Cureus Journal of Computer Science
https://doi.org/10.7759/s44389-026-00077-z

Original Article

Uncertainty-Aware Wound Localization and First-Aid Decision Support From Smartphone Images Using an Explainable Multimodal AI Framework

Sumit Barua [1], Ruth Bahre [2], Radu Babiceanu [2], Guan Y. Hong [1]

1. *Department of Computer Science, Western Michigan University, Kalamazoo, USA*
2. *Department of Electrical and Computer Engineering, Western Michigan University, Kalamazoo, USA*

Received: March 23, 2026 | Review began: March 29, 2026 | Review ended: May 07, 2026 | Published: May 18, 2026

Abstract

Purpose

Automated wound assessment from unconstrained smartphone images remains challenging due to limited annotated data, heterogeneous imaging conditions, and the lack of trustworthy AI mechanisms such as uncertainty handling and interpretable decision support. This study evaluates the feasibility of a safety-aware multimodal framework that integrates computer vision and language-based reasoning to provide structured first-aid guidance under open-world imaging conditions.

Methods

Two custom datasets were constructed from openly available images for wound segmentation and coarse anatomical region classification. The proposed pipeline integrates YOLOv11-Segmentation for wound localization, a ResNet-50 classifier with Gradient-weighted Class Activation Map visualization for anatomical context, and a lightweight offline Mistral large language model to generate structured first-aid recommendations. An uncertainty-aware routing mechanism implements selective prediction, allowing the system to abstain from automated guidance and request clarification when prediction confidence falls below a predefined threshold. Generated responses are constrained through structured prompting to avoid diagnostic claims and emphasize general first-aid principles. Confidence scores used for routing were not explicitly calibrated.

Results

On heterogeneous smartphone imagery, YOLOv11-Segmentation achieved an overall mean average precision@0.5 of 0.771 and mean average precision@0.5-0.95 of 0.685. A U-Net baseline trained on the same dataset achieved a pixel accuracy of 0.9623 and a mean intersection-over-union of 0.7004, providing contextual segmentation benchmarking. For anatomical classification, ResNet-50 achieved 82.11% accuracy and outperformed an EfficientNet-B0 baseline (74.42%). Sensitivity analysis of the routing threshold showed that increasing the confidence threshold reduced automated guidance and increased clarification requests, demonstrating how uncertainty-aware routing regulates language-model usage under ambiguous visual conditions. Evaluation of language-model outputs was limited to structured qualitative analysis due to the lack of standardized quantitative metrics for first-aid guidance.

Conclusion

These findings demonstrate the feasibility of combining wound localization, anatomical context modeling, and uncertainty-aware language generation within a transparent multimodal workflow for open-world smartphone imagery. The system is not intended for clinical use and has not been validated with clinical experts or external datasets. While the

How to cite this article:
Barua S, Bahre R, Babiceanu R, et al. (May 18, 2026) Uncertainty-Aware Wound Localization and First-Aid Decision Support From Smartphone Images Using an Explainable Multimodal AI Framework. Cureus J Comput Sci 3 : es44389-026-00077-z.
DOI https://doi.org/10.7759/s44389-026-00077-z

proposed framework illustrates a safety-oriented approach to integrating perception and controlled reasoning, further work is required to evaluate generalizability, calibrate uncertainty estimates, and establish rigorous quantitative and expert-based evaluation of generated recommendations before real-world deployment.

Categories: AI applications, Computer Vision, Mobile Health

Keywords: explainable artificial intelligence, trustworthy ai, uncertainty-aware decision support, selective prediction, smartphone-based image analysis, multimodal vision-language systems, wound segmentation, medical image analysis, first-aid decision support, deep learning

Introduction

Motivation and context

Acute and minor wounds are among the most frequently encountered injuries in both clinical and community settings and represent a common reason for healthcare utilization worldwide [1]. Although many wounds are non-severe and can be managed with basic first-aid practices, timely assessment and appropriate guidance remain challenging in settings where access to wound-care expertise is limited. These challenges are amplified by geographic barriers, healthcare workforce shortages, and the increasing adoption of telemedicine-based care models [2,3]. In many home-care scenarios, smartphone photographs serve as the primary means of documenting wound appearance and communicating wound status to caregivers or healthcare providers.

Recent advances in AI have enabled progress in medical image analysis and clinical decision support, motivating interest in image-based wound assessment [4]. However, translating these advances to real-world smartphone scenarios remains challenging due to variability in image quality, including lighting conditions, motion blur, occlusion, and heterogeneous backgrounds. These factors can reduce the reliability of models trained on curated clinical datasets when applied in open-world environments.

Addressing these challenges requires not only accurate visual perception but also mechanisms for managing uncertainty, preventing overconfident outputs, and communicating system behavior in a transparent and interpretable manner for end users [5]. These requirements align with broader efforts toward trustworthy and human-centered clinical AI, as reflected in emerging reporting and evaluation frameworks such as CONSORT-AI, SPIRIT-AI, CLAIM, and TRIPOD+AI [6-9].

To maintain ethical scope and focus on practical deployment scenarios, this study targets minor wounds such as superficial cuts and burns. These injuries are commonly captured using smartphones and typically require basic first-aid guidance rather than specialized clinical imaging or immediate medical intervention. This constraint enables the exploration of AI-assisted decision support in realistic home-care settings while avoiding diagnostic claims and reducing reliance on sensitive clinical data.

Related work

Recent advances in image-based wound assessment have demonstrated the potential of AI to support remote monitoring and telemedicine workflows. Deep learning models have been applied to wound classification and segmentation, achieving promising performance in controlled clinical or curated datasets [1-4]. In parallel, telemedicine-based approaches have improved access to wound care and enabled remote clinical decision-making [2,3]. However, most existing approaches are developed and evaluated under standardized imaging conditions, and their performance under real-world smartphone variability remains less explored.

To improve interpretability and reliability, there has been increasing focus on explainable and trustworthy AI in healthcare. Techniques such as Gradient-weighted Class Activation Map (Grad-CAM) provide visual explanations by highlighting image regions that influence model predictions, supporting interpretability and user trust [5]. Additionally,

reporting frameworks including CONSORT-AI, SPIRIT-AI, and TRIPOD emphasize transparency, reproducibility, and appropriate evaluation in AI-based clinical systems [6,7,9]. Despite these advances, limited attention has been given to integrating uncertainty-aware mechanisms and interpretable outputs into end-user decision-support workflows.

More recently, multimodal AI and large language models have been explored for clinical reasoning and decision-support applications. Multimodal frameworks that integrate visual and contextual information have shown potential for improving healthcare decision-making [10,11]. Similarly, large language models have demonstrated the ability to encode clinical knowledge and generate structured responses for healthcare-related queries [12]. However, these approaches are typically evaluated in controlled settings, with relatively limited focus on grounding outputs in uncertain real-world visual inputs or supporting safety-aware deployment in non-clinical environments.

Research gaps

Despite these advances, several limitations continue to hinder the development of practical and safety-aware wound assessment systems for real-world use.

First, many existing approaches focus on isolated tasks such as wound detection, classification, or description. In contrast, real-world decision-support applications require coordinated integration of perception, contextual reasoning, and downstream guidance within a unified workflow. Systems that treat these components independently may struggle to provide coherent end-to-end support for non-expert users [10,11].

Second, anatomical context is often overlooked in wound-analysis pipelines. First-aid guidance frequently depends on the location of an injury, yet most existing systems treat wound recognition independently of body-region information. Incorporating anatomical context may improve the relevance and safety of generated guidance but remains relatively underexplored.

Third, current explainability approaches primarily rely on visual saliency methods such as Grad-CAM. While these techniques highlight regions influencing predictions, they do not directly translate model outputs into actionable or human-understandable reasoning for end users [5]. Bridging the gap between visual explanations and practical decision support remains an open challenge.

Fourth, although large language models have shown promise for healthcare communication, their integration into wound-assessment workflows remains limited. Existing systems rarely incorporate structured prompting, uncertainty-aware routing, or safeguards designed to reduce the risk of overconfident or misleading recommendations in non-clinical settings [10,12].

Finally, many wound-analysis datasets rely on curated clinical images captured under controlled conditions. Such datasets may not reflect the variability present in smartphone-captured images used in home-care and telemedicine scenarios. As a result, models trained on controlled datasets may exhibit reduced reliability when deployed in open-world environments [3,11].

Proposed framework and contributions

To address the identified limitations, this study proposes a multimodal AI framework for wound assessment using smartphone imagery. The framework integrates wound segmentation, coarse anatomical classification, uncertainty-aware decision routing, and language-based first-aid guidance within a unified decision-support pipeline.

The primary objective of this work is to evaluate the feasibility of safety-aware multimodal reasoning under open-world smartphone conditions rather than to optimize standalone model performance. The system combines visual perception with selective prediction mechanisms that regulate downstream language generation based on prediction confidence, allowing the pipeline to defer guidance and request clarification when visual predictions are uncertain.

This study makes several contributions toward the development of safety-aware multimodal decision-support systems. First, it introduces an integrated pipeline that jointly models wound localization, anatomical context, uncertainty-aware routing, and language-based guidance, addressing the fragmentation observed in prior approaches. Second, the

How to cite this article:
Barua S, Bahre R, Babiceanu R, et al. (May 18, 2026) Uncertainty-Aware Wound Localization and First-Aid Decision Support From Smartphone Images Using an Explainable Multimodal AI Framework. Cureus J Comput Sci 3 : es44389-026-00077-z.
DOI https://doi.org/10.7759/s44389-026-00077-z

framework incorporates coarse anatomical classification to enable location-aware recommendations, an aspect that is often overlooked in existing wound-analysis systems. Third, it implements an uncertainty-aware decision-control mechanism that regulates guidance generation by abstaining under low-confidence conditions, improving safety in ambiguous scenarios. Fourth, the framework combines visual explanations through Grad-CAM with structured textual rationales to enhance interpretability for non-expert users. Finally, the study evaluates the proposed system under heterogeneous smartphone imaging conditions and provides comparative benchmarking against baseline architectures to contextualize performance.

Materials And Methods

Study design

This study evaluates a multimodal pipeline for automated wound assessment using smartphone-captured images. The system integrates computer vision models for wound localization and anatomical classification with a language model for first-aid guidance generation. A selective prediction mechanism is incorporated to regulate automated outputs under uncertainty.

Dataset description

The datasets used in this study are described in detail in the Results section. In brief, publicly available images were curated to construct datasets for wound segmentation and anatomical region classification under real-world conditions.

Wound segmentation model

A YOLOv11-based segmentation model [13] was used for wound localization and classification. The model produces pixel-level segmentation masks along with wound class predictions and confidence scores. YOLOv11 was selected for wound localization due to its ability to perform efficient instance-aware segmentation with low latency, making it suitable for real-time or near-real-time applications in resource-constrained environments such as mobile or edge-based systems.

The model was initialized with pretrained weights from the Ultralytics framework and optimized using a composite loss combining classification, localization, and mask reconstruction components. The segmentation workflow is shown in Figure *1*.

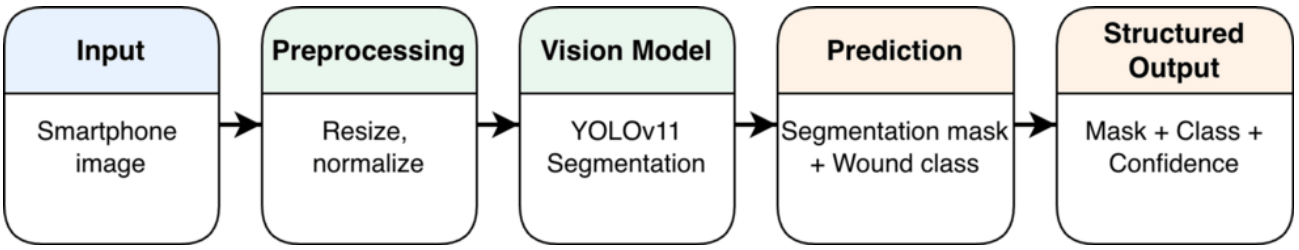

FIGURE 1: YOLOv11-based wound segmentation pipeline

A smartphone image is first preprocessed through resizing and normalization before being passed to the YOLOv11 segmentation model. The model generates a pixel-level wound mask and corresponding wound-class prediction along with a confidence score. These structured outputs are used as inputs for downstream multimodal reasoning.

Baseline segmentation model

A U-Net model with a ResNet34 encoder was trained as a baseline using the same dataset splits. U-Net was selected due to its established effectiveness in pixel-level segmentation tasks with limited data and its ability to capture fine-grained spatial features through encoder-decoder skip connections. Polygon annotations were converted to rasterized masks.

The model was optimized using pixel-wise cross-entropy loss with the Adam optimizer and early stopping. Performance was evaluated using pixel accuracy and mean Intersection-over-Union.

Although YOLOv11 performs instance-aware segmentation and U-Net is a semantic segmentation model, the comparison is intended to provide contextual benchmarking rather than direct architectural equivalence. Both models are evaluated on their ability to localize wound regions at the pixel level under identical dataset conditions. This comparison enables assessment of segmentation quality across different modeling paradigms and provides a reference point for understanding the practical performance of the proposed approach in real-world settings.

Anatomical region classification

A ResNet-50 [14] model was used for anatomical classification with four output classes. ResNet-50 was selected due to its strong generalization performance and stable training behavior in image classification tasks, particularly when initialized with ImageNet-pretrained weights. The model was initialized with ImageNet-pretrained weights and fine-tuned for coarse anatomical region classification using the training configuration described in the Experimental Setup section.

Baseline classification model

An EfficientNet-B0 model was trained using the same dataset splits and preprocessing pipeline to provide a comparative baseline. EfficientNet-B0 was selected as a lightweight architecture with strong parameter efficiency, enabling comparison between deeper residual networks and more compact convolutional designs under identical experimental conditions.

Design rationale for two-stage processing

Although the segmentation model provides wound class predictions alongside segmentation masks, the proposed framework adopts a two-stage design that separates wound localization from downstream reasoning. Direct multi-class segmentation of wound types requires highly detailed and consistently annotated datasets, which are limited in unconstrained smartphone imagery. By focusing the segmentation stage on robust localization and coarse categorization, the framework reduces sensitivity to noisy or ambiguous labels. This separation also enables independent confidence estimation and facilitates the integration of anatomical context and uncertainty-aware routing, which are critical for safety-aware decision support.

Explainability

Grad-CAM [15] was applied to the final convolutional layer of the classification model to generate attention heatmaps, highlighting spatial regions contributing to predictions. Grad-CAM was selected due to its compatibility with convolutional architectures and its ability to provide class-specific, spatially localized explanations without modifying the underlying model.

Multimodal framework integration

Outputs from the segmentation and classification models were combined into structured representations, including wound type, anatomical region, and confidence scores. These outputs were used as inputs to downstream reasoning components without direct image-level processing by the language model. The overall multimodal framework is shown in Figure *2*.

How to cite this article:
Barua S, Bahre R, Babiceanu R, et al. (May 18, 2026) Uncertainty-Aware Wound Localization and First-Aid Decision Support From Smartphone Images Using an Explainable Multimodal AI Framework. Cureus J Comput Sci 3 : es44389-026-00077-z.
DOI https://doi.org/10.7759/s44389-026-00077-z

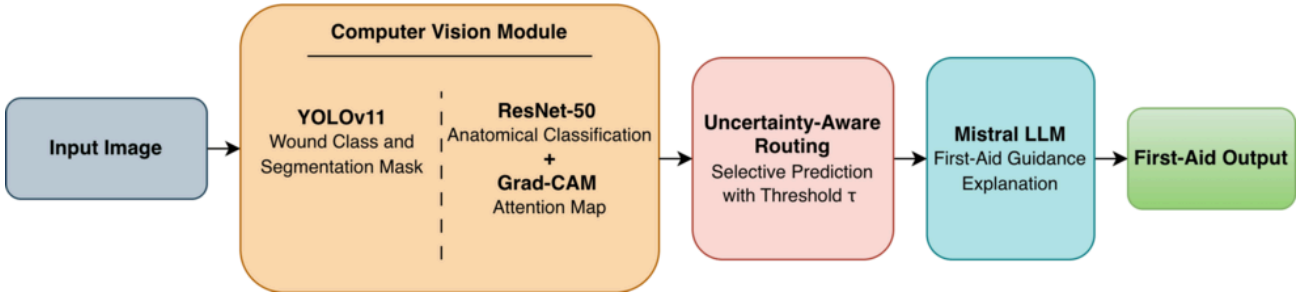

FIGURE 2: Overview of the proposed multimodal framework for uncertainty-aware wound assessment

A smartphone input image is processed by a computer vision module consisting of a YOLOv11 segmentation model for wound localization and classification and a ResNet-50 model for anatomical region classification with Grad-CAM attention cues. The resulting symbolic outputs, including predicted wound class, body region, confidence scores, and attention information, are evaluated using an uncertainty-aware selective prediction mechanism with threshold τ. When confidence is sufficient, structured representations of visual predictions are passed to a local Mistral language model for first-aid guidance generation. In cases of low confidence, the system defers automated recommendations and may request clarification. The framework integrates perception, uncertainty-aware decision control, and language-based reasoning to produce safe and interpretable first-aid outputs.

Language model integration

A Mistral [16] language model was executed locally using the Ollama [17] framework to generate first-aid guidance. The model was selected due to its lightweight architecture and support for local inference, enabling privacy-preserving deployment without reliance on external APIs while maintaining the ability to generate structured and controlled language outputs. The model operates on structured inputs derived from vision outputs, including wound type, anatomical region, and confidence scores. Prompt templates were designed to enforce safety constraints, including avoidance of diagnostic claims and inclusion of escalation recommendations.

Uncertainty-aware selective prediction

A selective prediction mechanism was implemented to regulate automated guidance. Predictions are accepted only when both wound and anatomical confidence scores exceed a predefined threshold. If confidence is below the threshold or uncertain classes are predicted, the system abstains from automated guidance and requests additional user input.

The operating threshold was set to $\tau = 0.70$ based on empirical inspection of model confidence distributions on the validation set. This value was selected to balance the trade-off between automated response coverage and the risk of low-confidence predictions, favoring conservative behavior in uncertain cases. The uncertainty-aware reasoning workflow is illustrated in Figure *3*.

How to cite this article:
Barua S, Bahre R, Babiceanu R, et al. (May 18, 2026) Uncertainty-Aware Wound Localization and First-Aid Decision Support From Smartphone Images Using an Explainable Multimodal AI Framework. Cureus J Comput Sci 3 : es44389-026-00077-z.
DOI https://doi.org/10.7759/s44389-026-00077-z

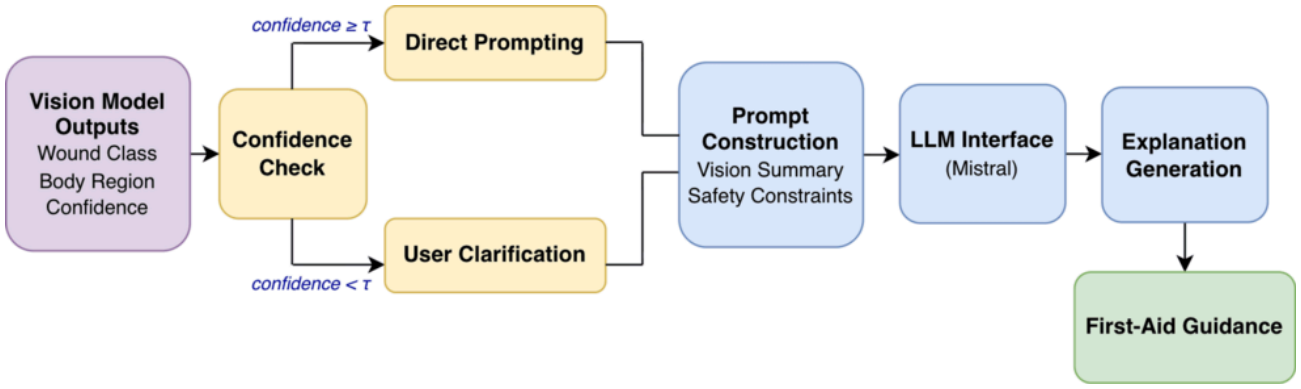

FIGURE 3: Uncertainty-aware multimodal reasoning workflow for first-aid guidance

Vision model outputs, including predicted wound class, anatomical region, and associated confidence scores, are first evaluated using a predefined confidence threshold τ. When confidence values exceed the threshold, predictions are routed directly to prompt construction. When confidence is below the threshold, the system enters a clarification stage and requests additional user input. The resulting structured prompt, incorporating vision outputs and safety constraints, is processed by a local Mistral language model to generate first-aid guidance along with concise textual explanations.

Experimental setup

All models were trained using Google Colab Pro+ with an NVIDIA A100 GPU. The YOLOv11 segmentation model was trained with an input resolution of 640 × 640, a batch size of 16, and a maximum of 100 epochs. Early stopping based on validation performance was applied, and training was terminated once performance plateaued, indicating convergence prior to reaching the maximum epoch limit. The model was initialized with pretrained weights and optimized using the Adam optimizer.

Classification models were trained with an input resolution of 224 × 224 using the Adam optimizer with a learning rate of 1×10^{-4} and cross-entropy loss. Early stopping was similarly applied based on validation performance. Training loss curves demonstrated rapid reduction in early epochs followed by stabilization, suggesting that the models reached convergence and that extending training beyond the selected epoch limit was unlikely to yield significant performance improvements.

Data augmentation included horizontal flips, small rotations, and brightness variations to improve robustness to real-world image variability. Dataset splits and preprocessing pipelines were kept consistent across all models to ensure fair comparison between architectures.

Definition of experimental settings

In this study, the term *proposed framework* refers to the primary experimental configuration combining YOLOv11-based wound segmentation, ResNet-50 anatomical classification, uncertainty-aware routing, and language-based guidance. In contrast, *baseline experiments* refer to independent model evaluations conducted for comparative benchmarking, including U-Net for segmentation and EfficientNet-B0 for classification. These baseline models were trained and evaluated under identical dataset splits and preprocessing conditions to provide contextual performance comparisons against the proposed framework.

Deployment considerations

The proposed pipeline is designed for modular inference and can be deployed in local, edge, or cloud-based environments. In this study, inference was performed using a local setup, where both vision and language components were executed without reliance on external APIs. For language-based guidance, a lightweight Mistral large language

model was deployed locally using the Ollama framework, enabling offline operation.

Inference is performed sequentially within a modular pipeline. Given an input smartphone image, the segmentation model first localizes the wound region, followed by anatomical classification. The predicted outputs and associated confidence scores are then passed to the uncertainty-aware routing mechanism, which determines whether to generate guidance or request clarification. When confidence exceeds the predefined threshold, structured prompts are used to generate first-aid guidance through the language model.

This modular design allows flexible deployment across different environments, including local, edge, or cloud-based systems. While the current implementation emphasizes offline operation for privacy and accessibility, further work is required to systematically evaluate latency, computational efficiency, and deployment trade-offs across different hardware configurations.

Evaluation metrics

Segmentation performance was evaluated using mean average precision at an Intersection-over-Union threshold of 0.5 (mAP@0.5) and averaged across thresholds from 0.50 to 0.95 (mAP@0.5-0.95), which are standard metrics for object detection and segmentation tasks. Pixel accuracy and mean Intersection-over-Union were additionally computed for the U-Net baseline to enable comparison with traditional segmentation evaluation measures.

Classification performance was evaluated using accuracy, precision, recall, and F1-score to capture both overall correctness and class-wise performance under potential class imbalance.

System-level behavior was analyzed across multiple confidence thresholds by measuring the proportion of cases routed to automated guidance versus clarification. This evaluation reflects the effectiveness of the uncertainty-aware routing mechanism in balancing automated coverage and conservative decision-making.

Results And Discussion

Dataset overview

The wound segmentation dataset consisted of 2,805 images collected from publicly available sources and curated to represent non-clinical, real-world conditions. The dataset was split into 2,375 training images, 273 validation images, and 157 test images. The images included four classes: *healthy_skin*, *wound_burn*, *wound_cut*, and *wound_unknown*. The *wound_unknown* class was included to capture ambiguous or low-quality visual regions, enabling uncertainty-aware behavior in downstream processing.

The class distribution of the wound segmentation training set is summarized in Table *1*.

How to cite this article:
Barua S, Bahre R, Babiceanu R, et al. (May 18, 2026) Uncertainty-Aware Wound Localization and First-Aid Decision Support From Smartphone Images Using an Explainable Multimodal AI Framework. Cureus J Comput Sci 3 : es44389-026-00077-z.
DOI https://doi.org/10.7759/s44389-026-00077-z

Class	Images containing class (non-exclusive)	Pixel count
healthy_skin	2,375	876,945,157
wound_burn	1,010	57,705,793
wound_cut	730	6,725,456
wound_unknown	190	31,423,594

TABLE 1: Class distribution in the wound segmentation training set

Image counts are non-exclusive because multiple annotated regions may occur within a single image. Pixel counts represent the total number of annotated pixels per class.

Experimental overview

Experiments were conducted using Google Colab Pro+ with an NVIDIA A100 GPU. Models were trained using consistent dataset splits and preprocessing pipelines across all experiments to ensure fair comparison. Detailed training configurations are provided in the Methods section.

Wound segmentation performance

The YOLOv11-Segmentation model demonstrated moderate but practically informative wound localization performance on a heterogeneous test set of 157 smartphone images. The evaluation was intentionally conducted under minimally controlled conditions to reflect realistic variability in lighting, occlusion, blur, and image composition commonly encountered in home-care scenarios. Performance metrics are summarized in Table *2*.

Class	Precision	Recall	mAP@50	mAP@50-95
healthy_skin	0.831	0.826	0.779	0.693
wound_burn	0.850	0.734	0.724	0.670
wound_cut	0.807	0.727	0.814	0.708
wound_unknown	0.776	0.667	0.767	0.671
Overall	0.816	0.738	0.771	0.685

TABLE 2: Performance of YOLOv11-Segmentation on wound detection

Values represent class-wise precision, recall, and mean average precision (mAP) at intersection-over-union thresholds of 0.5 and 0.5–0.95. The "Overall" row indicates aggregated performance across all classes.

How to cite this article:
Barua S, Bahre R, Babiceanu R, et al. (May 18, 2026) Uncertainty-Aware Wound Localization and First-Aid Decision Support From Smartphone Images Using an Explainable Multimodal AI Framework. Cureus J Comput Sci 3 : es44389-026-00077-z. DOI https://doi.org/10.7759/s44389-026-00077-z

Overall, the model achieved an mean average precision (mAP)@50 of 0.771 and an mAP@50-95 of 0.685. Despite substantial variability in image quality, these values indicate a reasonable consistency in identifying wound-related regions. Performance differed across classes, with stronger results for *wound_burn* and *healthy_skin* and comparatively lower performance for *wound_unknown*, which was designed to capture ambiguous or low-quality visual regions. This pattern reflects expected challenges in open-world image interpretability and supports the role of uncertainty-aware handling within the system. These findings are consistent with the feasibility objective defined in this study, emphasizing stable system behavior under heterogeneous real-world conditions rather than optimized predictive performance.

Class-wise segmentation performance is further illustrated in Figure *4*, which presents mAP across wound categories. The results highlight stronger performance for *wound_cut* and *healthy_skin*, with comparatively lower performance for *wound_burn* and *wound_unknown*, reflecting the challenges associated with visually ambiguous or low-quality inputs.

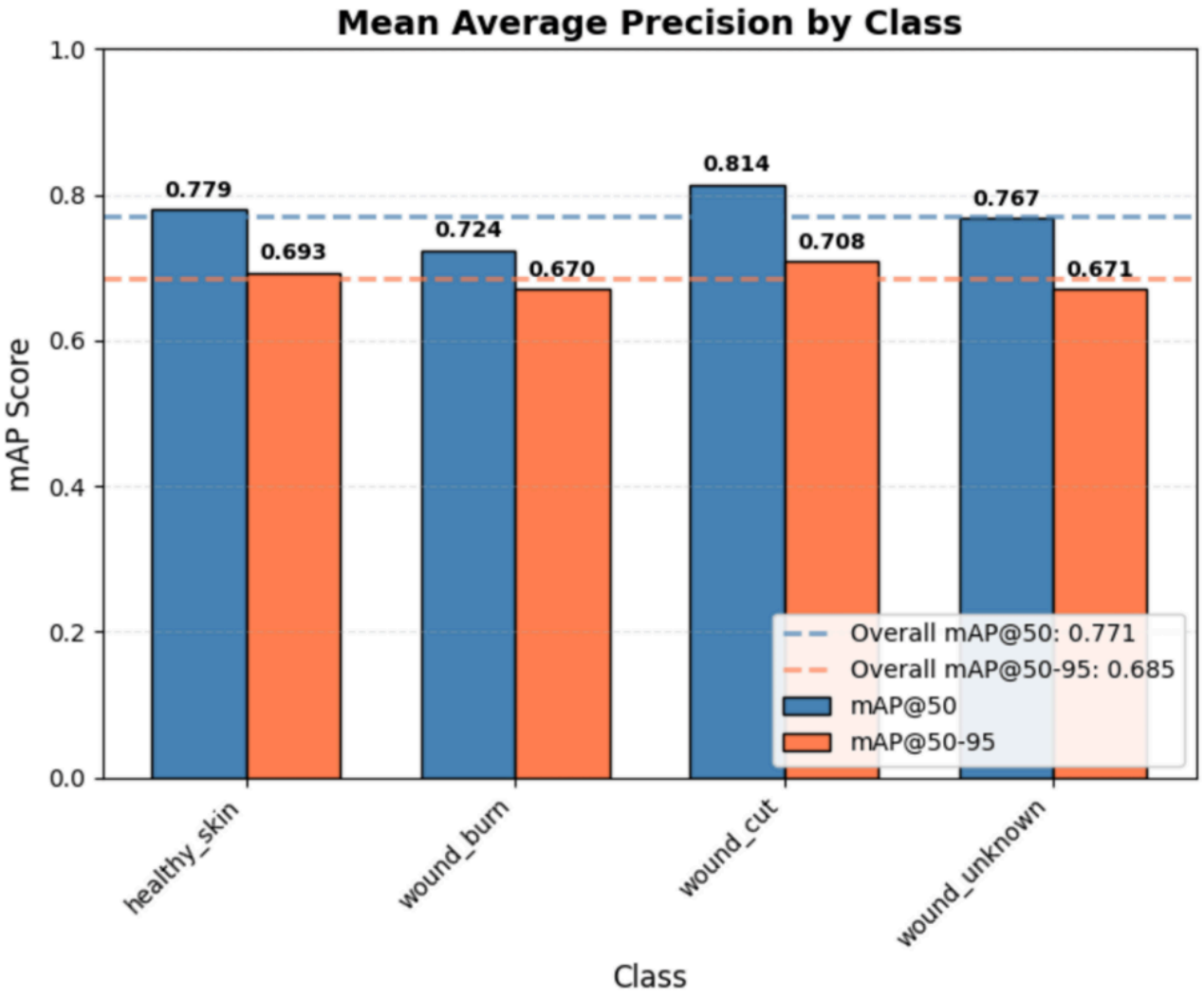

FIGURE 4: Mean average precision (mAP) by wound class for segmentation performance

Bar chart showing class-wise mean average precision at intersection-over-union thresholds of 0.50 and 0.50–0.95. Dashed lines indicate overall mAP performance across all classes.

How to cite this article:
Barua S, Bahre R, Babiceanu R, et al. (May 18, 2026) Uncertainty-Aware Wound Localization and First-Aid Decision Support From Smartphone Images Using an Explainable Multimodal AI Framework. Cureus J Comput Sci 3 : es44389-026-00077-z. DOI https://doi.org/10.7759/s44389-026-00077-z

Comparative segmentation benchmarking

A U-Net model with a ResNet34 encoder was trained using the same dataset split to provide contextual benchmarking against a widely used semantic segmentation architecture. While YOLOv11 performs instance-aware segmentation optimized for real-time applications, U-Net represents a classical pixel-wise segmentation approach commonly used in medical image analysis. The comparison is therefore intended to provide complementary performance context rather than a direct architectural equivalence.

The U-Net model achieved a pixel accuracy of 0.9623 and a mean intersection-over-union of 0.7004, as summarized in Table *3*.

Metric	Value
Pixel Accuracy	0.9623
Mean Intersection-over-Union (IoU)	0.7004

TABLE 3: Segmentation performance of the U-Net baseline model

Values represent the overall segmentation performance of the U-Net model evaluated on the test set. Pixel accuracy reflects the proportion of correctly classified pixels, while mean IoU measures the overlap between predicted and ground-truth segmentation regions.

Per-class segmentation performance is presented in Table *4*, where intersection-over-union values are reported for each wound category along with the mean intersection-over-union across all classes.

Class	Intersection-over-Union (IoU)
healthy_skin	0.9615
wound_burn	0.6418
wound_cut	0.6632
wound_unknown	0.5351
Mean IoU	0.7004

TABLE 4: Per-class IoU for the U-Net baseline model

Values represent per-class segmentation performance of the U-Net model on the test set. IoU measures the overlap between predicted and ground-truth regions for each class. The mean IoU is reported as the average across all classes.

How to cite this article:
Barua S, Bahre R, Babiceanu R, et al. (May 18, 2026) Uncertainty-Aware Wound Localization and First-Aid Decision Support From Smartphone Images Using an Explainable Multimodal AI Framework. Cureus J Comput Sci 3 : es44389-026-00077-z.
DOI https://doi.org/10.7759/s44389-026-00077-z

Across both models, a consistent pattern emerged: visually dominant and homogeneous regions such as *healthy_skin* were segmented more accurately than rare or ambiguous wound categories. This observation suggests that performance limitations are primarily driven by dataset variability and class imbalance rather than a specific architectural choice. Importantly, both models provide useful but distinct perspectives on segmentation performance under open-world conditions, supporting their use for contextual benchmarking within this study.

Body location classification performance

The ResNet-50 anatomical classifier achieved an overall accuracy of 82.11% on the test set, indicating stable performance on visually diverse, non-clinical images representative of real-world conditions. Performance metrics are summarized in Table *5*.

Class	Precision	Recall	F1-Score
arm	0.77	0.81	0.79
hand	0.81	0.79	0.80
leg	0.85	0.92	0.88
other	0.74	0.75	0.74
Overall	0.821	0.817	0.802

TABLE 5: Body location classification performance of the ResNet-50 model

Values represent class-wise precision, recall, and F1-score for anatomical region classification on the test set. The "Overall" row indicates aggregated performance across all classes.

Classification performance varied across anatomical regions, with the *leg* class demonstrating consistently strong performance across both precision and recall. In contrast, lower recall for the *hand* class and reduced precision for the *other* category indicate challenges in distinguishing visually similar or ambiguous anatomical regions.

A normalized confusion matrix for the ResNet-50 classifier is shown in Figure *5*. The matrix indicates strong class-wise performance along the diagonal, particularly for the leg class (0.92), reflecting high correct classification rates. Misclassifications are primarily observed between visually similar regions, such as *hand* and *arm*, and between ambiguous samples and the *other* category. The *other* class captures uncertain or partially visible anatomical regions, reducing forced misclassification into specific categories. These results provide a detailed view of class-wise prediction behavior and complement the aggregate performance metrics reported in Table *5*.

How to cite this article:
Barua S, Bahre R, Babiceanu R, et al. (May 18, 2026) Uncertainty-Aware Wound Localization and First-Aid Decision Support From Smartphone Images Using an Explainable Multimodal AI Framework. Cureus J Comput Sci 3 : es44389-026-00077-z.
DOI https://doi.org/10.7759/s44389-026-00077-z

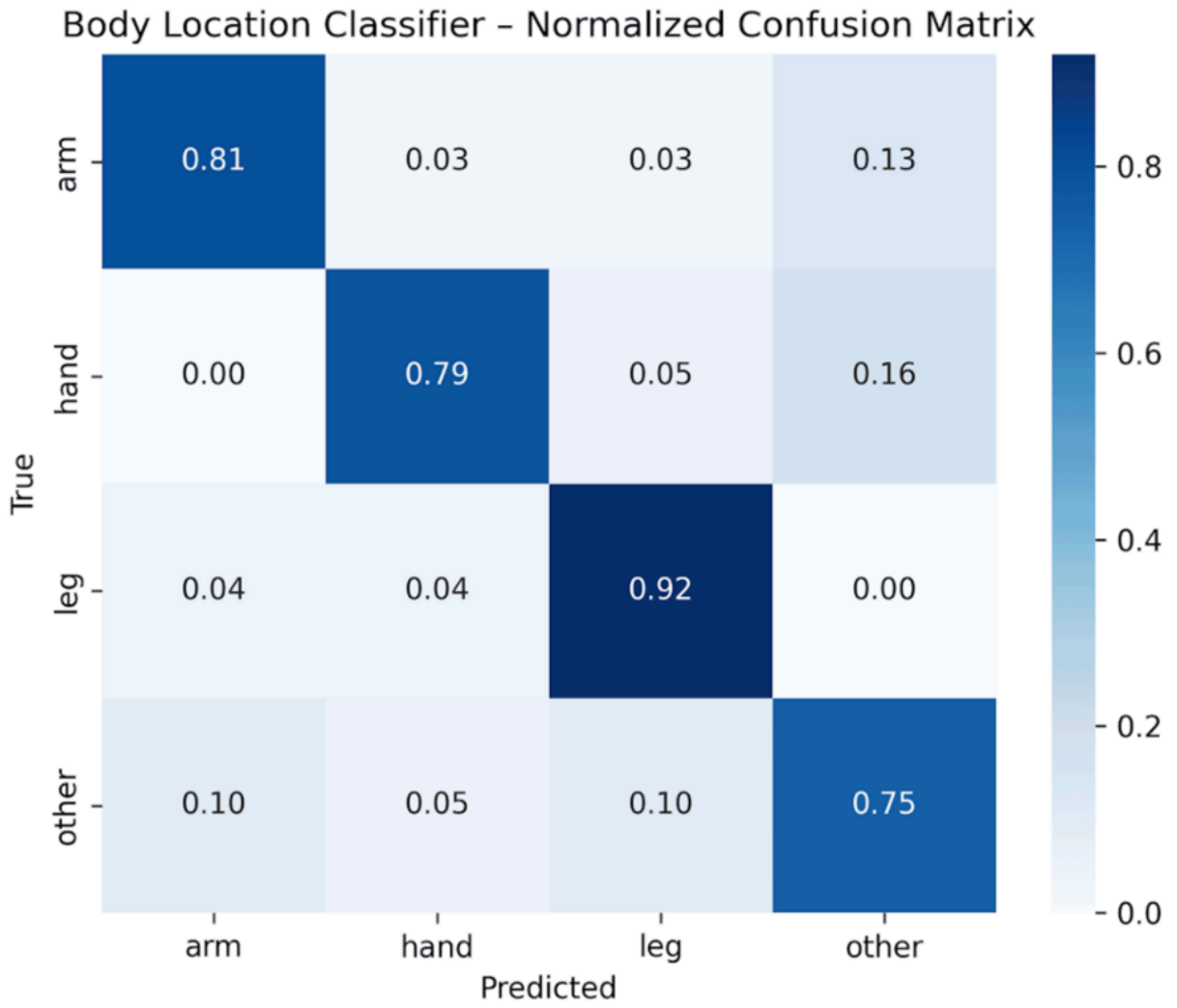

FIGURE 5: Normalized confusion matrix for anatomical region classification using ResNet-50

Rows represent true anatomical region labels, and columns represent predicted labels. Values are normalized proportions for each true class, where diagonal entries indicate correct classifications and off-diagonal values indicate misclassifications. Higher values along the diagonal reflect better class-wise performance.

Training loss for the ResNet-50 classifier is shown in Figure *6*. The curve demonstrates stable convergence, with rapid loss reduction in early epochs followed by gradual stabilization. Early stopping was applied based on validation performance, and training was terminated once performance plateaued, indicating that additional epochs were unlikely to yield significant improvement.

How to cite this article:
Barua S, Bahre R, Babiceanu R, et al. (May 18, 2026) Uncertainty-Aware Wound Localization and First-Aid Decision Support From Smartphone Images Using an Explainable Multimodal AI Framework. Cureus J Comput Sci 3 : es44389-026-00077-z.
DOI https://doi.org/10.7759/s44389-026-00077-z

FIGURE 6: Training loss curve for the ResNet-50 anatomical classification model

The curve illustrates cross-entropy loss across training epochs. Rapid initial loss reduction followed by stabilization indicates effective learning and convergence without prolonged overfitting.

Grad-CAM visualizations consistently highlighted anatomically meaningful regions, including limb contours and joint structures. This suggests that predictions were influenced by relevant spatial features, although this observation is qualitative and does not constitute formal validation of model reasoning.

A comparative experiment with EfficientNet-B0 (Table *6*) showed lower accuracy (0.7442), supporting the selection of ResNet-50 for this application.

How to cite this article:
Barua S, Bahre R, Babiceanu R, et al. (May 18, 2026) Uncertainty-Aware Wound Localization and First-Aid Decision Support From Smartphone Images Using an Explainable Multimodal AI Framework. Cureus J Comput Sci 3 : es44389-026-00077-z.
DOI https://doi.org/10.7759/s44389-026-00077-z

Model	Overall test accuracy
ResNet-50	0.8211
EfficientNet-B0	0.7442

TABLE 6: Comparative anatomical classification accuracy of evaluated models

Values represent test-set classification accuracy for anatomical region prediction using different model architectures under identical experimental conditions.

LLM-based first-aid guidance and safety evaluation

The language-model component generated structured first-aid recommendations conditioned on predicted wound type, anatomical region, and associated confidence values. A key feature of the system is the incorporation of uncertainty-aware routing, which determines whether guidance is generated directly or deferred pending additional user input.

Threshold sensitivity analysis (Table *7*) demonstrated that increasing the confidence threshold resulted in more conservative system behavior. As the threshold increased, the proportion of cases routed directly to automated guidance decreased, while clarification requests increased. A threshold of 0.70 provided a balanced but safety-oriented operating point, with clarification slightly exceeding direct responses. This behavior is desirable in the context of first-aid support, where avoiding overconfident or potentially misleading recommendations is more important than maximizing automated coverage.

Threshold	Direct guidance (%)	User clarification (%)
0.50	70.50	29.50
0.60	65.23	34.77
0.70	62.13	37.87
0.80	47.58	52.42

TABLE 7: Threshold sensitivity analysis for uncertainty-aware routing

Values represent the proportion of test cases routed directly to automated first-aid guidance versus those requiring user clarification at different confidence thresholds. Increasing the threshold results in more conservative system behavior, with a higher proportion of cases routed to clarification.

Since standardized quantitative metrics for first-aid language outputs are not well established in this setting, the generated recommendations were evaluated qualitatively (Table *8*). Across representative scenarios, the system produced concise, non-diagnostic, and safety-aligned guidance consistent with basic first-aid practices. In ambiguous

cases, the system appropriately deferred guidance and requested clarification, reinforcing its conservative design. These observations should be interpreted as qualitative system behavior rather than a formal evaluation of language model correctness.

Scenario	LLM output (summary)	Analysis
Burn on arm	Cooling with running water, light covering, and avoidance of ointments.	Guidance aligns with common first-aid principles for superficial burns and emphasizes conservative care.
Cut on hand	Gentle cleaning, application of pressure, and monitoring for swelling or redness.	Instructions prioritize hygiene and bleeding control while identifying relevant warning signs.
Unknown wound on leg	Requests clarification regarding wound appearance and severity before providing guidance.	System appropriately abstains under uncertainty, demonstrating safety-aware behavior.
Cut on other region	Cleaning and compression with a request to confirm the anatomical location.	Guidance remains constrained until sufficient contextual information is available.

TABLE 8: Qualitative examples of large language model (LLM)-based first-aid guidance and system behavior

Examples illustrate representative system behavior across different wound scenarios. Outputs are generated by the language model based on structured predictions from vision modules and are constrained to non-diagnostic, first-aid guidance. The analysis column summarizes alignment with basic first-aid principles and the system's uncertainty-aware decision behavior.

To improve transparency, the system also generated short textual explanations linking predicted attributes and confidence signals to the resulting recommendations. These explanations were constrained to reflect observable inputs rather than internal reasoning processes. An illustrative example is shown below:

> Based on the detected burn wound on the arm region (82% confidence), I recommend cooling with running water because burns require immediate temperature reduction, and the arm's accessibility makes this feasible. The moderate confidence suggests monitoring for blister formation.

Overall, these results indicate that language-based guidance was tightly coupled to upstream confidence signals and system-level safety controls, supporting the feasibility of uncertainty-aware first-aid support in open-world wound assessment.

Computational efficiency

The system demonstrated practical computational performance under experimental conditions. Average end-to-end inference time was approximately four to five seconds per image, including segmentation (1.8 seconds), anatomical classification (0.3 seconds), and language-based guidance generation (2.5 seconds).

How to cite this article:
Barua S, Bahre R, Babiceanu R, et al. (May 18, 2026) Uncertainty-Aware Wound Localization and First-Aid Decision Support From Smartphone Images Using an Explainable Multimodal AI Framework. Cureus J Comput Sci 3 : es44389-026-00077-z.
DOI https://doi.org/10.7759/s44389-026-00077-z

Failure case analysis

To better understand system limitations, several representative failure cases were qualitatively analyzed. Common failure scenarios included low image quality (e.g., motion blur or poor lighting), partial occlusion of the wound region, and visually ambiguous wound appearances. In such cases, segmentation masks were often incomplete or imprecise, and anatomical classification confidence decreased.

The *wound_unknown* category frequently captured ambiguous or low-quality visual inputs, reflecting the intended design of the dataset to account for uncertainty. In these situations, the uncertainty-aware routing mechanism correctly reduced automated guidance and instead prompted user clarification, demonstrating appropriate conservative behavior.

Failure cases also revealed challenges in distinguishing between visually similar wound types and in handling background textures that resemble wound regions. These observations suggest that performance is influenced not only by model architecture but also by dataset variability and inherent visual ambiguity in real-world images.

Qualitative end-to-end case study

A representative example is presented in Figure *7*. The system successfully localized a superficial burn and identified the anatomical region, with segmentation outputs and Grad-CAM visualizations highlighting relevant spatial features. These visual results provide interpretation for both wound detection and anatomical classification.

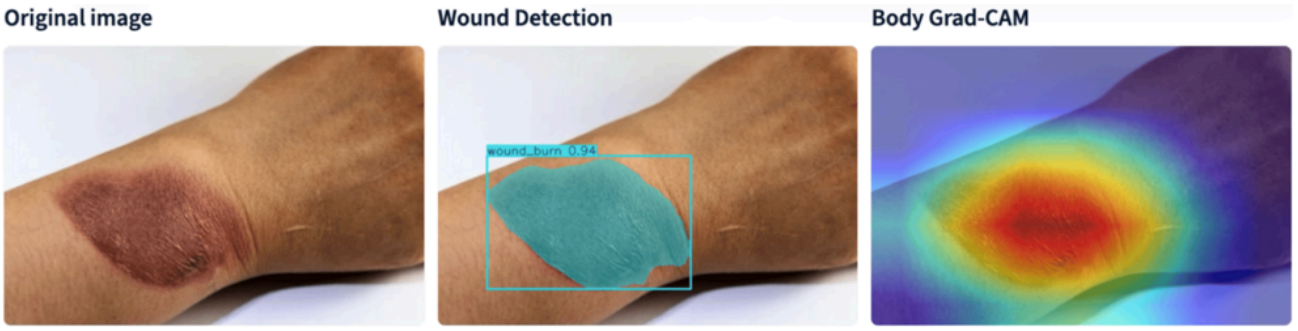

FIGURE 7: Qualitative end-to-end example illustrating system outputs

The first panel shows the original smartphone image. The second panel shows YOLOv11 wound segmentation with predicted class and associated confidence. The third panel shows the Grad-CAM heatmap used for anatomical classification. Grad-CAM highlights regions contributing to the model's prediction and should be interpreted as an explanatory visualization rather than clinical ground truth.

Figure *8* displays the corresponding structured model outputs. These outputs include predicted wound type, anatomical region, and associated confidence scores, which are used by the uncertainty-aware routing mechanism to determine whether to generate direct guidance or request user clarification.

How to cite this article:
Barua S, Bahre R, Babiceanu R, et al. (May 18, 2026) Uncertainty-Aware Wound Localization and First-Aid Decision Support From Smartphone Images Using an Explainable Multimodal AI Framework. Cureus J Comput Sci 3 : es44389-026-00077-z. DOI https://doi.org/10.7759/s44389-026-00077-z

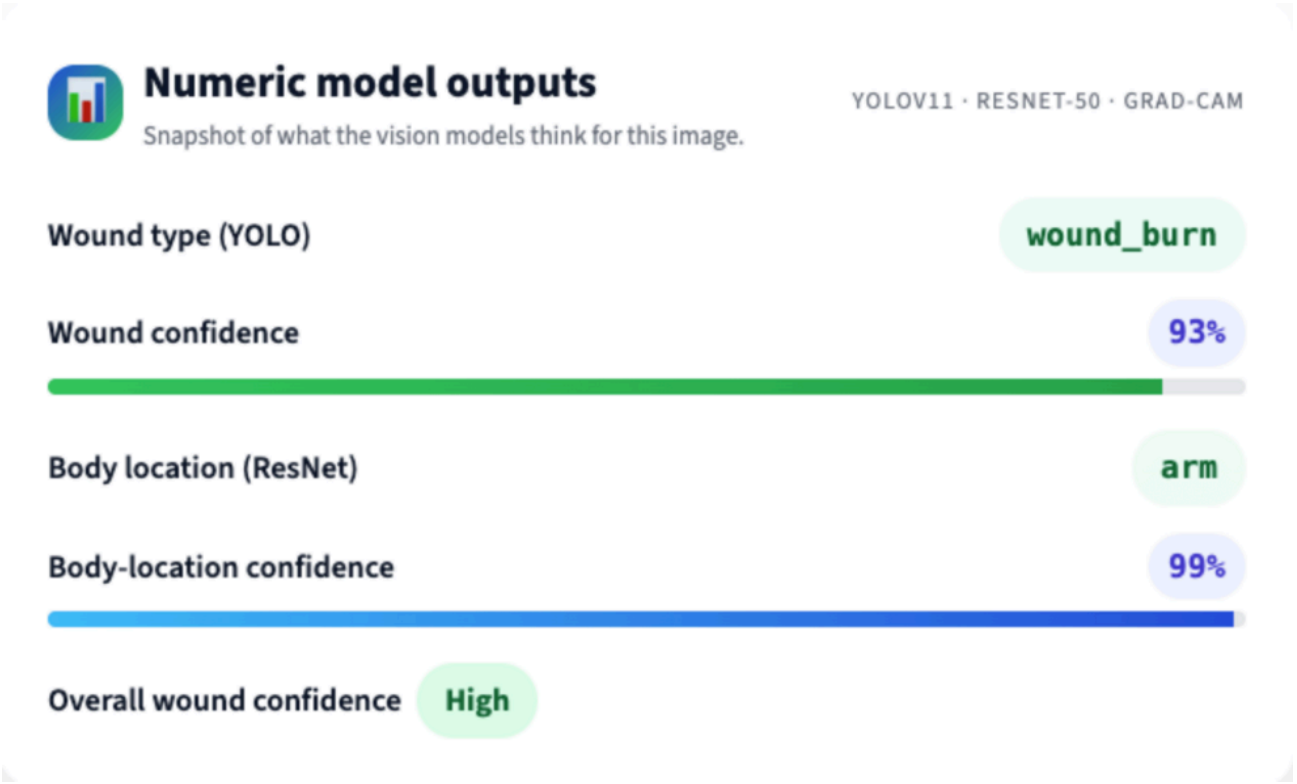

FIGURE 8: Numeric model outputs and confidence-based routing signals

The panel summarizes structured outputs from the vision models, including wound-type prediction from YOLOv11 and anatomical classification from ResNet-50, along with their associated confidence scores. These values are used by the uncertainty-aware routing mechanism to determine whether the system proceeds directly to first-aid guidance or enters a clarification stage. The reported confidence values represent model prediction scores and should not be interpreted as clinical certainty.

Together, these results illustrate how segmentation, anatomical context, interpretability through Grad-CAM, and confidence-aware decision-making are integrated into a coherent workflow. This example demonstrates the feasibility of combining perception and controlled reasoning for safety-oriented first-aid support.

Discussion

Interpretation of Findings

The findings of this study support the feasibility of safety-aware wound assessment using smartphone images captured under real-world conditions. Unlike controlled clinical datasets, the images used in this study exhibit substantial variability in lighting, framing, and background content, which more closely reflects how non-expert users would capture wound images in practice.

The segmentation results indicate that wound-related visual features can be identified with moderate reliability under these conditions. While performance is lower for ambiguous or visually complex categories, this behavior is expected and reinforces the importance of incorporating uncertainty-aware mechanisms rather than relying solely on predictive accuracy.

The anatomical classification results further suggest that coarse body-region context can be recovered with sufficient consistency to support preliminary first-aid reasoning. In this application, broad anatomical categorization is more relevant than fine-grained precision, as guidance primarily depends on general region-specific considerations.

How to cite this article:
Barua S, Bahre R, Babiceanu R, et al. (May 18, 2026) Uncertainty-Aware Wound Localization and First-Aid Decision Support From Smartphone Images Using an Explainable Multimodal AI Framework. Cureus J Comput Sci 3 : es44389-026-00077-z. DOI https://doi.org/10.7759/s44389-026-00077-z

A central contribution of this work is the integration of uncertainty-aware routing with language-based guidance. Rather than generating recommendations indiscriminately, the system explicitly accounts for prediction confidence and defers guidance when uncertainty is high. This behavior is particularly important in safety-sensitive applications, where overconfident or unsupported recommendations may pose a risk. This work distinguishes itself by emphasizing safety-aware integration, where perception, uncertainty handling, and controlled language generation are explicitly coupled within a unified decision-making framework rather than treated as independent components.

Importantly, the results should be interpreted within the context of a feasibility-oriented study. The objective was not to achieve state-of-the-art predictive performance but to evaluate whether perception, uncertainty handling, and constrained language generation can be combined into a coherent and safety-aware workflow. Within this scope, the observed results provide evidence supporting the viability of the proposed approach. These findings are consistent with the failure cases analyzed in the Results section, where image quality degradation, occlusion, and visual ambiguity were key factors affecting model performance.

Practical and System-Level Significance

The proposed workflow aligns with several principles relevant to real-world first-aid support systems. First, it incorporates anatomical context rather than treating wound recognition as an isolated task. Second, it integrates uncertainty-aware decision-making to reduce the risk of overconfident automation. Third, it provides both visual and textual explanations to improve transparency for non-expert users.

The use of local inference for both vision and language components also has practical implications. Offline operation can improve accessibility in low-resource or connectivity-constrained environments while reducing privacy concerns associated with cloud-based processing. However, latency, resource constraints, and deployment trade-offs were not systematically evaluated and require further investigation.

At the same time, the system is intentionally limited to minor, non-diagnostic wound scenarios. This constraint is appropriate for an early-stage feasibility study and helps ensure that the system is not interpreted as a clinical diagnostic tool.

Limitations

Several limitations should be acknowledged. The datasets were compiled from publicly accessible images instead of curated clinical repositories, thereby constraining direct clinical generalizability. The wound taxonomy was intentionally narrow and does not include many clinically relevant wound types or severity levels. The system has not been clinically validated, and its outputs should not be interpreted as medical diagnoses or treatment recommendations. Evaluation of the language-model component was limited to structured qualitative assessment and did not involve clinician review or formal human-subject studies. Confidence scores were not formally calibrated, and the selected threshold should be interpreted as an empirically motivated setting rather than a universally optimal value. Additionally, explainability methods such as Grad-CAM and textual rationales provide useful transparency but do not fully capture internal model reasoning. External validation on independent datasets was not performed, and generalization across diverse populations and imaging conditions remains to be established. These limitations are further reflected in the qualitative failure cases presented in the Results section, highlighting the impact of image quality, occlusion, and visual ambiguity on system performance.

Future Work

Future work should focus on expanding dataset quality and diversity through clinically annotated data collected under appropriate ethical protocols. Extending the wound taxonomy and incorporating severity estimation would improve real-world applicability. Additional evaluation should include clinician-led assessment of generated guidance, formal clinical validation, and external validation using independent datasets. In addition, future studies should incorporate quantitative evaluation frameworks for language-model outputs and investigate calibration techniques such as temperature scaling and reliability analysis to improve the interpretability and reliability of confidence estimates. Further

work is also needed to evaluate deployment trade-offs and optimize the system for mobile and edge environments. These directions are directly motivated by the failure cases observed in this study and aim to improve robustness under challenging real-world conditions.

Summary

This study demonstrates that wound localization, anatomical context modeling, uncertainty-aware abstention, and constrained language-based guidance can be integrated into a unified and interpretable workflow for image-based wound assessment under real-world conditions. While the system is not intended for clinical use, it establishes a foundation for safety-aware decision-support approaches in first-aid settings and highlights the importance of combining perception, uncertainty handling, and controlled reasoning in real-world AI applications.

Conclusions

This study presents a feasibility-oriented evaluation of an explainable multimodal workflow for image-based wound assessment that integrates wound segmentation, anatomical region classification, uncertainty-aware routing, and language-based first-aid guidance. The results indicate that stable and practically informative performance can be achieved on heterogeneous, smartphone-captured images, supporting the potential of such approaches for real-world, unconstrained conditions.

A central contribution of this work is the integration of uncertainty-aware selective prediction with controlled language generation, enabling the system to defer guidance in low-confidence scenarios rather than producing potentially unreliable recommendations. This design, combined with Grad-CAM-based visual explanations and structured textual rationales, provides a transparent and safety-oriented framework for decision support. While the system is not intended for clinical use and remains an exploratory proof-of-concept, the findings suggest that perception, uncertainty handling, and constrained reasoning can be combined into a coherent and interpretable pipeline. This approach provides a foundation for the future development of more robust, explainable, and uncertainty-aware first-aid support systems, subject to further validation and evaluation.

Additional Information

Author Contributions

All authors have reviewed the final version to be published and agreed to be accountable for all aspects of the work.

Concept and design: Sumit Barua

Acquisition, analysis, or interpretation of data: Sumit Barua, Ruth Bahre, Radu Babiceanu, Guan Y. Hong

Drafting of the manuscript: Sumit Barua, Ruth Bahre

Critical review of the manuscript for important intellectual content: Sumit Barua, Radu Babiceanu, Guan Y. Hong

Supervision: Radu Babiceanu, Guan Y. Hong

Disclosures

Human subjects: All authors have confirmed that this study did not involve human participants or tissue. **Animal subjects:** All authors have confirmed that this study did not involve animal subjects or tissue. **Conflicts of interest:** In compliance with the ICMJE uniform disclosure form, all authors declare the following: **Payment/services info:** All authors have declared that no financial support was received from any organization for the submitted work. **Financial relationships:** All authors have declared that they have no financial relationships at present or within the previous three years with any organizations that might have an interest in the submitted work. **Other relationships:** All authors have declared that there are no other relationships or activities that could appear to have influenced the submitted work.

Data Availability Statements

The wound-detection and anatomical-classification datasets used in this study were constructed by the authors through manual annotation and aggregation of openly available images obtained from Roboflow Universe repositories. Due to mixed licensing terms across the original sources, the combined annotated datasets cannot be redistributed as a single package. However, all original source datasets remain publicly accessible, and links to these datasets are provided in the reference list under the corresponding dataset citations.

References

1. Han G, Ceilley R: Chronic wound healing: a review of current management and treatments. Advances in Therapy. 2017, 34:599-610. 10.1007/s12325-017-0478-y
2. Chen L, Cheng L, Gao W, Chen D, Wang C, Ran X: Telemedicine in chronic wound management: systematic review and meta-analysis. JMIR Mhealth Uhealth. 2020, 8:e15574. 10.2196/15574
3. Høyland SA, Holte KA, Islam K, et al.: A cross-sector systematic review and synthesis of knowledge on telemedicine interventions in chronic wound management-Implications from a system perspective. International Wound Journal. 2023, 20:1712-1724. 10.1111/iwj.13986
4. Anisuzzaman DM, Wang C, Rostami B, Gopalakrishnan S, Niezgoda J, Yu Z: Image-based artificial intelligence in wound assessment: a systematic review. Advances in Wound Care. 2022, 11:687-709. 10.1089/wound.2021.0091
5. Holzinger A, Langs G, Denk H, Zatloukal K, Muller H: Causability and explainability of artificial intelligence in medicine. WIREs Data Mining and Knowledge Discovery. 2019, 9:e1312. 10.1002/widm.1312
6. Liu X, Cruz Rivera S, Moher D, et al.: Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: the CONSORT-AI extension. Nature Medicine. 2020, 26:1364-1374. 10.1038/s41591-020-1034-x
7. Cruz Rivera S, Liu X, Chan AW, et al.: Guidelines for clinical trial protocols for interventions involving artificial intelligence: the SPIRIT-AI extension. Nature Medicine. 2020, 26:1351-1363. 10.1038/s41591-020-1037-7
8. Mongan J, Moy L, Kahn CE Jr: Checklist for artificial intelligence in medical imaging (CLAIM): a guide for authors and reviewers. Radiology: Artificial Intelligence. 2020, 2:e200029. 10.1148/ryai.2020200029
9. Collins GS, Reitsma JB, Altman DG, Moons KG: Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. Annals of Internal Medicine. 2015, 162:55-63. 10.7326/m14-0697
10. Soenksen LR, Ma Y, Zeng C, et al.: Integrated multimodal artificial intelligence framework for healthcare applications. npj Digital Medicine. 2022, 5:149. 10.1038/s41746-022-00689-4
11. Warner E, Lee J, Hsu W, Syeda-Mahmood T, Kahn CE Jr, Gevaert O, Rao A: Multimodal machine learning in image-based and clinical biomedicine: survey and prospects. International Journal of Computer Vision. 2024, 132:3753-3769. 10.1007/s11263-024-02032-8
12. Singhal K, Azizi S, Tu T, et al.: Large language models encode clinical knowledge. Nature. 2023, 620:172-180. 10.1038/s41586-023-06291-2
13. Ultralytics YOLO. (2024). Accessed: November 26, 2025: https://github.com/ultralytics/ultralytics.
14. He K, Zhang X, Ren S, Sun J: Deep residual learning for image recognition. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA. 2016, 770-778. 10.1109/CVPR.2016.90
15. Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D: Grad-CAM: visual explanations from deep networks via gradient-based localization. 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy. 2017, 618-626. 10.1109/ICCV.2017.74
16. Jiang AQ, Sablayrolles A, Mensch A, et al.: Mistral 7B [PREPRINT]. arXiv. 2023, 10.48550/arXiv.2310.06825
17. Ollama: local model serving framework. (2024). Accessed: December 7, 2025: https://github.com/ollama/ollama.

How to cite this article:
Barua S, Bahre R, Babiceanu R, et al. (May 18, 2026) Uncertainty-Aware Wound Localization and First-Aid Decision Support From Smartphone Images Using an Explainable Multimodal AI Framework. Cureus J Comput Sci 3 : es44389-026-00077-z. DOI https://doi.org/10.7759/s44389-026-00077-z