VenaPath Guardian: Multimodal Vision and Uncertainty-Aware Augmented Guidance for Clinician-in-the-Loop Hand Peripheral IV Cannulation

A. Maria, Immaculate College for Women, Cuddalore, mariasamy23@gmail.com

Abstract—Peripheral intravenous (PIV) cannulation is among the most common invasive clinical procedures, yet first-attempt failure remains frequent and can be clinically consequential, particularly for patients with difficult intravenous access (DIVA). Multiple attempts increase pain and anxiety, bruising and hematoma, infiltration/extravasation risk, staff workload, and time-to-therapy. This paper presents *VenaPath Guardian*, a real-time decision-support system that augments clinician judgement by enhancing hand-vein visibility, segmenting candidate veins, and overlaying conservative puncture-site guidance on a live camera preview. The approach combines RGB context and near-infrared (NIR) vessel contrast (with optional depth/ToF and thermal inputs) and is explicitly non-autonomous: it does not puncture, does not prescribe an action, and can withhold recommendations when information is unreliable. Two safety mechanisms are central to the design: (i) *quality and contraindication gating* that detects blur, glare, motion, and avoid-regions (e.g., bruising or dressings) and returns actionable acquisition prompts; and (ii) *uncertainty-aware segmentation* that estimates model confidence and triggers abstention under ambiguity. Vein topology is extracted as centerline graphs, and a learning-to-rank model scores feasible candidate segments to produce interpretable top-k suggestions (entry point, tolerance radius, and an insertion-angle band). We also describe a multi-center dataset and evaluation protocol with demographic and skin-tone diversity, subgroup reporting, and prospective endpoints such as first-attempt success and time-to-cannulate.

Index Terms—Peripheral intravenous access, difficult IV access (DIVA), near-infrared imaging, vein segmentation, uncertainty estimation, learning-to-rank, augmented reality guidance, clinical decision support.

I. INTRODUCTION

Peripheral intravenous (PIV) access enables infusion therapy across emergency, inpatient, outpatient, and infusion-center workflows. While the task is routine for many patients, it becomes challenging when veins are poorly visible or fragile. DIVA is commonly associated with pediatric anatomy, older age, dehydration, obesity/edema, low perfusion states, and situations where visible-light contrast is reduced (including darker skin tones or challenging illumination). When access is difficult, repeated needle insertions delay care and increase patient discomfort and complication risk.

Near-infrared (NIR) imaging often improves superficial vessel visibility because blood absorbs NIR light differently than

Maria is with Immaculate College for Women, Cuddalore, India (email: mariasamy23@gmail.com).

surrounding tissue. However, visibility alone is not the same as safe and efficient cannulation. Clinicians must still choose a puncture site that is (i) stable under motion, (ii) away from contraindicated regions such as bruising or suspected infection, and (iii) compatible with a feasible needle trajectory. The decisions are time-sensitive and must be made under imperfect images (motion blur, glare, occlusion, and variable lighting).

This paper proposes *VenaPath Guardian*, a clinician-in-the-loop guidance system that turns multimodal imaging into *conservative decision support*. The system outputs an interpretable overlay when it is confident, and otherwise abstains with clear acquisition prompts. Unlike autonomous robotics, the clinician retains full control over site selection and insertion.

A. Contributions

VenaPath Guardian is motivated by three design principles: interpretability, abstention under uncertainty, and robustness across populations. The contributions of this paper are:

- A modular multimodal pipeline for hand-vein enhancement, segmentation, topology extraction, and ranking-based puncture-site suggestion with calibrated confidence.
- A safety layer that combines measurable image-quality checks, contraindication detection, and uncertainty-based abstention.
- An interpretable overlay design (entry point, angle band, confidence indicator, and reason codes) suitable for clinical decision support.
- A dataset and evaluation protocol emphasizing demographic diversity (including Fitzpatrick skin types I–VI), robustness stress tests, and prospective clinical endpoints.

II. RELATED WORK

Thin-structure segmentation is commonly addressed with encoder–decoder architectures such as U-Net [1] and DeepLab variants [2], which preserve multi-scale context while retaining fine detail. In safety-critical applications, model confidence must be communicated and used for risk control; uncertainty estimation methods such as Monte Carlo dropout and Bayesian approximations [3] and deep ensembles [4] provide tools for measuring epistemic uncertainty.

Vessel enhancement filtering (e.g., multi-scale vesselness) has a long history in medical imaging [5] and remains a strong
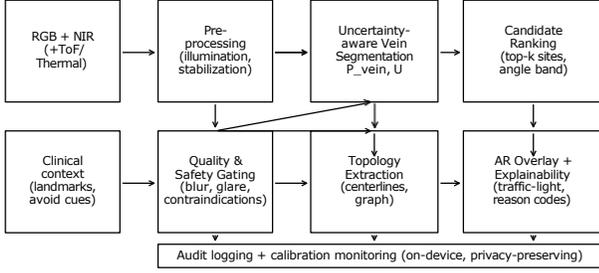
Fig. 1. System overview of VenaPath Guardian. Quality gating, contraindication detection, and uncertainty estimation can suppress guidance under low-quality or unsafe conditions.

baseline for vein highlighting when training data are limited. For the decision-support component, ranking is a natural formulation when multiple candidates are plausible. Learning-to-rank methods from information retrieval provide practical losses and robust inference; in particular, gradient-boosted ranking has strong performance and interpretability properties (e.g., feature importance and monotonic constraints).

Finally, AR overlays for procedural assistance require careful calibration and latency control. For live video, stabilization and optical flow methods such as RAFT [6] support temporal coherence, which is critical to prevent overlay drift that could mislead users.

III. SYSTEM OVERVIEW

Fig. 1 summarizes the proposed architecture. RGB and NIR streams provide complementary information: RGB supports context and contraindication cues (bruising, redness, dressings, scarring) while NIR improves vessel contrast. Optional depth/ToF improves geometric stabilization and supports feasibility constraints for insertion angles; thermal can provide a coarse perfusion proxy. The pipeline is explicitly conservative: quality and safety checks can suppress all recommendations and return only prompts and reason codes.

A. Sensing Modalities

Table I summarizes intended modality roles and limitations. The system is designed to degrade gracefully when optional modalities are absent (e.g., RGB+NIR only).

IV. METHODS

The system takes a live hand video stream and outputs either: (i) an augmented overlay containing enhanced veins, ranked candidate puncture segments, and a recommended entry point with an insertion-angle band, or (ii) an abstention state with acquisition prompts (e.g., remove glare, stabilize, adjust lighting) and reason codes.

A. Preprocessing and Stabilization

Preprocessing reduces nuisance variation due to lighting and motion. For NIR, contrast-limited adaptive histogram equalization can improve local vessel contrast without saturating

TABLE I
PROPOSED SENSING CONFIGURATIONS AND PRACTICAL
CONSIDERATIONS.

Modality	Primary role	Notes / limitations
RGB video	Context and contraindication cues	Sensitive to lighting and glare; supports bruise/redness/dressing detection and landmarks.
NIR imaging	Superficial vein contrast	Contrast may drop for deeper veins or low perfusion; requires illumination and camera calibration.
Depth/ToF (optional)	Surface geometry for stabilization	Can reduce overlay drift and help estimate feasible insertion angles; can be noisy on reflective/dark surfaces.
Thermal (optional)	Perfusion proxy	Ambient-dependent and often low-resolution; treated as a weak auxiliary signal.
Ultrasound (subset)	Validation/labeling	Provides depth/diameter ground truth; used on a subset rather than required at inference.

highlights; for RGB, photometric normalization can reduce illumination changes. Motion stabilization aligns frames using feature tracking or optical flow, while hand keypoints provide a canonical coordinate system to stabilize the overlay over time and to define exclusion zones near joints and creases.

B. Quality and Contraindication Gating

A hard-gating layer computes measurable quality metrics:
- *Blur*: focus/edge-energy or Laplacian variance.
- *Glare*: detection of saturated regions and specular highlights.
- *Exposure*: under/over-exposure checks.
- *Motion*: frame-to-frame displacement magnitude.
- *NIR contrast*: vessel-to-background contrast proxy.

In parallel, a lightweight detector identifies avoid-regions such as bruising/hematoma-like discoloration, erythema suspicious for infection, dressings, and occlusions. When any gate fails, the system abstains and returns an acquisition prompt. This behavior is not a failure case; it is a safety feature.

C. Uncertainty-Aware Vein Segmentation

A convolutional encoder–decoder network predicts a per-pixel vein probability map $P_{\text{vein}}(x)$ from the multimodal input x. For multimodal fusion, two configurations are considered:

1) *Early fusion*: stack RGB and NIR channels at the input (simple and fast).
2) *Mid-level fusion*: separate modality encoders with feature concatenation or attention (more robust to missing modalities).

To support conservative decision support, the model estimates uncertainty $U(x)$ using Monte Carlo dropout or deep ensembles [3], [4]. Let Ω be a candidate region (e.g., the dorsum ROI). The system computes a summary uncertainty statistic (e.g., mean or 95th percentile) and abstains if it exceeds a threshold:

$$\text{abstain} \quad \text{if} \quad \text{Agg}\left(U(x) \mid \Omega\right) > \tau. \tag{1}$$

Threshold τ is chosen on validation data to trade off coverage and risk.

D. Losses and Calibration

Segmentation is trained with a combination of overlap- and imbalance-aware losses, for example,

$$L_{seg} = \lambda_{Dice}(1 - Dice) + \lambda_{Focal}\, Focal. \qquad (2)$$

Because the user interface communicates confidence, calibration is treated as a first-class objective. Post-hoc temperature scaling or isotonic regression can be used to align predicted probabilities to empirical accuracy, and calibration is monitored across demographic strata.

E. Topology Extraction and Candidate Generation

From high-confidence pixels in $P_{vein}(x)$, a centerline is extracted via skeletonization and converted into a graph $G = (V, E)$ with nodes (endpoints/bifurcations) and edges (candidate vein segments). For each edge, features are computed:

- geometric features (length, curvature/tortuosity, branching proximity),
- a diameter proxy from the distance transform around the centerline,
- anatomical distances to landmarks and exclusion zones,
- temporal stability across a short time window,
- uncertainty statistics along the segment.

Candidate puncture points are sampled along each segment while masking avoid-regions and high-uncertainty pixels.

F. Learning-to-Rank for Puncture-Site Suggestions

Selecting a single "best" puncture point is inherently preference-based and depends on context. Therefore, the system outputs a *ranked list* and communicates uncertainty. A learning-to-rank model scores candidates c with features $\phi(c)$:

$$S(c) = f(\phi(c)), \qquad (3)$$

where $f(\cdot)$ can be a gradient-boosted ranker (e.g., LambdaMART-style objectives) or a neural ranker. Training labels include expert-preferred top-1 sites and acceptable top-3 sites, enabling pairwise or listwise ranking losses. Additional constraints can be imposed (e.g., monotonicity: closer to joint crease $\Rightarrow$ lower score).

G. Augmented Overlay and Explainability

Fig. 2 illustrates the inference flow including abstention. When the system does provide guidance, the overlay includes: (i) enhanced veins, (ii) centerlines with uncertainty indication, (iii) a recommended entry point with tolerance radius, and (iv) an insertion-angle band that reflects feasible trajectories under the current pose and sensor geometry. To avoid over-trust, the UI uses traffic-light confidence and reason codes that explain why the recommendation was selected or why it was withheld.
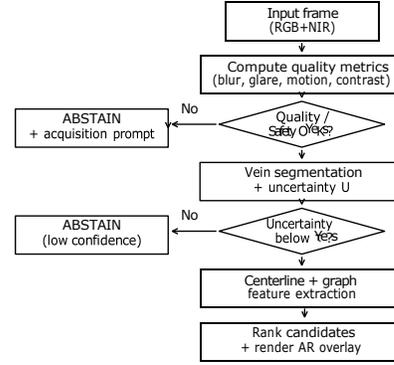

Fig. 2. Inference flow with abstention. Quality/safety checks and uncertainty thresholds can suppress guidance and return acquisition prompts.

V. DATASET AND ANNOTATION PROTOCOL

Robust clinical performance requires diversity in capture conditions and patient characteristics. The proposed dataset plan is multi-center and explicitly stratified by age, sex, and Fitzpatrick skin types I–VI, with intentional enrichment for DIVA scenarios (e.g., edema, low perfusion, obesity, dehydration). Each sample includes synchronized RGB and NIR video of the dorsum of the hand under varied conditions: pose changes, different lighting, and tourniquet/no-tourniquet settings. Optional depth/ToF and thermal channels can be collected where available. A subset of cases can be paired with ultrasound for validation of depth and diameter, but ultrasound is not assumed to be present at inference.

Annotations align to the end-to-end tasks:

- *Vein masks*: pixel-level binary labels, optionally with soft boundaries.
- *Centerlines and graphs*: skeletons with bifurcations/endpoints.
- *Puncture sites*: preferred and acceptable entry points with tolerance radii.
- *Angle bands (optional)*: ranges/bins reflecting recommended insertion angles.
- *Avoid-regions*: masks for bruising, redness suspicious for infection, dressings, scars, and exclusion zones near creases.
- *Clinical outcomes (prospective)*: attempts, success, time-to-cannulate, pain score, and complications.

Because puncture-site choice is preference-based, inter-rater variability should be measured and included in evaluation (e.g., agreement distributions rather than single-point "truth").

VI. EVALUATION PLAN

The evaluation protocol separates technical model performance from clinical utility, and treats abstention as a core outcome.

A. Offline Technical Evaluation
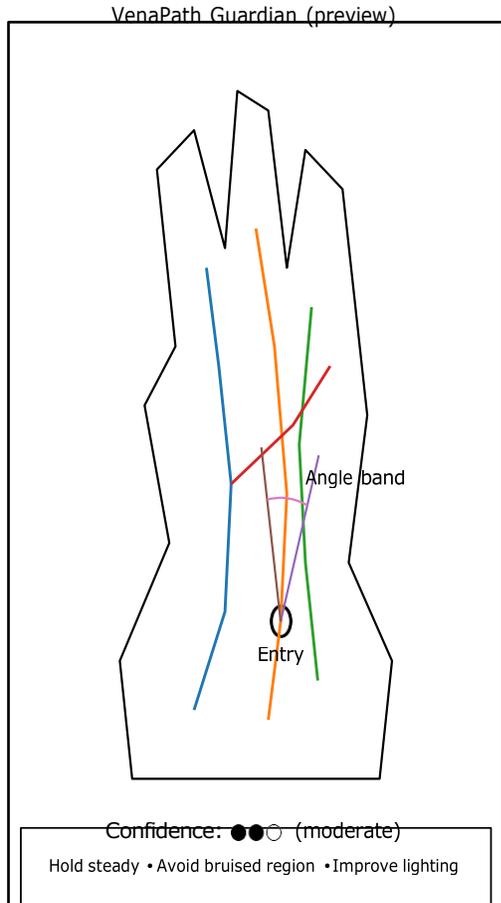
Key technical measures include:

Fig. 3. Illustrative augmented overlay concept for clinician decision support. The interface shows an entry point, an insertion-angle band, and confidence prompts; clinicians remain in control.

- *Segmentation*: Dice/IoU for pixel overlap; additional thin-structure measures for continuity.
- *Topology*: centerline distance (mean/95th percentile) and connectivity/fragmentation.
- *Recommendation*: top-1 and top-3 agreement with expert sites under a tolerance radius; coverage-aware performance when abstention is allowed.
- *Robustness*: stress tests under motion blur, low light, glare, occlusion, and edema-like appearance; error taxonomy.

All metrics should be reported overall and stratified by skin type and key DIVA factors. Importantly, subgroup reporting must include abstention rates, since disparate abstention can reflect inequitable access to tool support.

B. Prospective Clinical Evaluation

A prospective study can compare two arms: clinician-only versus clinician-plus-tool decision support. Primary endpoints may include first-attempt success rate and time-to-cannulate. Secondary endpoints may include number of at-tempts, pain scores, bruising/hematoma incidence, and infiltration/extravasation. Operating thresholds (quality gates and τ) should be pre-registered, and audit logs should support post hoc analysis of failure modes and drift.

VII. Discussion

A. Safety, Ethics, and Governance

VenaPath Guardian is designed as conservative clinical decision support. The system is intended to *withhold* guidance when information is unreliable. This differs from typical "always-on" overlays and is more aligned with selective prediction in high-risk domains. For governance, versioned models, calibration monitoring, and audit trails are essential, especially if the system is deployed across multiple sites and camera configurations.

Privacy and data protection can be supported by on-device inference, minimizing storage of identifiable imagery, and encrypting any logs required for monitoring. If deployed as software as a medical device (SaMD), risk management, usability/human factors testing, and cybersecurity practices must accompany performance validation.

B. Limitations

Surface imaging cannot reliably infer vein depth for all patients; therefore, depth-related cues should be treated as conservative ranges and validated against ultrasound on a subset. In severe edema or low perfusion, even NIR contrast may be inadequate; abstention is preferable to overconfident recommendations. Finally, the "best" puncture site depends on therapy type, catheter gauge, and clinician preference; the ranker must be adaptable and should expose its reason codes to support clinician trust and override.

VIII. Conclusion

This paper presented VenaPath Guardian, an uncertainty-aware multimodal vision system for conservative guidance of hand PIV cannulation. By combining RGB+NIR sensing, thin-structure segmentation, topology-based candidate generation, learning-to-rank decision support, and explicit quality gating with abstention, the approach aims to reduce repeated attempts while keeping clinicians fully in control. Future work includes multi-center data collection, robust subgroup analysis, and prospective clinical validation.

References

[1] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. MICCAI*, 2015.

[2] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs," *IEEE Trans. Pattern Anal. Mach. Intell.*, 2018.

[3] A. Kendall and Y. Gal, "What uncertainties do we need in Bayesian deep learning for computer vision?" in *Proc. NeurIPS*, 2017.

[4] B. Lakshminarayanan, A. Pritzel, and C. Blundell, "Simple and scalable predictive uncertainty estimation using deep ensembles," in *Proc. NeurIPS*, 2017.

[5] A. F. Frangi, W. J. Niessen, K. L. Vincken, and M. A. Viergever, "Multiscale vessel enhancement filtering," in *Proc. MICCAI*, 1998.

[6] Z. Teed and J. Deng, "RAFT: Recurrent all-pairs field transforms for optical flow," in *Proc. ECCV*, 2020.