Integrating articial intelligence-driven wound detection into simulation-based forensic medicine training

Boonsak Hanterdsith

Forensic Medicine Department, Maharat Nakhonratchasima Hospital, Nakhonratchasima

Abstract

Forensic wound detection and classification is a fundamental skill in forensic medicine education, enabling medical students to interpret injury mechanisms and provide medico-legal opinions. Traditional teaching is often limited by insufficient exposure to real forensic cases and lack of interactive, feedback-driven learning opportunities. This study aimed to develop, validate, and implement a real-time, Artificial Intelligence (AI)-assisted wound detection and classification system using You Only Look Once version 8 nano (YOLOv8n) and deploy it as a web-based application to enhance simulation-based forensic teaching for medical students. A custom dataset comprising 595 images of eight wound types (gsw_entrance, gsw_exit, wound_burn, wound_hanging, wound_hesitation, wound_laceration, wound_open_fracture, wound_strangulation) was collected from physical wound simulations and augmented using Roboflow. Images were annotated in YOLOv8 format and split into training (70%), validation (20%), and test (10%) sets. The YOLOv8n model was trained for 100 epochs on Google Colab with an NVIDIA T4 Graphics Processing Unit (GPU). Model performance was evaluated using mean average precision at IoU 0.50 (mAP@50), mAP@50–95, precision, recall, F1-score, and confusion matrix analysis. The best-performing model (best.pt, 6 MB) was deployed via a Streamlit-based web application, enabling image upload and real-time camera detection. The YOLOv8n model achieved high detection performance, with overall mAP@50 = 0.99, mAP@50–95 = 0.64, precision = 0.98, and recall = 0.99 on the test set (n = 24 images, 27 instances). Per-class F1-scores exceeded 0.99 for most wound types, with highest performance observed for gunshot exit and hanging (1.00). Real-time application testing demonstrated smooth inference on CPU-only devices, enabling immediate visual feedback during simulation sessions. This study demonstrates the feasibility and educational value of integrating Artificial Intelligence-based wound detection and classification into simulation-based teaching. The deployed web application supports cross-platform access without installation requirements, making it a practical and scalable tool for medical education. Future work will focus on expanding the dataset, improving generalizability across wound severities, and conducting formal evaluations of student learning outcomes to quantify the pedagogical impact of this approach.

Keywords: Forensic wound detection; Medical education; Simulation-based teaching; Computer vision; AI-driven learning

Received: 17 September 2025, Revised: 20 November 2025, Accepted: 30 November 2025

INTRODUCTION

Forensic medicine plays a key role in bridging medical science and the justice system. Wound detection is one of the fundamental skills required for interpreting the mechanism of injury, estimating the timing of trauma, and providing medico-legal opinions. However, teaching wound detection and classification to medical students is often limited by insufficient exposure to real forensic cases, legal and ethical

Correspondence: Boonsak Hanterdsith, Forensic Medicine Department, Maharat Nakhonratchasima Hospital, Nakhonratchasima 30000, Thailand, Tel.: 0-4423-5000, email: trapezius60@gmail.com

restrictions in observing autopsies, and variability in expert interpretation. Traditional lecture-based approaches may not sufficiently prepare students for real-world forensic practice because they lack opportunities for hands-on experience and repeated exposure to diverse wound patterns. Simulation-based education has been shown to enhance experiential learning, allowing students to safely practice decision-making skills in a controlled environment [1]. Artificial intelligence (AI) and computer vision (CV) technologies, particularly deep learning–based object detection models such as You Only Look Once version 8 (YOLOv8), have the unique capability to simultaneously detect the location of wounds in an image and classify them into predefined categories [2]. This dual capability mirrors the cognitive process that forensic practitioners use in real cases: first identifying the wound site, then reasoning about its type and potential cause. Integrating AI into simulation-based teaching not only provides students with immediate, objective feedback but also exposes them to emerging technology that is increasingly relevant in forensic casework and digital evidence analysis [3]. By combining simulation-based teaching with AI-powered wound detection and classification, educators can create a more interactive, feedback-driven learning environment that encourages students to critically evaluate both their own decisions and the algorithm's output. This blended approach has the potential to bridge the gap between theoretical knowledge and practical forensic skills while familiarizing students with the limitations and ethical considerations of AI-assisted forensic work.

The primary aim of this study was to design, develop, and integrate an artificial intelligence–based wound detection and classification model into simulation-based forensic scene investigation teaching for undergraduate medical students. Simulation-based teaching was chosen to ensure control, ethical feasibility, and consistency in forensic learning environments—conditions under which real-world wound images may not be ethically or legally accessible. The specific objectives were:1) To develop a computer vision model using YOLOv8 for simultaneous wound detection and classification from simulated forensic mannequins. 2) To deploy the trained model as a web-based application capable of providing real-time feedback during simulation sessions.

MATERIALS AND METHODS

Definition of Terms

Wound types:

- **gsw_entrance** (Gunshot wound entrance): GSW entrance has distinctive characteristics including a punch-out lesion. The entrance may show firing residue components such as soot/gun powder tattooing as shown in the image.
- **gsw_exit** (Gunshot wound exit or GSW exit): GSW exit has distinctive characteristics including a punch-out lesion. The exit wound does not contain firing residue components and is typically larger than the entrance wound. May have an irregular stellate shape resembling a laceration.
- **wound_burn** (Burn wound): Wound resulting from thermal injury. Assess the depth and distribution pattern of the wound to determine consistency with the reported incident.
- **wound_hanging** (Neck compression from hanging): Typically characterized by an oblique upward compression pattern around the neck. This is characteristic of hanging and must be evaluated in conjunction with other evidence.
- **wound_hesitation** (Hesitation wound): Commonly found in individuals attempting self-harm. Characterized by multiple superficial wounds that are parallel or nearly parallel to each other, located near the main wound or in easily accessible areas such as the inner wrist.
- **wound_laceration** (Laceration, irregular wound edges): Caused by blunt hard objects striking the skin surface. Characterized by irregular edges, abrasions and bruising at wound margins, and tissue bridging/undermining.
- **wound_open_fracture** (Open fracture): Occurs when fractured bone pierces through the skin surface. In vehicle collision cases, it can be used for accident reconstruction.
- **wound_strangulation** (Neck compression from strangulation): Characterized by horizontal compression

pattern around the neck. If both types of neck compression patterns are found, consider the possibility of disguised homicide.

Technical terms:

- **YOLOv8n:** "You Only Look Once" version 8 (nano variant), a convolutional neural network architecture for object detection.
- **mAP@50–95:** Mean Average Precision averaged across Intersection over Union thresholds from 0.50 to 0.95.
- **Confidence Threshold:** Minimum probability for considering a prediction valid.
- **Non-Maximum Suppression (NMS):** A post-processing technique to eliminate overlapping bounding boxes.

Dataset and Pre-processing

A total of 595 images depicting eight forensic wound types (gsw_entrance, gsw_exit, wound_burn, wound_hanging, wound_hesitation, wound_laceration, wound_open_fracture, wound_strangulation) were collected from simulated wounds created using forensic mannequins and moulage techniques. For each wound type, a single physical simulation was prepared. High-resolution digital photographs were then taken from multiple angles (such as frontal and lateral), under a range of lighting conditions (including natural light and artificial light at varying intensities), and at distances spanning from close-up (10 cm) to mid-range (50 cm). This approach was designed to maximize diversity and enhance the robustness of the model.

Images were resized to 640x640 pixels (stretch mode) and converted to grayscale using a CRT phosphor filter using Roboflow Annotate (https://app.roboflow.com/). This approach enhanced contrast in wound margins and reduced dependence on variable lighting or skin tone. YOLOv8 default preprocessing applied normalization, random horizontal flips (p=0.5), and light contrast-limited adaptive histogram equalization (CLAHE).

Data Augmentation

To increase generalizability and mitigate class imbalance, a combination of geometric and photometric augmentations was applied to create three additional versions of each source image. Augmentation was applied only to training and validation sets, not to the test set, to prevent data leakage. The following augmentations were processed with Roboflow and applied probabilistically:

- 50% probability of horizontal flip
- Equal probability of one of the following 90-degree rotations: none, clockwise, counter-clockwise
- Randomly crop between 0 and 20 percent of the image
- Random rotation of between -15 and +15 degrees
- Random shear of between -10° to +10° horizontally and -10° to +10° vertically
- Random brigthness adjustment of between -15 and +15 percent
- Random exposure adjustment of between -10 and +10 percent
- Random Gaussian blur of between 0 and 2.5 pixels
- Salt and pepper noise was applied to 0.1 percent of pixels

Data Splitting

The dataset was divided into training, validation, and test sets following a 70:20:10 ratio before augmentation. Stratification ensured that all wound classes appeared in each split.

Model Architecture and Training

Model training utilized the You Only Look Once version 8 nano (YOLOv8n) architecture (Ultralytics YOLO v8.3.191) [4], the smallest and most computationally efficient model in the YOLOv8 family, containing approximately 3.2 million parameters. YOLOv8n was selected to balance accuracy, generalization, real-time inference performance in abrowser-based educational environment. This model has been proven performance in medical object detection research.

A pretrained YOLOv8n backbone (trained on the COCO dataset) was used for transfer learning. Although COCO does not contain wound images, its rich representation of natural images provides transferable low- and mid-level features such as edges, shapes, and textures. Fine-tuning on the custom wound dataset allowed the model to adapt these features to forensic wound characteristics, enabling robust detection while avoiding overfitting.

Training was conducted on Google Colab with an NVIDIA Tesla T4 GPU using automatic mixed precision (AMP) to improve memory efficiency and training speed. The learning rate was set to 0.01, with AdamW optimizer and a weight decay of 0.0005 to enhance generalization. Inference thresholds were confidence = 0.25 and IoU (NMS) = 0.45, consistent across all evaluation sets. Training proceeded for 100 epochs with early stopping based on validation loss to minimize unnecessary computation and reduce overfitting. The final model checkpoint (best.pt, 6 MB) was selected based on the highest mean average precision at IoU 0.50 (mAP@50) on the validation set.

To provide a clear overview of the methodological approach, Fig. 1 below illustrates the complete workflow of the AI-based wound detection and classification system. This diagram visually summarizes each key step, from dataset preparation and augmentation through model training, validation, deployment, and application in educational settings.

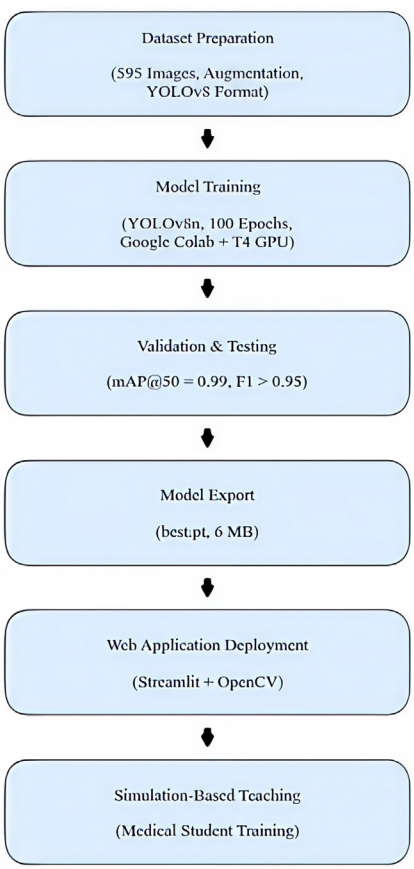


Fig. 1. *Workflow of the AI-based wound detection and classification system.*

Model Variant Comparison

Both You Only Look Once version 8 extra (YOLOv8x) and YOLOv8n models were trained and evaluated on the same dataset. In comparing the two YOLO variants, both models achieved identical mAP@50 scores (0.99), indicating equivalent performance in this evaluation. However, the YOLOv8n model demonstrated superior performance during real-time testing within the web application, likely due to reduced overfitting and more stable predictions across unseen images. Furthermore, YOLOv8n's lightweight architecture (≈ 6 MB) enabled smoother inference speed on CPU -only devices, supporting real-time feedback without hardware acceleration. To quantitatively assess real-world performance, inference testing on standard student hardware (Intel Core i5 CPU, 8 GB RAM) yielded an average latency of approximately 45–60 ms per image, corresponding to 16–22 frames per second (FPS). This confirms that the model maintains smooth real-time visualization and detection performance even without GPU support, making it suitable for classroom or simulation environments. Given the educational aim of this project, where interactivity and accessibility are critical, YOLOv8n was selected for deployment in the final application. This training protocol provided a balanced trade-off between accuracy and deployment feasibility, yielding a lightweight yet high-performing model suitable for real-time inference in a web-based teaching application.

Model Evaluation

The model weights with the lowest validation loss were chosen for the final evaluation. To thoroughly assess performance, the selected model was tested on an independent dataset using several metrics: mean Average Precision (mAP) at Intersection over Union (IoU) thresholds of 0.5 (mAP@50) and a range from 0.5 to 0.95 (mAP@50–95), as well as F1-score, precision, and recall. Performance for each wound type was also analyzed to highlight the model's ability to detect and classify individual categories. Specifically, mAP@50 measures detection accuracy at a single IoU threshold of 0.50, reflecting correct detections with reasonable localization tolerance, while mAP@50–95

averages performance across ten IoU thresholds (0.50–0.95 in 0.05 increments), providing a stricter measure of localization precision. The model achieved mAP@50 = 0.99, indicating high detection accuracy, while the mAP@50–95 = 0.64 reflects the increased challenge of achieving precise bounding-box localization. To ensure internal validity, all wound simulations were prepared by the same forensic medical team using standardized lighting, camera settings, and imaging distance. Each wound type was photographed from multiple perspectives to enhance representativeness and robustness of feature learning. Construct validity was supported by expert verification of class labels by two forensic specialists prior to training. Reliability was established by repeating the model training three times with identical hyperparameters, yielding consistent performance (mAP@50 variation < 0.01).

The reproducibility of results was verified by re-running inference using the exported model (best.pt) on the independent test set and within the deployed web application, confirming stable prediction outcomes.

Integration into Teaching

The trained model was deployed as a web-based application that enabled students to upload images from simulation scene capture photos in real time using their device's camera, or use a live video stream for immediate wound detection and classification. This functionality allowed integration into different educational scenarios, supporting both simulated and live demonstrations with prompt feedback. Confidence scores were displayed alongside predictions to encourage critical thinking and comparison with students' own observations. Facilitators guided discussions around cases where the AI classification differed from the students' assessments, promoting reflective learning.

RESULTS

The YOLOv8n model demonstrated excellent detection performance across all wound categories, with an overall mAP@50 of 0.99 and mAP@50–95 of 0.64. Despite the limited number of physical wound examples (one per type), the model successfully recognized these wound types under different perspectives and lighting conditions, supporting its utility for simulation-based forensic medicine teaching.

Training and Validation Performance

Model convergence was achieved within 100 epochs, with box loss, classification loss, and distribution focal loss steadily decreasing (Fig. 2). On the validation set, the model achieved mAP@50: 0.98, mAP@50–95: 0.65, Precision: 0.95, and Recall: 0.98.

The training curves (Fig. 2.) demonstrated stability without evidence of overfitting, suggesting good generalization capability.

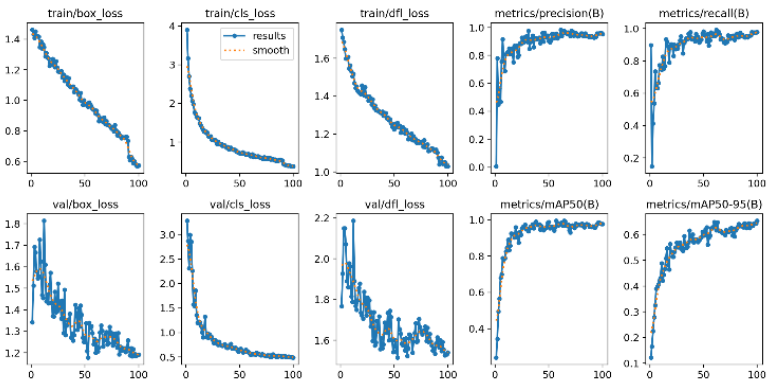


Fig. 2. *Training and validation curves of the YOLOv8n model are shown, depicting training loss curves (box, classification, and distribution focal loss) alongside validation metrics (precision, recall, and mAP), all tracked across 100 epochs.*

Test Set Performance

The YOLOv8n model demonstrated robust detection performance across all eight wound categories following dataset correction and class balancing. Training and validation curves showed stable convergence, with no signs of overfitting. On the independent test set (n = 24 images, 27 instances), the model achieved an overall precision of 0.98, recall of 0.99, mAP@50 of 0.99, and mAP@50–95 of 0.64. The additional evaluation at mAP@75 (0.72) confirmed consistent localization accuracy at stricter IoU thresholds, reflecting the model's robustness in delineating wound boundaries. Per-class performance is summarized in Table I, with all classes achieving near-perfect mAP@50 ($\geqslant$0.99). The lowest mAP@50–95 was observed for

gsw_entrance (0.46), while wound_burn achieved the highest (0.80). The per-class F1-scores ranged from 0.95 to 1.00, and all classes maintained recall ⩾ 0.99, indicating balanced sensitivity across categories. Precision–Recall (PR) curves (Fig. 3) demonstrated that most classes maintained high precision across a wide recall range, confirming strong classifier confidence and minimal false positives. Bootstrap confidence intervals (95%) for AP@50 (Table I) were uniformly narrow ([0.99–0.99] across all classes), signifying high stability of the model's predictions under resampling. Calibration analysis yielded an Expected Calibration Error (ECE) of 0.22, suggesting slight overconfidence in prediction probabilities but within acceptable limits for practical applications, especially in educational and non-clinical contexts. Crucially, after retraining on the corrected dataset, the model demonstrated improved multi-class detection capability, successfully identifying multiple wound types in the same image — a significant improvement over the earlier YOLOv8x model, which occasionally failed to detect all classes when multiple wounds co-occurred in a frame. In real-world testing using the web-based Streamlit application, the model consistently detected multiple wound types in the same image, even under variations in orientation and background.

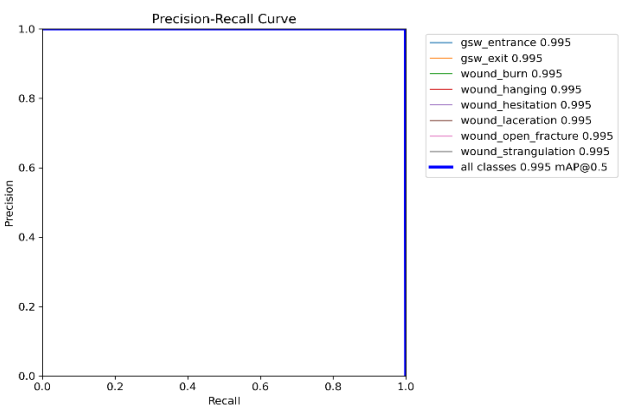


Fig. 3. *Precision–Recall (PR) curves of the YOLOv8n model for each wound class.*

Dataset Distribution

A total of 595 images were included after export from Roboflow, representing eight wound categories. The YOLOv8 dataset visualization reported 569 annotated instances (Fig. 4A). This small discrepancy is likely due to a small number of images being excluded during preprocessing (e.g., malformed or empty label files) and does not affect the proportional representation of classes, because this represents <5% of the dataset, and all classes remained proportionally represented, no further corrections were made for the present analysis. Bounding box dimensions were well distributed across the input image space (Fig. 4B–C), supporting model robustness to variations in object scale and location.

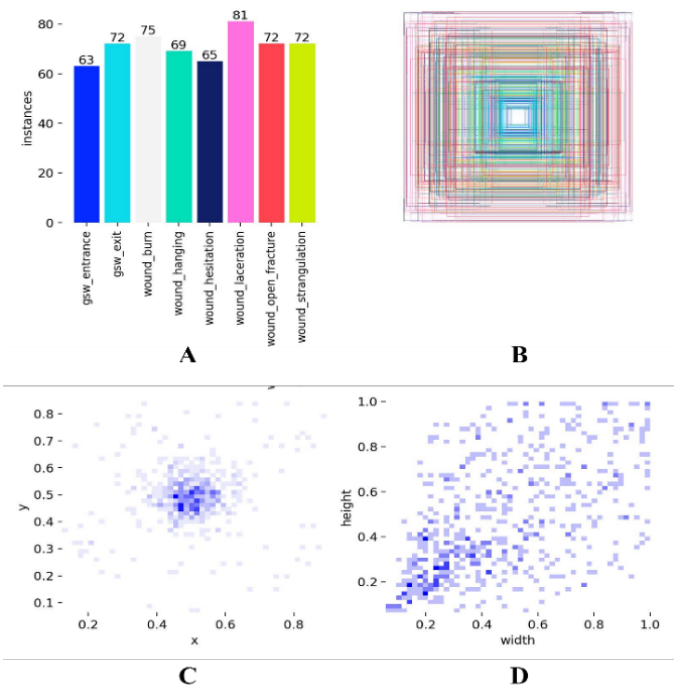


Fig. 4. *Dataset overview visualization generated by YOLOv8. (A) Number of instances per class, showing object count distribution across eight categories; (B) Composite bounding box visualization, illustrating size and aspect ratios; (C) Heatmap of bounding box centers, highlighting object concentration and variability; (D) Bounding box dimension heatmap, indicating the range of scales present in the dataset.*

TABLE I. PER-CLASS DETECTION PERFORMANCE OF THE YOLOV8N MODEL ON THE INDEPENDENT TEST SET (N = 24 IMAGES, 27 INSTANCES).

Class	Images	Instances	Precision (P)	Recall (R)	F1-Score	mAP@50	AP@75	mAP@50–95	95% CI (Bootstrap AP@50)
gsw_entrance	2	2	0.96	1.00	0.98	0.99	0.54	0.46	[0.99–0.99]
gsw_exit	3	3	1.00	0.99	1.00	0.99	0.20	0.50	[0.99–0.99]
wound_burn	1	1	0.91	1.00	0.95	0.99	0.99	0.80	[0.99–0.99]
wound_hanging	5	5	0.99	1.00	0.99	0.99	0.32	0.48	[0.99–0.99]
wound_hesitation	6	6	1.00	0.99	1.00	0.99	0.99	0.73	[0.99–0.99]
wound_laceration	4	4	0.98	1.00	0.99	0.99	0.71	0.68	[0.99–0.99]
wound_open_fracture	3	3	0.98	1.00	0.99	0.99	0.99	0.69	[0.99–0.99]
wound_strangulation	3	3	0.98	1.00	0.99	0.99	0.99	0.77	[0.99–0.99]
Overall	24	27	0.98	0.99	0.99	0.99	0.72	0.64	-

The normalized confusion matrix (Fig. 5) demonstrated perfect classification performance across all eight wound types, with 1.00 along the diagonal, confirming the absence of significant false positives or inter-class confusions.

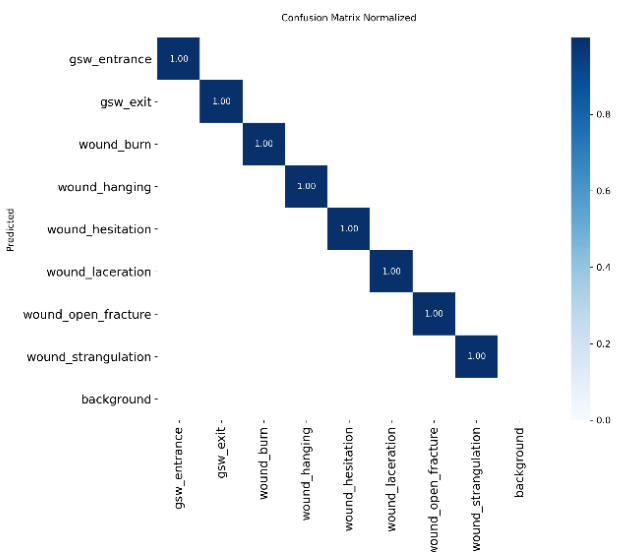


Fig. 5. *Normalized confusion matrix for YOLOv8n wound detection on the independent test set (n = 24 images, 27 instances). Each row represents the true class, and each column represents the predicted class. Darker diagonal cells indicate higher classification accuracy.*

Fig. 6 displays wound type detection images with bounding boxes and YOLO8n confidence scores on the validation set.



Fig. 6. *Example YOLOv8n model predictions on the validation set.*

Sixteen representative images are displayed, each showing a simulated wound on a forensic mannequin or model. Bounding boxes and labels indicate the predicted wound type and confidence score. Predicted classes include wound_hesitation (parallel superficial incisions), wound_strangulation (two patterns of neck compression marks), wound_laceration (irregular tears),

wound_hanging (upward ligature marks), wound_burn (deep burn), and gsw_exit (gunshot exit wounds). The model accurately localized and classified wounds under different lighting conditions and viewing angles, supporting its potential for real-time educational use.

These qualitative results strongly align with the high mAP and F1-scores reported in quantitative analyses, underscoring the reliability and interpretability of the model's outputs. The correspondence between visual examples and metric-based outcomes strengthens confidence in the model's ability to generalize across diverse scenarios. While these results demonstrate strong overall performance, a closer examination reveals areas for further improvement and real-world impact, as discussed below.

Summary of Findings

Overall, these results indicate that the YOLOv8n model provides robust wound type classification in a simulated forensic setting. Its high mAP and precision make it a reliable tool for real-time feedback during simulation-based teaching, potentially improving medical students' recognition of wound characteristics.

IMPLEMENTATION AND APPLICATION

To translate the trained model into a usable educational tool, we developed a cross-platform web application using Streamlit. The application integrates the YOLOv8n model to enable real-time wound detection and classification for simulation-based scene investigation training.

Technical Framework

- **Model:** YOLOv8 (Ultralytics) fine-tuned on eight wound types, deployed using the best.pt weights file.
- **Framework:** Python with Streamlit and OpenCV for video capture and real-time inference.
- **Frontend:** HTML5, CSS, JavaScript for responsive UI.
- **Backend:** Streamlit-based Python server for inference and visualization.
- **Platform:** Cross-platform support (desktop, tablet, mobile) accessible via a standard web browser.
- **Deployment:** Streamlit web application (no local installation required).
- **Open-Source Code:** Available on GitHub for reproducibility and further development (https://github.com/trapezius60/forensic_wound_classification).

Educational Objectives

The application was designed with the following objectives:

1. Enable real-time detection and classification of eight wound types from still images or live webcam video during forensic simulation sessions.
2. Serve as a prototype for future applications in forensic medicine, traffic injury prevention, and medical training.
3. Promote active learning and skill development in machine learning, computer vision, and forensic medicine among medical students.
4. Provide a browser-based, installation-free platform for easy accessibility on multiple devices.
5. Deliver a user-friendly interface suitable for medical students, faculty, and researchers.

Key Features

- Image upload, snapshot capture, and live webcam video detection.
- Bounding boxes and class labels with confidence scores displayed in real-time.
- Short textual descriptions of wound types shown alongside detections for learning reinforcement.
- Central Processing Unit (CPU) processing meets most needs and works flexibly on any device, even mobiles. Web browser access enables use across PCs, tablets, and smartphones for both classroom and remote settings.
- Fast inference speed (depending on internet connection), providing immediate feedback to students.

Fig. 7 presents a sample user interface of the Streamlit application (https://object-detection-yolo8.streamlit.app), illustrating an example of wound detection.

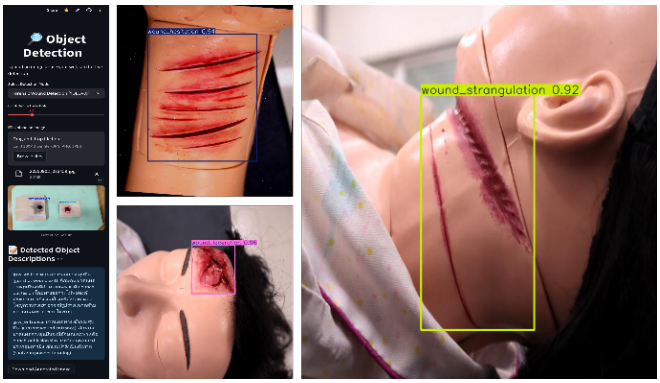


Fig. 7. *Web-based application demonstration of forensic wound detection and classification.*

Left: user interface displaying detected objects, bounding boxes, and brief wound descriptions.

Right: example outputs on simulation mannequins, with bounding boxes and confidence scores overlaid for wound_strangulation (0.92) and other wound types. This real-time feedback supports interactive, simulation-based teaching and allows students to immediately verify and discuss wound detection and classifications during class.

DISCUSSION

The integration of AI-based wound detection into simulation-based teaching offers a novel approach to bridging theoretical learning and hands-on forensic practice [1], preparing them for future medico-legal practice in an increasingly digital healthcare environment. This study demonstrates the feasibility and educational value of applying object detection for forensic wound recognition in simulation-based teaching. A pretrained YOLOv8n model (trained on the COCO dataset) was used as the backbone for feature extraction. Although COCO does not contain wound images, its diverse representation of natural images provides generalizable low- and mid-level visual features (e.g., edges, textures, shapes) that are transferable to wound detection tasks. Fine-tuning on the custom wound dataset allowed the higher network layers to adapt to domain-specific patterns, such as wound contours and tissue characteristics, while retaining robust generic features. This transfer learning strategy is widely adopted in medical image analysis, where data scarcity limits the feasibility of training from scratch, and it significantly accelerated convergence while reducing computational cost [5].

Beyond classification models such as ResNet50 [6], GoogLeNet [7], and ShuffleNet-V2 [8], other architectures have been widely applied in medical imaging, including DenseNet, EfficientNet, MobileNet, and transformer-based approaches such as ViT and Swin-Transformer. Object detection frameworks such as Faster R-CNN, SSD, RetinaNet, and DETR have also gained prominence across medical domains. A recent systematic review of deep learning-based object detection in medical imaging provides a comprehensive overview of these models and their applications, underscoring their adaptability and growing impact in healthcare [9]. Unlike traditional classification models—commonly used in forensic image research—YOLOv8n performs both localization and classification simultaneously, enabling the detection of multiple wound types in a single image. This capability is particularly valuable in forensic simulation scenarios, where students must interpret multiple injuries concurrently, rather than classify one cropped image at a time.

The YOLOv8n model achieved strong detection performance across all wound types, with mAP@50 values approaching 1.00 despite the limited dataset size. The "gsw_entrance" and "wound_hanging" categories had relatively low mAP@50–95 scores (0.46 and 0.48), indicating less precise bounding box localization. This highlights the need for more diverse training data, especially examples with varying lighting, anatomical positions, and rope patterns. Expanding the dataset for wound_hanging could reduce confusion with wound_strangulation and improve detection. A similar approach is recommended for gsw_entrance and gsw_exit. Although the YOLOv8n model achieved excellent real-time performance without further optimization, additional techniques such as quantization or structured pruning may further enhance inference speed and reduce model size. These approaches could be valuable for future deployment on lower-end or mobile devices, broadening accessibility in educational and field applications.

The inclusion of richer evaluation metrics—such as PR curves, per-class AP, bootstrap confidence intervals, and calibration—confirmed that the YOLOv8n model maintained consistent precision and recall across wound categories. Calibration evaluates whether the model's predicted confidence scores accurately reflect the true likelihood of being correct. For example, if the model predicts "wound_hanging: 0.90 confidence," it should be correct approximately 90% of the time. A perfectly calibrated model would have an Expected Calibration Error (ECE) of 0.0; in this study, the observed ECE = 0.22, indicating mild overconfidence yet remaining within acceptable limits for educational applications. These results reinforce the reliability and practical utility of the YOLOv8n model for real-time forensic wound detection and classification in simulation-based teaching environments. Minor discrepancies between the total number of images and annotated instances were observed, likely due to preprocessing exclusions. Although this did not meaningfully affect class balance or model performance, future work should include systematic verification of annotation completeness to further strengthen dataset quality. These results underscore the potential for lightweight deep learning models to function as educational tools, providing instant visual feedback and thereby fostering active learning and student engagement [2]. Furthermore, AI-based crime scene simulation frameworks integrating 3D image processing and semantic segmentation have demonstrated real-time, high-resolution modeling capabilities that enhance forensic training through precise object detection and interactive visualization, supporting immersive and flexible educational experiences even in variable lighting and occlusion conditions [3]. Importantly, the model was deployed as a web-based application, ensuring cross-platform accessibility and eliminating the need for specialized hardware or software installation — a key consideration for resource-limited settings. This study also compared YOLOv8n with YOLOv8x. While YOLOv8x achieved slightly higher mAP@50 during validation, the YOLOv8n model performed better in real-time testing, providing smoother inference and more consistent predictions on CPU-only devices. These findings highlight a critical consideration for educational applications: real-time interactivity and accessibility may be more important than achieving marginally higher offline accuracy [4], especially when students are expected to interact with the system on personal devices during class. Additional research involving comprehensive benchmarking of various detection frameworks such as EfficientDet, Faster R-CNN, DETR (DEtection TRansformer) or other versions of YOLO could yield valuable insights into the balance between computational efficiency and classification robustness within forensic simulation environments.

From an educational perspective, the system enhanced student participation and confidence in wound detection and classification, suggesting that such AI-based tools may help compensate for the limited exposure to real-world forensic cases commonly seen in undergraduate curricula [10]. This mirrors successful applications of AI-assisted learning in other medical fields such as radiology, pathology, and dermatology, where automated classification systems have been shown to improve diagnostic reasoning and stimulate peer discussion [11][12]. However, this study has several important limitations. First, each wound type was represented by only a single physical simulation, albeit captured under multiple angles and lighting conditions. As a result, the model's performance primarily reflects its ability to generalize to variations of the same wound instance rather than to entirely new cases. Additional wound simulations, with different anatomical locations, weapons, skin tones, and severity levels, should be included in future datasets to improve model generalizability, reduce class-specific bias [13], and optimizing calibration through temperature scaling or ensemble methods. Although the model achieved strong generalization under stretch-mode grayscale preprocessing, further ablation studies comparing color vs. grayscale and letterbox vs. stretch transformations would clarify their influence on fine-grained texture representation, calibration, and inference latency. Second, although the validation and test results were robust, formal educational outcome evaluation has not yet been completed. Future research should include pre- and

post-intervention assessments of student learning, as well as usability surveys, to quantify the pedagogical impact of AI-assisted teaching [14]. Furthermore, comparative studies could evaluate whether students trained with AI feedback outperform those trained using traditional methods in objective structured clinical examinations (OSCEs) or written assessments [15]. In addition, future evaluations should incorporate automation bias analysis to assess whether AI-generated feedback influences students' independent clinical judgment and diagnostic reasoning. Finally, the current model has not been tested under suboptimal conditions such as poor lighting, partial occlusion, or motion blur, which may occur during real-world scene investigations. Incorporating such challenging scenarios into the dataset and exploring data augmentation strategies may further enhance robustness [16]. Importantly, the wound detection model developed in this study is intended solely as an educational aid to support interactive learning and case-based discussion. It is not designed or validated for clinical diagnosis or forensic adjudication. All simulations and interpretations are performed under instructor supervision, ensuring that human expertise remains central to decision-making. This human-in-the-loop approach aligns with the principles of responsible AI, emphasizing transparency, accountability, and pedagogical safety in educational settings.

CONCLUSION

This study demonstrates the feasibility and educational value of deploying an AI-based object detection model for forensic wound detection and classification in a simulation-based teaching environment. The YOLOv8n model, selected for its speed and deployment efficiency, provided high-accuracy classification of eight wound types and enabled real-time feedback in a browser-based application accessible across devices. The integration of this system into the curriculum represents a novel educational innovation that bridges theoretical knowledge and practical skills while exposing students to applied AI technology. Despite the dataset's limitations, the model's strong performance supports its use as a teaching aid to enhance exposure to diverse wound patterns and improve diagnostic reasoning. Future work will focus on expanding the dataset, integrating student evaluation results, and assessing measurable learning outcomes. If proven effective, this approach could be scaled to other medical schools and extended to additional forensic or trauma-related training modules, contributing to a new paradigm of AI-augmented medical education.

ACKNOWLEDGMENT

The author gratefully acknowledges the invaluable assistance of the forensic team for their support during the data collection and simulation preparation phases of this project. In particular, sincere thanks are extended to Mr. Thanaphong Klumkrathok, Mr. Uthen Thubkrathok, Mr. Chakkree Petchrak, and Mr. Anatchai Rattana for their dedicated efforts in photographing wound specimens during the pre–data processing stage and in preparing forensic mannequins used for student learning and simulation-based instruction. Their expertise and collaboration were essential to the success of this study.

REFERENCES

[1] V.A. Cărcălе, "The future of artificial intelligence applications in forensics," rais.education, 2025. [Online]. Available: https://rais.education/wp-content/uploads/2025/05/0523.pdf.

[2] S. Prithwish and R. Pinki, "A lightweight end-to-end system for wound tissue analysis," Sci. Direct, 2025. [Online]. Available: https://www.sciencedirect.com/science/article/abs/pii/S1746809425002459.

[3] O. Hajare and M. Thalor, "AI-based crime scene simulation through 3D image processing and semantic segmentation," IJSRET, vol. 11, no. 2, pp. 1576–1580, Mar.–Apr. 2025. [Online]. Available: https://ijsret.com/wp-content/uploads/2025/03/IJSRET_V11_issue2_552.pdf.

[4] J. Jocher et al., "Explore Ultralytics YOLOv8," 2025. [Online]. Available: https://docs.ultralytics.com/models/yolov8/.

[5] F. Vazquez, J. A. Nu⊛ez, X. Fu, P. Gu, and B. Fu, "Exploring transfer learning for deep learning polyp detection in colonoscopy images using YOLOv8," arXiv preprint arXiv:2502.00133, Jan. 2023. [Online]. Available: https://arxiv.org/html/2502.00133v1

[6] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), 2016, pp. 770–778. [Online]. Available: https://arxiv.org/abs/1512.03385

[7] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), 2015, pp. 1–9. [Online]. Available: https://arxiv.org/abs/1409.4842

[8] N. Ma, X. Zhang, H.-T. Zheng, and J. Sun, "ShuffleNet V2: Practical guidelines for efficient CNN architecture design," in Proc. Eur. Conf. Comput. Vis. (ECCV), 2018, pp. 116–131. [Online]. Available: https://arxiv.org/abs/1807.11164New

[9] C. Albuquerque, D. M. de Oliveira, A. Oliveira, and G. Rodrigues, "Deep learning-based object detection algorithms in medical imaging: A review," Heliyon, vol. 10, no. 9, e27379, 2024. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S240584402417168X

[10] J. Smith, "Simulation-based education in the artificial intelligence era," Cureus, vol. 15, 2023. [Online]. Available: https://www.cureus.com/articles/161951-simulation-based-education-in-the-artificial-intelligence-era.

[11] E. Topol, "Applications of artificial intelligence in medical education: a systematic review," Cureus, 2025. [Online]. Available: https://www.cureus.com/articles/334959-applications-of-artificial-intelligence-in-medical-education-a-systematic-review.

[12] S. Hayden, B. Travis, and S. Benjamin K., "Ethics of artificial intelligence-assisted image interpretation in dermatology and radiology," NCBI, 2025. [Online]. Available: https://pmc.ncbi.nlm.nih.gov/articles/PMC11889981/.

[13] V. Vidit, M. Engilberge, and M. Salzmann, "A single domain generalization approach for object detection," in Proc. CVPR, 2023. [Online]. Available: https://openaccess.thecvf.com/content/CVPR2023/papers/Vidit_CLIP_the_Gap_A_Single_Domain_Generalization_Approach_for_Object_CVPR_2023_paper.pdf.

[14] R. D. A. Budiman et al., "Effectiveness of AI-driven assessments in enhancing learning outcomes," Int. J. Inform. Educ. Technol., vol. 15, no. 7, 2025. [Online]. Available: https://www.ijiet.org/vol15/IJIET-V15N7-2342.pdf.

[15] D. Lattas, C. Davis, C. Creamer, and G. Jefferies, "The use of simulation in forensic social work education: a scoping review," J. Soc. Work Educ., 2024. [Online]. Available: https://journals.sagepub.com/doi/abs/10.1177/14999013241301097.

[16] M. Hu, Z. Li, J. Yu, X. Wan, H. Tan, and Z. Lin, "Efficient lightweight YOLO: improving small object detection for robustness under real-world conditions," Cureus, 2023. [Online]. Available: https://pmc.ncbi.nlm.nih.gov/articles/PMC10385816/.