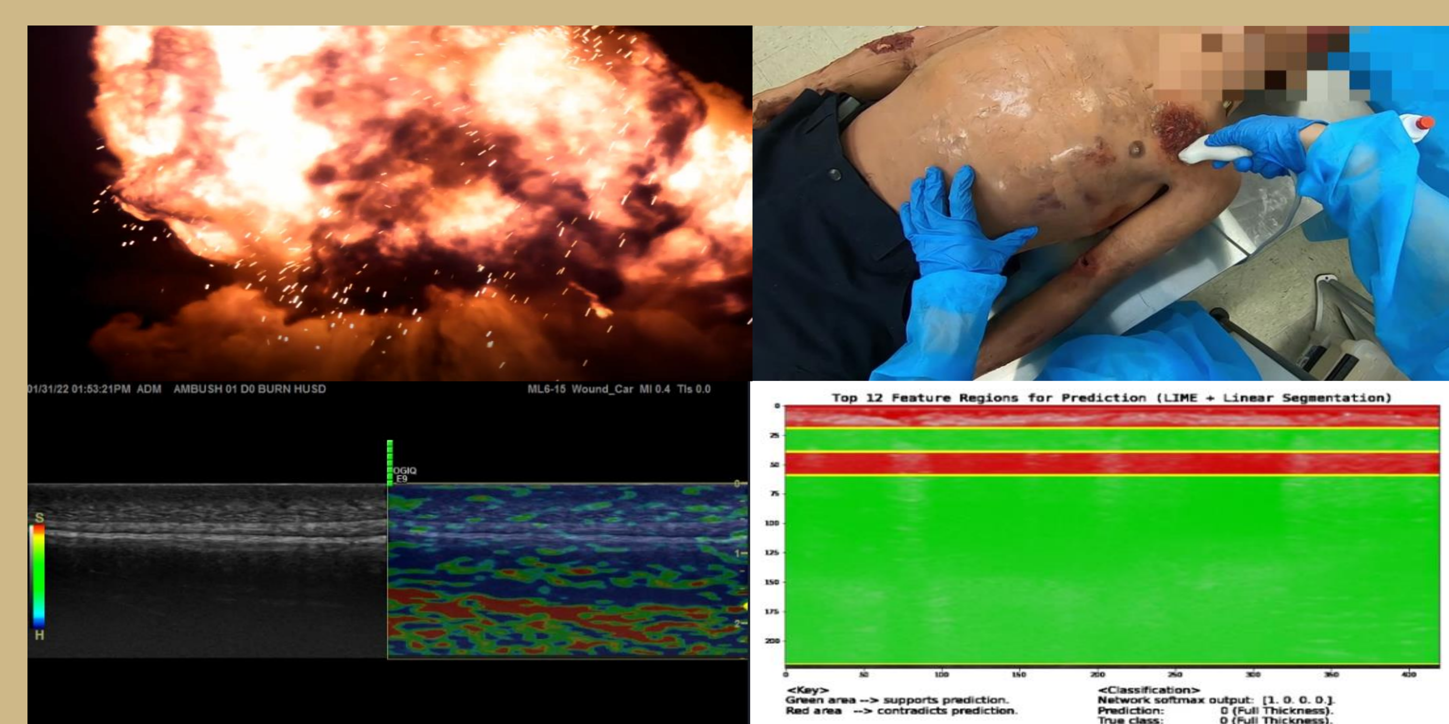
XAI algorithms can be split into global and local – explaining a whole model or a single data point prediction – or into model/algorithm agnostic and specific.

- The outcome for any burn patient is determined greatly by the percent total body surface area (%TBSA) and the depth of the burn.
- In the military setting, burn casualties correspond to roughly 8% of combat-related injuries.
- Four classes of burn depth determine the treatment options for the patient.
- Explainable AI (XAI) is often used to understand and validate AI decision-making in the medical domain.



Explainable AI

Deep Learning

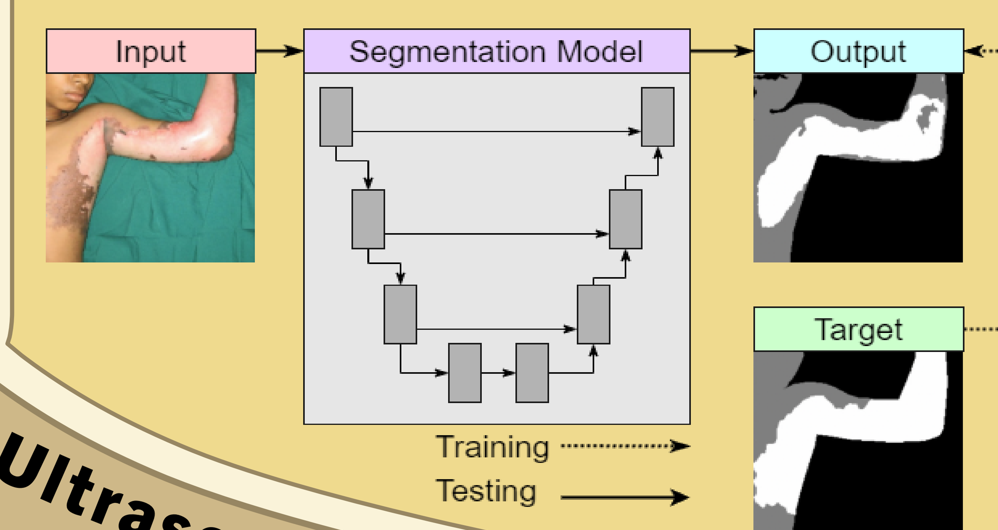


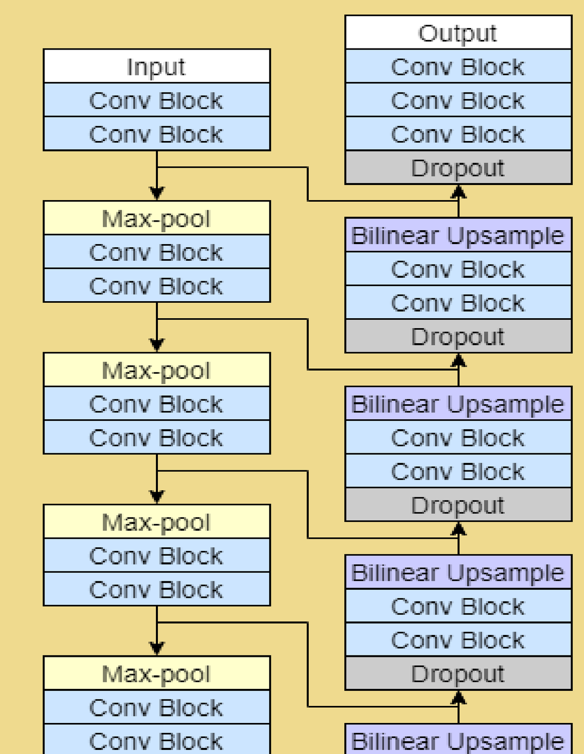
Computer Vision

Ultrasound

Burn Segmentation

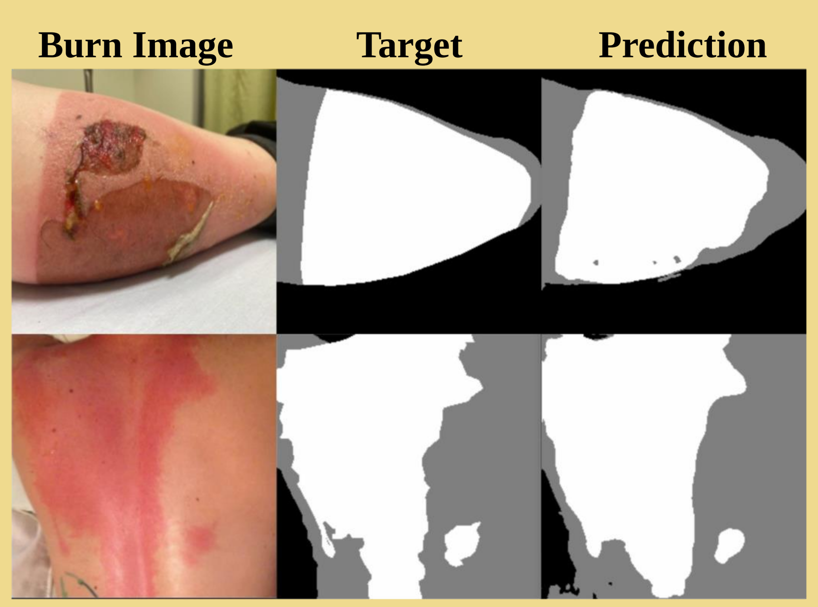
- To predict the surface area of the burn, we must segment the body and the burned region.
- We used a modified U-net[3] to learn burn segmentation.
- The general U-net architecture has shown consistently good results on medical image segmentation tasks[1].
- Batch normalization and dropout operations were added to reduce the risk of overfitting given our small dataset.





Our segmentation U-net model. Each convolutional block is composed of a 2D convolutional layer, a batch normalization, and a ReLU activation. The final block omits the normalization and activation.

Results



Some segmentation results. The U-net prediction closely matches the ground truth target. This can be used to calculate the burn area of the displayed skin.

- Classify burn wounds with a mean accuracy greater than 90%.
- Segment burn wounds with a mean global accuracy greater than 0.87, and a mean intersection-over-union (IoU) score of ~0.

Segmentation Results

Metric	Stock U-Net (Mean, Stddev)		Custom U-Net (Mean, Stddev)	
Global Accuracy	0.783	0.033	<u>0.872</u>	0.008
IoU	0.664	0.041	<u>0.784</u>	0.012

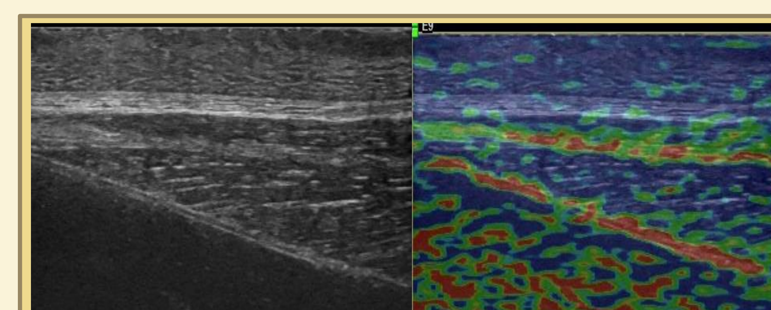
Classification Results at 3 Time Points

Days	Accuracy	Precision	Recall	F1 Score
Day 0	0.9379	0.9474	0.9403	0.9414
Day 3	0.8821	0.9072	0.8714	0.8787
Day 7	0.9497	0.9466	0.9412	0.9397
Total	0.9252	0.9373	0.9172	0.9235

Classification Results Algorithm Comparison

Method	Features (img, tex)		Accuracy	Precision	Recall	F1 Score
RBF SVM	✗	✓	0.4830	0.5433	0.4654	0.4745
VGG16	✓	✗	0.8898	0.9217	0.8790	0.8837
ResNet34	✓	✗	0.8963	0.9219	0.8860	0.8933
Ours	✓	✓	<u>0.9257</u>	<u>0.9374</u>	<u>0.9182</u>	<u>0.9242</u>

Data



One frame of B-mode ultrasound data (left) and the corresponding TDI data (right).

- In-vivo B-mode ultrasound and Tissue Doppler Imaging (TDI) scans collected from porcine subjects.
- The GE Logiq E9 device was used to generate HUSD B-mode and Tissue Doppler elastography Imaging videos simultaneously.
- RGB images for segmentation were manually collected from Google Images, and hand-annotated with the supervision of clinicians.

References

[1] N. S. Punn and S. Agarwal, "Modality specific u-net variants for biomedical image segmentation: a survey," AI Review, pp. 1–45, 2022.
[2] M. T. Ribeiro, S. Singh, and C. Guestrin, "'why should i trust you?' explaining the predictions of any classifier," in Proceedings of the 22nd ACM SIGKDD, 2016, pp. 1135–1144.
[3] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in MICCAI. Springer, 2015, pp. 234–241.
[4] R. M. Haralick, K. Shanmugam, and I. H. Dinstein, "Textural features for image classification," IEEE Transactions on SMC, no. 6, pp. 610–621, 1973.

Acknowledgments

This work was supported by the Office of the Assistant Secretary of Defense for Health Affairs under Award No. 6W81XWH-21-2-0030 and by the National Science Foundation under Grant NSF #2140612.

Opinions, interpretations, conclusions and recommendations are those of the authors and are not necessarily endorsed by the Department of Defense.