Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Surgical Vision ROS2: a ROS2 Humble pipeline for intraoperative wound assessment and variable-autonomy arm control

Lamia Islam

lamia.210611@s.pust.ac.bd

Pabna University Of Science and Technology

Research Article

Keywords: ROS2, surgical robotics, wound assessment, image segmentation, variable autonomy, MoveIt2

Posted Date: July 16th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10362132/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: The authors declare no competing interests.

Abstract

Surgical Vision ROS2 is a ROS2 Humble pipeline for intraoperative wound assessment. A vision node scores a wound using a five-part ABCDE framework adapted from dermatology. A behaviour manager converts the score into one of three autonomy tiers for a MoveIt2-controlled arm: full execution, reduced-speed execution with per-motion confirmation, and a hold state in which the arm freezes until the operator types CONFIRM. On a 45-image synthetic benchmark the original grayscale adaptive-threshold segmentation reached 6.7% three-class verdict accuracy. Inspection showed the method was locking onto suture stitches. A replacement HSV colour-distance segmentation with Otsu binarisation raised accuracy to 64.4% on the same images.

1. Motivation and significance

Teleoperated surgical robots execute commands issued by a human operator. In most systems the arm has no independent view of the tissue it is acting on, and it does not modify its own behaviour based on what the camera sees at the surgical site. The operator is the only party in the control loop that inspects the wound, judges its state, and decides how the arm should move. This works well when the operator is experienced and the tissue is behaving normally. It is a weaker arrangement in training and simulation settings, where trainees may miss visual cues that a supervisor would catch, and in long procedures where operator attention fades.

An intraoperative monitoring layer that watches the wound independently and constrains what the arm is allowed to do would add a second, cheaper judgement to the control loop. That second judgement does not need to be as accurate as a trained surgeon. It only needs to catch the cases the operator misses and slow the arm down when something looks wrong. The surgical data science community has argued for perception-conditioned assistance of this general shape [1], but few open-source implementations exist that a robotics researcher can pick up and modify.

Surgical Vision ROS2 is one such implementation. The software takes a wound image, computes a five-part clinical-style score, converts the score into one of three autonomy tiers, and enforces the tier on a MoveIt2-driven arm [3]. The pipeline is built on ROS2 Humble [2] and OpenCV [4]. The scoring uses the ABCDE lesion-assessment framework from dermatology [5], adapted for surgical wounds. The five criteria are Asymmetry, Border sharpness, Colour, Diameter, and Evolution over time.

The software is built as a domain retargeting of the DecodeLabs Robotics & Automation Industrial Training Kit (Project 2), which was originally written for gear defect detection. The four-node ROS2 architecture and the general Input-Process-Output structure are inherited from that kit. The vision logic, the autonomy tier definitions, and the wound generator are new to this work.

The contribution of the software is in two parts. The first is the runnable pipeline itself: a four-node ROS2 graph that couples wound perception to arm-autonomy gating, with clean topic-level separation between perception and control. Any of the four nodes can be replaced without touching the others. The second

contribution is empirical. When the pipeline was evaluated on a controlled 45-image synthetic benchmark, the originally implemented segmentation was found to fail badly, not by a small margin but by an order of magnitude in accuracy. The failure was traced to a specific property of the segmentation method. A replacement method was implemented and validated on the same benchmark. The improvement is not clinical in scale, and the fixed pipeline still misclassifies one whole class. Both the working parts and the residual failure are reported here.

2. Software description

2.1. Software architecture

The software is organised as four ROS2 nodes connected in a linear producer-consumer chain (Fig. 1). The nodes and their behaviour are as follows.

lesion_detector_node reads an image from a directory that simulates a live camera feed. It calls analyze_wound in abcde_analyzer.py to compute the five ABCDE sub-scores. It publishes a std_msgs/String message on the topic /scene_state in the format VERDICT|SCORE|REASON, for example FAIL|4|Asymmetry irregular, Tissue color abnormal.

behavior_manager_node subscribes to /scene_state. It maps the verdict string to an autonomy tier using a fixed lookup table. PASS maps to the FULL tier at speed factor 1.00, CAUTION maps to REDUCED at speed factor 0.50, and FAIL maps to HOLD at speed factor 0.00. The result is published on /autonomy_mode in the format MODE|SPEED|REASON.

surgeon_console_node subscribes to /autonomy_mode. It implements the operator-facing gating logic. In FULL mode the node passes arm-command intents through. In REDUCED mode it prompts the operator for confirmation before each motion. In HOLD mode it blocks all motion and waits for the operator to type the literal token CONFIRM before allowing motion to resume. Approved commands are published on /arm_command.

moveit_bridge_node subscribes to /arm_command. It forwards the requested motion to MoveIt2 [3] for planning and execution, and MoveIt2 in turn drives the arm and updates the RViz visualisation.

The four nodes exchange only string messages on standard ROS2 topics. There is no shared state, no direct method calls between nodes, and no locking. Each stage can be run in isolation with a mocked upstream publisher for testing. In principle, either the perception node or the arm-control nodes could be replaced independently. In practice, the perception node has more coupling to the analyzer module (abcde_analyzer) than to the ROS2 layer, so replacing the analyzer is a Python-level task rather than a ROS2 task.

The architecture has known limitations. There is no watchdog on /scene_state: if the vision node stops publishing, the behaviour manager continues to publish the last autonomy mode indefinitely rather than

falling back to a safe default. There is no message versioning or explicit schema. All messages are std_msgs/String with a pipe-separated payload, which is convenient during prototyping but should be replaced with a defined message type before any use beyond a testbed.

2.2. Software functionalities

ABCDE scoring. The vision logic in abcde_analyzer.py computes five per-criterion sub-scores, each in the range 0 (normal), 1 (mild), or 2 (severe). The sub-scores are summed to a total in the range 0–10, and the total is thresholded to a verdict. Table 1 lists the criteria and the parameters used.

Table 1
ABCDE criteria and their default thresholds in abcde_analyzer.py.

Criterion	Metric	Method	Threshold(s)
A: Asymmetry	Contour solidity (area / hull area)	Convex-hull ratio	ASYMMETRY_THRESHOLD = 0.82; sub-score 0 if solidity ≥ 0.82, 1 if ≥ 0.67, 2 otherwise
B: Border	Sharpness of an 8-pixel strip around the contour	Laplacian variance	BORDER_SHARP_MIN = 18.0; sub-score 0 if variance ≥ 36.0, 1 if ≥ 18.0, 2 otherwise
C: Colour	Fraction of contour interior classified as healthy pink	HSV pixel classification	Sub-score 0 if healthy-pink ratio ≥ 0.25 and necrotic ratio ≤ 0.15, 1 if only pink is low, 2 if necrotic is high
D: Diameter	Minimum enclosing circle diameter (px)	Compared against target 180 px	DIAMETER_TOLERANCE = 0.30; sub-score 0 if within ± 30%, 1 if within ± 60%, 2 otherwise
E: Evolution	Frame-to-frame percent change in contour area	Comparison against previous area	EVOLUTION_ALERT_PCT = 0.25; sub-score 0 if change ≤ 12.5%, 1 if ≤ 25%, 2 otherwise

The verdict is assigned by determine_verdict(total) as PASS for a total of 0 or 1, CAUTION for a total of exactly 2, and FAIL for any total of 3 or above. This is the mapping actually used by the code and reflected in the published results. The README in the repository documents a wider CAUTION band. That wider band appears in a lookup table in the source but is not currently invoked, and should be reconciled in a future revision.

Autonomy gating. The three tiers implement a monotonic reduction of arm authority as the vision score worsens. In the FULL tier the arm behaves as a normal teleoperated system. In REDUCED the arm still executes commands but at half speed and with an explicit per-motion sign-off, which slows the operator and forces them to look at the arm before each move. In HOLD the arm is frozen and the operator has to type CONFIRM to lift the freeze. The speed factors and the CONFIRM token are hard-coded constants and are intended to be tuned per hardware setup.

Original segmentation. All five ABCDE metrics are computed on a binary wound mask. The originally implemented preprocessing pipeline abcde_analyzer.preprocess converts BGR to grayscale, then applies a Gaussian blur with a 7×7 kernel and σ = 2. It then applies cv2.adaptiveThreshold with ADAPTIVE_THRESH_GAUSSIAN_C, THRESH_BINARY_INV, block size 31, and constant C = 6. A morphological close (5×5 elliptical kernel, 2 iterations) is applied, followed by a morphological open (same kernel, 1 iteration). The largest external contour of the resulting mask was taken as the wound..

Replacement segmentation. The replacement pipeline is a global colour-distance segmentation. The input BGR image is converted to HSV. The background/drape colour is sampled from the four image corners (each corner is a 20×20 patch). The per-pixel Euclidean distance from the median background colour of those patches in HSV space is computed. Otsu's method is applied to the resulting distance map to produce a binary foreground mask. A morphological close (13×13 elliptical kernel, 2 iterations) fills small internal holes in the foreground. A morphological open with the same kernel (1 iteration) then removes thin high-frequency clutter, such as the suture stitches that broke the original method. The largest external contour of the cleaned mask is taken as the wound.

The two segmentations are drop-in alternatives at the level of abcde_analyzer.preprocess. The rest of the analyzer, all five ABCDE metrics, and the ROS2 nodes are unchanged.

3. Illustrative examples

Synthetic benchmark

The evaluation uses 45 synthetic wound images generated by the script generate_wounds.py included in the repository. The generator is deterministic under a fixed NumPy PCG64 seed (default_rng(2024)). Each image is 512×512 pixels, JPEG. The generator produces three classes, 15 images each.

Healthy images contain a centred ellipse with semi-axes rx = 90 and ry = 85 pixels, giving an analytic area of π × 90 × 85 ≈ 24,032 px². The tissue is a uniform healthy-pink fill (BGR base 100, 130, 210) with per-pixel additive noise in ± 14 and a fine specular-glint overlay. A sharp elliptical outline is drawn on the boundary, and eight evenly spaced suture stitches are drawn at radius rx + 8.

Concerning images contain an irregular closed polygon around a mean radius of 100 pixels with 22% radial irregularity, and one or two pale ischemic patches inside the wound. The border is deliberately blurred, using a Gaussian blur of the polygon outline with an 11-pixel kernel. Sutures are drawn at radius 108.

Critical images contain a more irregular polygon around a mean radius of 130 pixels with 48% radial irregularity, a central necrotic-dark zone, and several fibrinous slough patches. An erythema ring sits outside the main wound, along with three to five satellite tissue spots simulating dehiscence.

The benchmark is small (n = 45) and single-seed. It has value as a controlled, reproducible sanity check with a known ground-truth class label per image, and, in the healthy case, a known analytic wound area.

It is not a substitute for evaluation on realistic or clinical imagery.

Single-image diagnostic

The initial diagnostic was run on healthy_01.jpg, for which the analytic wound area is known. Under the original adaptive-threshold pipeline, the largest external contour of the segmentation mask had an area of 804 px². Its bounding box, 148 × 54 pixels, was centred on the lower half of the sutured ring rather than on the wound body. Under the replacement colour-distance pipeline, the same image yielded a contour of 25,236 px² (solidity 0.971), within 5% of the analytic wound area, and the drawn contour visually traces the wound boundary.

The single-image test is what motivated the fix. The 30-fold underestimate of contour area was one reason. The other was that the reported contour was located on a piece of tissue that is not the wound. Together, this made it clear that the failure was not a parameter-tuning issue. Any small change to the adaptive-threshold block size or constant would still be responding to local contrast, and the local-contrast maximum in this image is at the suture ring.

Aggregate results

Table 2 reports three-class verdict accuracy over the full 45-image benchmark for both segmentation methods.

Table 2
Three-class verdict accuracy on the 45-image synthetic benchmark.

Segmentation method	Accuracy	Correct predictions
Original (grayscale adaptive threshold)	6.7%	3 / 45
Replacement (HSV colour-distance + Otsu + morphology)	64.4%	29 / 45

The original method's confusion matrix has no consistent diagonal structure. It predicts almost no images as "healthy", only four out of 45 across all ground-truth classes, and it splits both healthy and critical images almost equally between the "concerning" and "critical" predicted classes. The replacement method's confusion matrix has a clear diagonal on the healthy and critical rows. The healthy class is resolved 14 out of 15 times (the single miss is predicted as concerning), and the critical class is resolved 15 out of 15 times. The concerning class is resolved 0 out of 15 times, and all 15 concerning images are predicted as healthy.

Solidity distributions

Figure 6 plots the contour solidity (contour area divided by convex hull area) by ground-truth class for both methods. Under the original method, the healthy class collapses into a tight, anomalously low band, diagnostic of contours drawn around thin suture-like structures rather than filled wound regions. Under the replacement method, the three classes separate cleanly along a single ordered axis, from the most

solid (healthy) to the least (critical), matching the generator's design. This separation is the clearest evidence that the replacement segmentation is measuring the actual wound rather than an artifact of the imaging.

The solidity distributions also explain the residual failure. The default ASYMMETRY_THRESHOLD in the analyzer is 0.82. Under the replacement segmentation, every concerning-class solidity value is above 0.82, so every concerning image is scored 0 on the asymmetry criterion. The border-sharpness threshold and the diameter threshold are similarly missed by most concerning images. The concerning class therefore accumulates a total ABCDE score of 0 or 1 (median 1), which the verdict function maps to PASS. The class does exist as a separate cluster in the underlying solidity and area distributions, and a small rotation of the thresholds would recover it. What the current run shows is that the thresholds baked into the analyzer were fit to the distributions produced by the broken segmentation. They no longer discriminate the class distributions produced by the working segmentation

Reproducing the evaluation

To reproduce the aggregate numbers, generate the 45-image benchmark with python generate_wounds.py, then run python compare_segmentation.py, which evaluates both the original and replacement preprocessing pipelines over the same images in a single run. Per-image raw results in the format file, ground_truth, predicted, total_score, solidity, contour_area are written to track_a_results/raw_original.csv and raw_fixed.csv, and are the source of Table 2. The repository additionally logs the two issues found and corrected during evaluation (an earlier generator bug and the pending fixed-segmentation commit) in track_a_results/reproduction_notes.md

4. Impact

Surgical Vision ROS2 is not a clinical tool. It has been tested only on 45 synthetic images generated from a fixed seed on the author's development machine. No patient data was used, and no IRB review has been sought or is applicable. The software should not be used in any setting that involves a real patient, and it is not validated for use with a physical robotic arm.

Within those limits the software is intended to be useful in three settings.

For robotics researchers working on variable-autonomy teleoperation, the four-node graph is a template for coupling any perception signal to any MoveIt2-driven arm. The /scene_state → /autonomy_mode → /arm_command chain is not specific to wound assessment. It could be reused for gear defect detection, which is what the original DecodeLabs training kit was written for. It could equally be reused for suture-tension monitoring, or for any other task where a scalar or categorical perception signal should modulate arm authority. Because the topics are string-typed, the perception node can be swapped without changing any downstream code.

For ROS2 education, the four-node graph is a compact worked example of topic-based decoupling. It demonstrates a perception-to-actuation loop with three intermediate topics, one operator-in-the-loop node, and a single point of integration with a real motion-planning library. The whole pipeline is under 700 source lines of Python across the four nodes, the analyzer, and the generator, which makes it small enough to be read in a single course session.

Limitations

The dataset is 45 images from a single deterministic seed. Cross-seed variance has not been measured, and sensitivity of the reported 64.4% accuracy to the generator parameters has not been measured. The benchmark contains no images with backgrounds that differ from the drape colour used by the generator, and the corner-sampling assumption in the replacement segmentation would fail on such images.

No baseline segmentation methods other than the original adaptive-threshold pipeline were compared. Otsu thresholding on grayscale, k-means clustering in colour space, and simple background subtraction would all be reasonable baselines and are not evaluated here.

The residual failure on the concerning class is a threshold-calibration problem, not a segmentation problem. Recalibrating the ABCDE numeric thresholds against the distributions produced by the replacement segmentation is expected to recover most of the concerning class. This has not yet been done, because doing it against a 45-image single-seed benchmark would over-fit to the benchmark. It is more honest to defer recalibration until a larger and more varied benchmark is available.

The Evolution (E) criterion requires a previous-frame area to compare against. In the batch evaluation the previous-frame area is fed from the immediately preceding image in filename order, which is arbitrary. The E sub-score is therefore noisy in this benchmark and should not be given weight in the current results. It is retained in the code because it will be meaningful in a real streaming setting where the previous frame is genuinely the previous frame from the same wound.

The MoveIt2 integration has been tested only against the RViz visualisation. It has not been tested against a physical arm, and no timing measurements are reported.

Two packaging issues were found during validation and should be fixed before wider reuse. First, lesion_detector_node.py currently reads its image directory from a hard-coded absolute path on the author's development machine rather than a ROS2 parameter or package-relative default, so the node will not locate images on another machine without a source edit. Second, package.xml declares cv_bridge as a dependency although no node currently imports it, and omits moveit_msgs, which moveit_bridge_node.py does import; a clean workspace without MoveIt2 already installed would fail to resolve this dependency automatically.

Future work

Two changes to the software are planned. First, the ABCDE numeric thresholds should be recalibrated on a larger benchmark, ideally with class-specific thresholds learned from a labelled set rather than hand-picked constants. Second, the autonomy tier assignment should be extended from a fixed lookup to a hysteresis-based rule. That change means a single anomalous frame will not immediately trigger a HOLD state and cause an oscillation when the vision node returns to PASS on the next frame

5. Conclusions

Surgical Vision ROS2 provides a working ROS2 Humble pipeline that couples camera-based wound assessment to variable-autonomy control of a MoveIt2-driven arm. The four-node architecture is small, decoupled at the topic level, and reusable for other perception-conditioned autonomy tasks.

The main empirical finding is that the originally implemented segmentation, which used grayscale adaptive thresholding, reached only 6.7% three-class verdict accuracy on a 45-image synthetic benchmark. Direct inspection of the segmented contours on a single controlled image showed that the method was locking onto suture stitches rather than the wound body. This was traceable to the local-contrast response of adaptive thresholding on synthetic wound interiors that are uniformly coloured and low-contrast. A replacement segmentation that thresholds on global colour distance from a corner-sampled background, followed by Otsu binarisation and morphological cleanup, raised accuracy to 64.4% on the same benchmark.

The improvement is real but incomplete. All 15 concerning-class images are misclassified as healthy under the replacement segmentation. The cause is that the numeric ABCDE thresholds were originally hand-tuned against distributions produced by the broken segmentation, and they no longer discriminate the class distributions produced by the working one. Recalibrating those thresholds against a larger benchmark is the next step. Evaluation on non-synthetic imagery is a further step before any clinical relevance can be considered.

The software is offered as an open reference for researchers, simulation labs, and educators. It is not offered as a clinical tool, and no part of the evaluation should be read as such.

Declarations

Ethical approval

Not applicable. This work used only synthetic images generated by the included generate_wounds.py script. No human or animal subjects, patient data, or clinical data were used, and no IRB or ethics committee review was required.

CRediT authorship contribution statement

Lamia Islam: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Data curation, Writing – original draft, Writing – review & editing, Visualization, Project administration.

Declaration of generative AI in scientific writing

During the preparation of this work, the author used AI tools for language assistance. After using these tools, the author reviewed and edited the content as needed and takes full responsibility for the content of the published article.

Preprint

A preprint of this work has been deposited on Research Square.

Declaration of competing interest

The author declares that she has no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Data availability

All data and code are available in the GitHub repository referenced in the metadata table (row C2).

References

1. Maier-Hein L, Vedula SS, Speidel S, Navab N, Kikinis R, Park A et al (2017) Surgical data science for next-generation interventions. Nat Biomed Eng 1(9):691–696. https://doi.org/10.1038/s41551-017-0132-7
2. Macenski S, Foote T, Gerkey B, Lalancette C, Woodall W (2022) Robot Operating System 2: design, architecture, and uses in the wild. Sci Robot 7(66):eabm6074. https://doi.org/10.1126/scirobotics.abm6074
3. Coleman D, Sucan I, Chitta S, Correll N (2014) Reducing the barrier to entry of complex robotic software: a MoveIt! case study. J Softw Eng Robot 5(1):3–16
4. Bradski G (2000) The OpenCV library. Dr Dobb's. J Softw Tools 25(11):120–125
5. Friedman RJ, Rigel DS, Kopf AW (1985) Early detection of malignant melanoma: the role of physician examination and self-examination of the skin. CA Cancer J Clin 35(3):130–151. https://doi.org/10.3322/canjclin.35.3.130

Figures

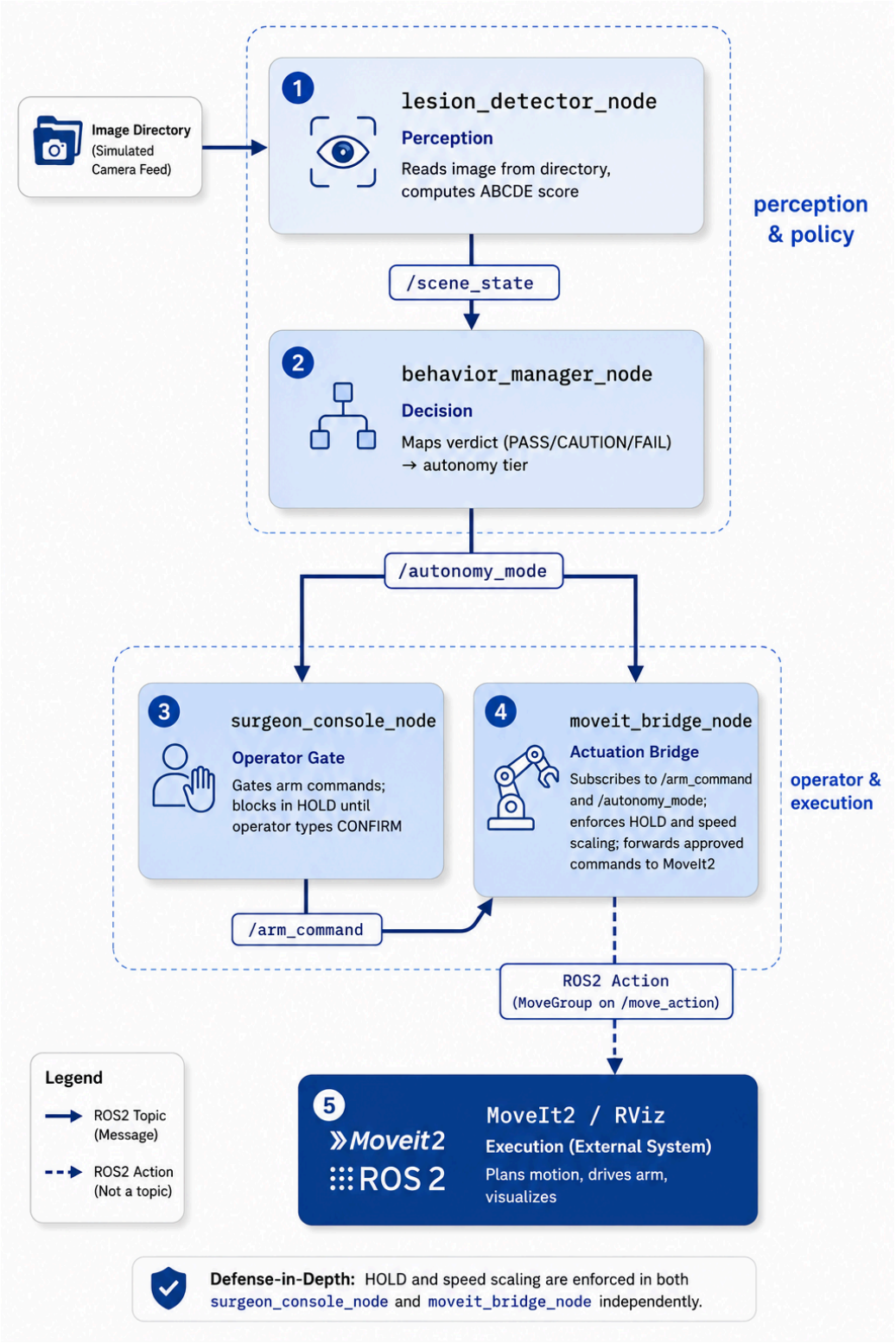


Figure 1

ROS2 node graph of Surgical Vision ROS2. Four nodes — lesion_detector_node, behavior_manager_node, surgeon_console_node, moveit_bridge_node — are connected by the topics /scene_state, /autonomy_mode, and /arm_command; /autonomy_mode fans out to both surgeon_console_node and moveit_bridge_node, which enforce HOLD and speed-tier gating

independently. The chain terminates in a ROS2 action call (MoveGroup on /move_action) to MoveIt2, visualized in RViz.

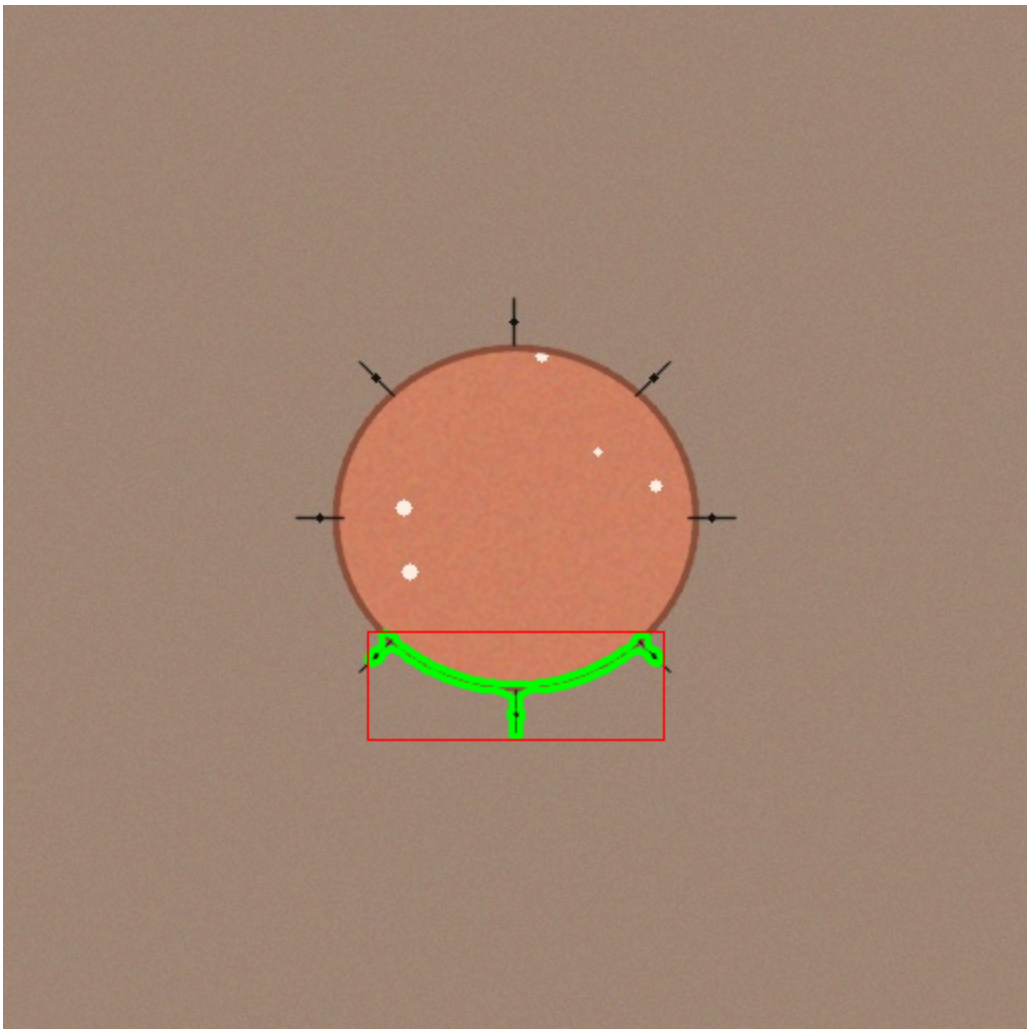


Figure 2

Original adaptive-threshold segmentation on healthy_01.jpg. The largest external contour has an area of 804 px² and is located on a suture cluster in the lower half of the image. The analytic wound area (ellipse with rx = 90, ry = 85) is approximately 24,032 px².

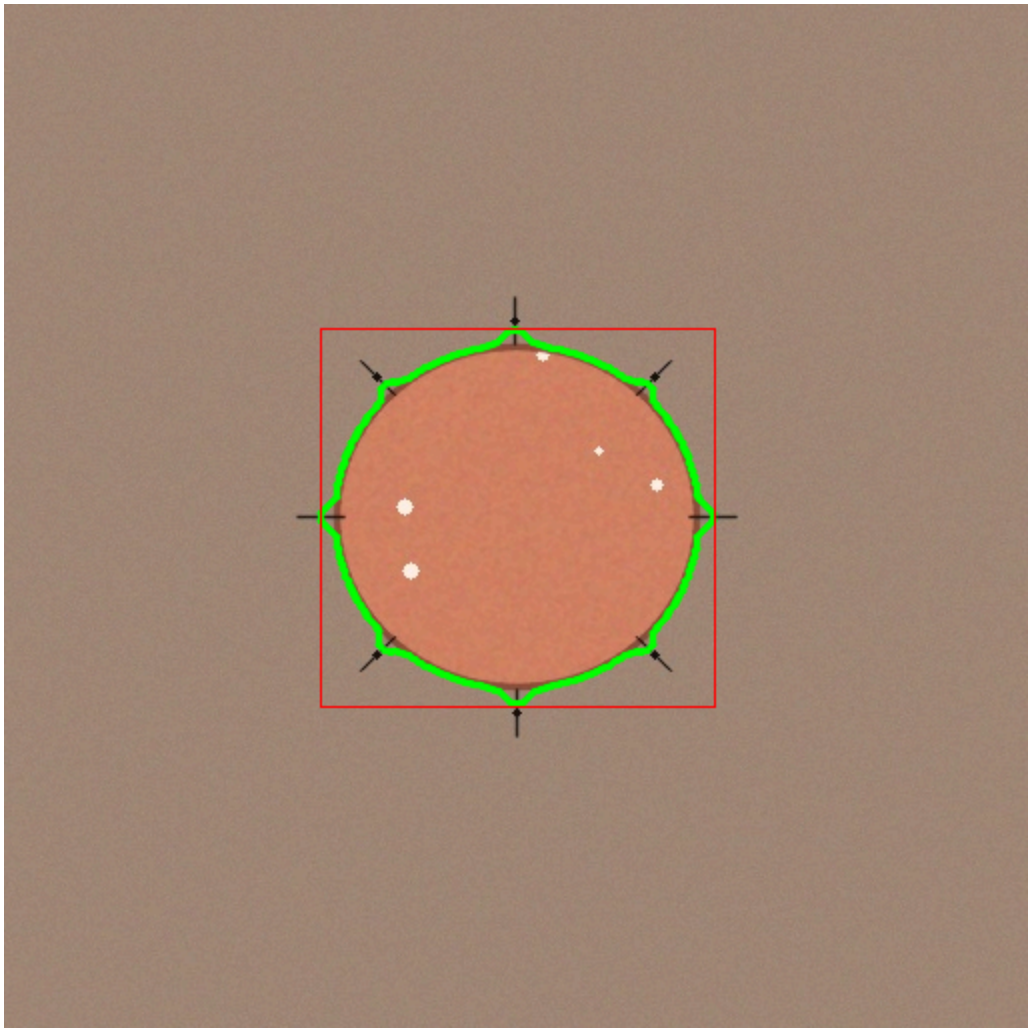


Figure 3

Replacement colour-distance segmentation on the same image. The largest external contour has an area of 25,225 px² and traces the wound boundary.

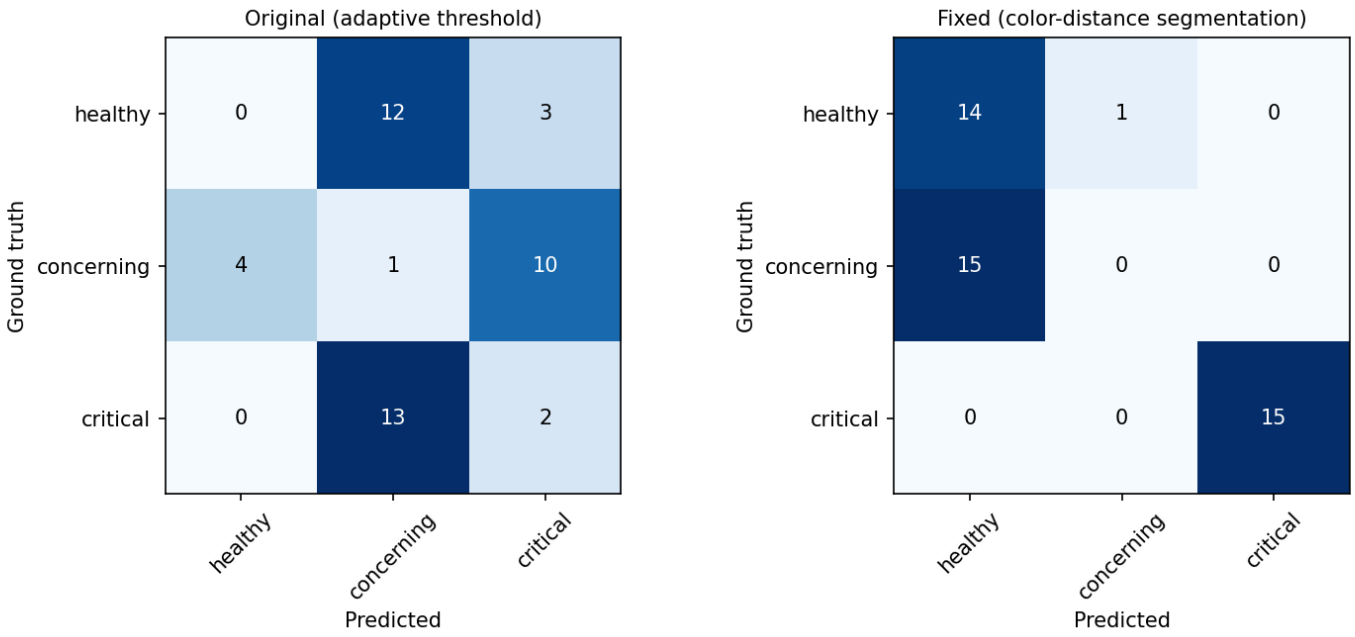


Figure 4

Confusion matrices on the 45-image benchmark, three-class verdict (healthy / concerning / critical). Left: original adaptive-threshold pipeline. Right: replacement colour-distance pipeline. The original has no consistent diagonal structure. The replacement resolves the healthy and critical classes but predicts all concerning images as healthy.

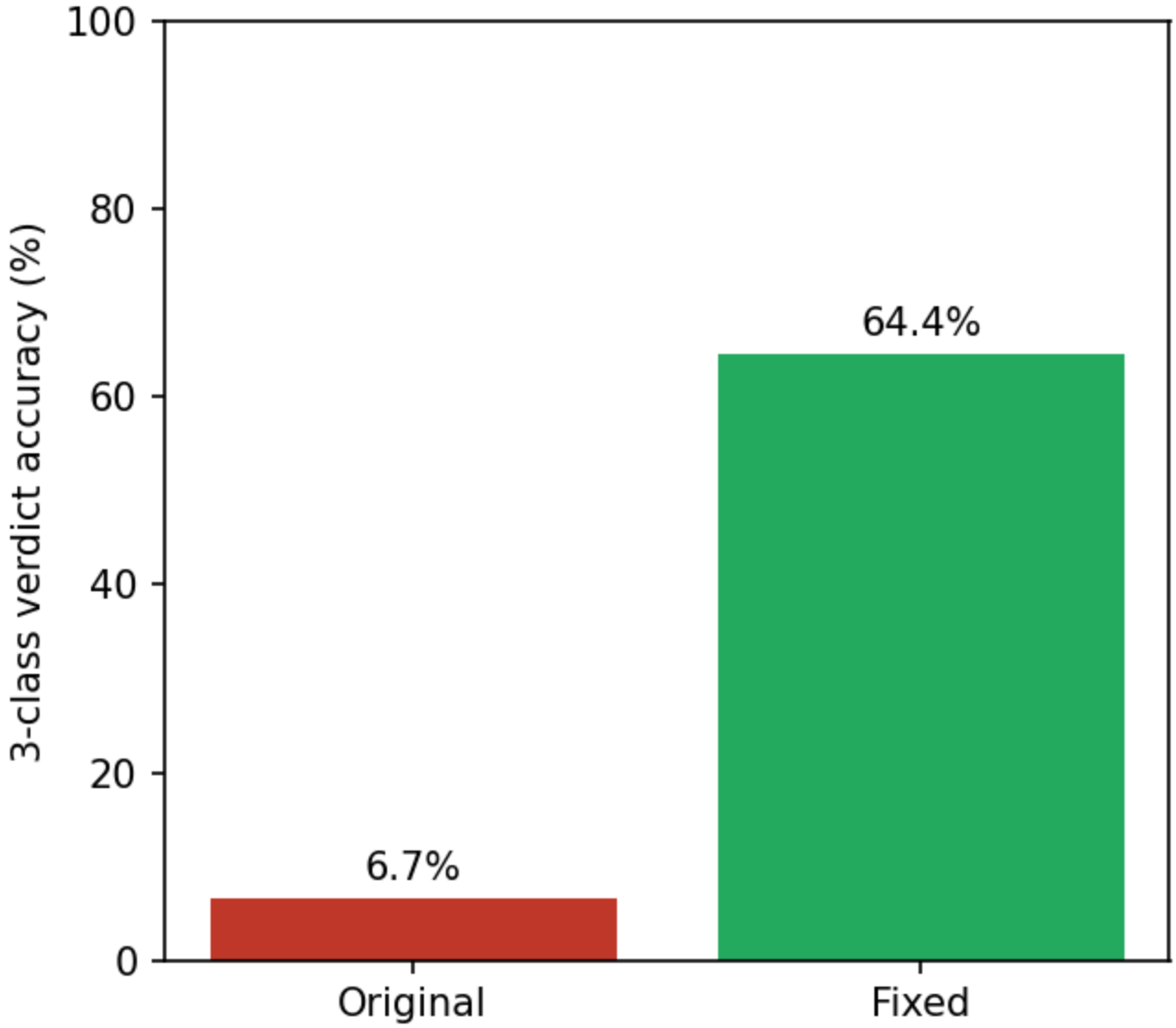


Figure 5

Three-class verdict accuracy on the 45-image benchmark: 6.7% (3/45) for the original pipeline and 64.4% (29/45) for the replacement pipeline.

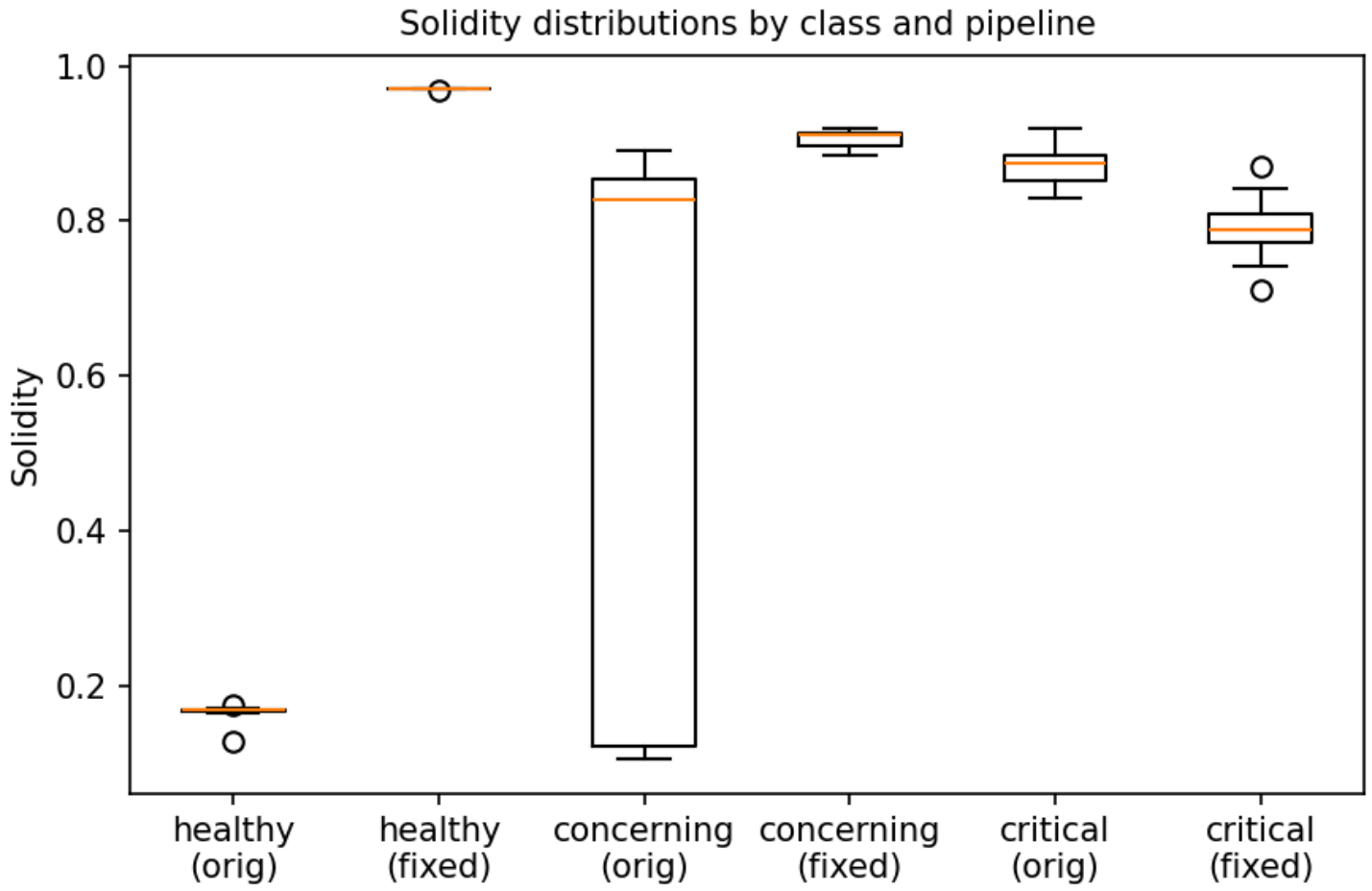


Figure 6

Contour solidity distributions by ground-truth class. Left: original pipeline. The healthy class shows anomalously low solidity clustered around 0.16, consistent with contours drawn around suture rings rather than filled wound regions. Right: replacement pipeline. Healthy has median 0.971 (range 0.968–0.971), concerning has median 0.910 (range 0.884–0.920), critical has median 0.788 (range 0.711–0.869).