Pseudo-Morphology-Guided Learning for Cotton Leaf Curl Disease Classification

Renu[1], Neha Gaba[2]
[1]Research Scholar, Department of Computer Science and Applications, Baba Mastnath University, Rohtak, Haryana, India; [2]Associate Professor, Baba Mastnath University, Rohtak, Haryana, India
Email id: renubaghel09876@gmail.com[1], nehabaga05@gmail.com[2]

Article Received: 20 June 2025 *Article Accepted: 25 July 2026* *Article Published: 10 Aug 2026*

Citation
Renu, Neha Gaba, "Pseudo-Morphology-Guided Learning for Cotton Leaf Curl Disease Classification", Journal of Next Generation Technology (ISSN: 2583-021X), 6(4), pp. 17-42. Aug 2026.

Abstract

Cotton Leaf Curl Disease (CLCuD) alters venation, lamina shape and leaf margins, yet classification accuracy alone cannot establish whether a model uses these disease-relevant signs. This paper proposes MorphoReli-CLCuD, a pseudo-morphology-guided framework combining bounded adaptive disease-preserving enhancement, a vein-aware attention module, and quality-gated fusion of local texture, pseudo-venation, leaf-boundary and global-geometry features. The retained cohort contains 472 healthy-labelled and 2,990 CLCuD-labelled images. Reconstructed capture grouping preceded a locked 70/15/15 allocation, and eight controlled configurations were trained for 12 epochs at three prespecified seeds. C2 achieved the highest mean macro-F1, 0.916 (SD 0.006). In a corrected crossed bootstrap that applies each resampled test-group set to all sampled seeds and paired models, its mean differences were 0.0217 versus C0 (95% CI -0.0027 to 0.0497) and 0.0209 versus P1 (-0.0048 to 0.0483); the Holm-adjusted one-sided p-value was 0.090 for both comparisons, so the prespecified confirmatory criterion was not met. A1 achieved the highest mean AUROC, 0.974, and the lowest mean Brier score, 0.032. Quantitative XAI and cue interventions provide behavioural, rather than anatomical, evidence. The study supplies an auditable architecture and a positive repeated-seed trend that requires confirmation with adjudicated grouping, same-label permutation controls and independent field data.

Keywords: Cotton leaf curl disease, pseudo-morphology supervision, pseudo-venation, reliability-gated fusion, repeated-seed evaluation, explainable artificial intelligence

I. Introduction

Cotton Leaf Curl Disease (CLCuD) is a major viral disease complex of cotton and can visibly alter veins, the lamina, leaf margins and overall leaf geometry [1]–[3]. These signs occur at different spatial scales. Fine vein-like changes, local surface deformation and whole-leaf curling therefore cannot be represented adequately by a single hand-defined cue.

Deep convolutional and Transformer models have already been used for CLCuD and related cotton-disease recognition [4]–[13]. Attention, feature fusion, boundary processing, shape-aware convolution and post-hoc explanations are also established. The unresolved question addressed here is narrower: can automatically extracted, correctly paired morphology proxies improve a CLCuD classifier beyond the effect of additional architecture and auxiliary losses alone?

This study proposes MorphoReli-CLCuD, a pseudo-morphology-guided framework built around a shared ResNet34 feature pyramid. It combines bounded adaptive disease-preserving enhancement

(ADPE-Lite), a vein-aware attention module (VAAM), and separate local-texture, pseudo-venation, leaf-boundary and global-geometry branches. A quality-aware residual gate combines the branches while retaining ungated evidence for robustness and auditability.

The study addresses three research questions:

RQ1: Does correctly paired automatic morphology supervision improve macro-F1 relative to a same-capacity diagnosis-only control and a mismatched-target control?

RQ2: What system-level behaviour is observed across adaptive preprocessing, pseudo-venation, simple fusion and quality-gated fusion under an equal training budget?

RQ3: Do the integrated morphology-aware features show greater foreground and structural-region sensitivity than a standard ResNet34 classifier under quantitative XAI and cue interventions?

The contributions are: (1) a CLCuD-specific decomposition of local texture, pseudo-venation, leaf boundary and global geometry; (2) bounded ADPE-Lite preprocessing regularized by image identity and edge preservation; (3) a VAAM implementation with an inspectable pseudo-vein output and visibility estimate; (4) quality-gated residual fusion with explicit branch outputs; and (5) a reconstructed-group, repeated-seed evaluation with same-capacity and mismatched-target controls. The novelty claim concerns this controlled coupling and evaluation, not the invention of CNNs, feature pyramids, attention, gating or attribution methods.

A. Abbreviations and operational terms

ADPE-Lite denotes adaptive disease-preserving enhancement; VAAM denotes the vein-aware attention module; MorphoReli denotes morphology-reliability fusion; CNN denotes convolutional neural network; FPN denotes feature pyramid network; RGB denotes red-green-blue input; and XAI denotes explainable artificial intelligence. ViT denotes Vision Transformer, ConViT denotes a convolutional Vision Transformer, SE denotes squeeze-and-excitation, and RISE denotes Randomized Input Sampling for Explanation. U²-Net is the nested U-structured segmentation network used for leaf isolation; pHash and dHash denote perceptual hash and difference hash; AdamW is Adam optimization with decoupled weight decay; and Grad-CAM++ denotes Gradient-weighted Class Activation Mapping++. AUROC is area under the receiver operating characteristic curve, AP is average precision, BA is balanced accuracy, ECE is expected calibration error, CI is confidence interval and SD is standard deviation. BCE denotes binary cross-entropy. A pseudo-label or morphology prior is an automatically estimated training target rather than an expert anatomical annotation. Configuration identifiers B0, A1, V1, C0, C1, C2, P1 and F1 are model labels; the F1 configuration is distinct from the F1 classification metric.

II. Related Work

A. CLCuD morphology and image-based diagnosis

Biological and phenotyping studies describe CLCuD as a multi-sign progression involving secondary and main-vein thickening, curling, enation and, at the plant scale, stunting [1]–[3]. Because the present collection contains detached or isolated leaf images, the model cannot infer whole-plant stunting, yield loss or agronomic resistance. Its output is consequently restricted to image-level diagnosis.

Magsi et al. combined colour, texture and morphological features with a CNN for four disease stages and reported that early-stage recognition was difficult [4]. Nazeer et al. studied susceptibility-scale classification using a self-collected expert-labelled dataset [5], while Ahmad et al. used RGB and

multispectral information for severity-level diagnosis [6]. Lightweight CNNs and transfer-learning systems have also classified leaf curl among multiple cotton conditions [7]. Most directly, Biju et al. compared CNN, ViT and ConViT variants for healthy-versus-CLCuD classification [9]. These works establish that neither CNN/Transformer use nor severity prediction alone is new for CLCuD. They do not, however, evaluate a joint proxy-guided separation of pseudo-venation, leaf-boundary and global-geometry evidence for CLCuD.

B. Cotton attention, fusion, shape and XAI

The broader cotton literature further constrains the novelty claim. MultiCotNet combines mobile inverted bottleneck convolution (MBConv) and SE operations, side-layer fusion and spatial attention [10]. Ullah et al. evaluated dual CNN–Transformer fusion and quantitative XAI [11]. Chen et al. combined high/low-frequency enhancement, boundary cues and attention-guided cross-scale fusion [12], and EnsembleNet used shape convolution with multi-scale hierarchical features [13]. Our claim is therefore not the invention of attention, fusion, boundary processing, enhancement, shape-aware convolution or XAI. Two additional 2026 plant-disease systems further narrow the claim. LeafFusionNet combines Gabor texture priors, dual attention, local/global feature fusion and Grad-CAM across PlantVillage classes [14]. LeafMorph-X combines restoration and illumination normalization with lesion/vein segmentation and cross-attention between lesion appearance and vein morphology for tomato disease recognition [15]. Thus, morphology-aware naming, vein features, disease-oriented enhancement and local/global attention are not unique to the present work. To our knowledge, among the literature reviewed through 10 September 2026, this is an early CLCuD-specific framework that jointly operationalizes pseudo-venation, leaf-boundary and global-geometry evidence with explicitly inspectable gated fusion and matched shortcut evaluation.

Recent same-domain studies sharpen this positioning. CottonLeafNet combines multiple attentive CNNs, evaluates two public cotton datasets and reports cross-dataset experiments [16]. Saha et al. trained an Inception-style seven-class model on the augmented SAR-CLD-2024 categories [17], [18]. Swapno et al. evaluated a lightweight CNN–ViT framework on SAR-CLD-2024 and other cotton or fabric datasets with post-hoc explanations [19], while Naeem et al. compared transfer-learning models on field and public cotton datasets [20]. These works establish strong prior art for multiclass cotton recognition, attention, transformers, cross-dataset evaluation and deployed interfaces. Table I states the narrower distinction tested here: explicit, separately supervised CLCuD pseudo-venation and boundary outputs, matched supervision controls, locked source grouping and repeated-seed matched-control evaluation.

Table I. Positioning against the closest direct and adjacent prior work

Study	Cotton task and evidence	Structural representation	Generalization or control design	Relation to this study
Magsi et al. [4]	Direct CLCuD severity, 1,600 images	Colour, texture and morphological features with CNN	Stage classification	Establishes direct CLCuD severity work
Nazeer et al. [5]	Direct CLCuD susceptibility classes	CNN representation	Self-collected and public cohorts	Establishes ordered susceptibility classification
Biju et al. [9]	Healthy versus CLCuD	CNN, ViT and ConViT comparison	Model comparison	Establishes direct CNN/Transformer baselines
MultiCotNet [10]	Multiclass cotton disease	Multi-spatial attention and feature fusion	Cotton and PlantVillage evaluation	Establishes cotton attention and fusion
EHDM-Net [12]	Similar cotton disease detection	Boundary/frequency enhancement and cross-scale fusion	Self-constructed and public data	Establishes boundary-aware cotton detection
EnsembleNet [13]	Multiclass cotton disease	Shape convolution and hierarchical features	Eight-class evaluation	Establishes explicit cotton shape modelling
CottonLeafNet [16]	Multiclass cotton disease	Attentive CNN ensemble	Cross-dataset and imbalance analysis	Establishes attentive cotton ensembles and transfer testing
Saha et al. [18]; Swapno et al. [19]	Seven-class SAR-CLD-2024 recognition	Inception CNN and lightweight CNN–ViT	Same repository categories	Establish multiclass modelling on the external repository
MorphoReli-CLCuD	Binary healthy-versus-CLCuD classification	Explicit pseudo-vein, boundary and geometry outputs with gated fusion	Grouped split and matched supervision controls	Tests whether correctly paired pseudo-morphology supervision adds internal value

Note. Reported results are not compared as a leaderboard because datasets, splits, tasks and augmentation protocols differ.

C. Validity and explanation evaluation

Image-level random splitting is unsafe when a collection contains burst frames, near duplicates or offline augmentations. Related samples on both sides of a split can convert the test into source-instance recognition rather than generalization [21]. Plant-disease models also commonly lose accuracy between controlled and field domains [22]. These concerns motivate source grouping, locked manifests, explicit background interventions and testing on an independently sourced external collection.

Gradient-weighted Class Activation Mapping++ (Grad-CAM++), Integrated Gradients and Randomized Input Sampling for Explanation (RISE) provide complementary gradient- and perturbation-based explanation families [23]–[25]. In this work, heatmaps are not treated as proof of biological reasoning. They are evaluated through overlap with automatic morphology pseudo-regions, attribution outside the leaf and confidence changes after masking individual cues. The implemented evaluation combines attribution overlap with foreground, background and structural-region interventions.

III. Materials and Dataset

A. Data origin and diagnostic labels

The recovered source contains cotton-leaf photographs assigned to healthy and CLCuD classes. The researcher reports that every internal image was manually photographed from cotton plants in the field before the cleaning operations described below; no web-sourced image was intentionally added to this internal collection. The supplied high-resolution acquisition archive contains 1,044 uniquely named field photographs (2.15 GB). The researcher reports that the full field collection comprised approximately 1,200 healthy and 4,500 CLCuD photographs (approximately 5,700 images and 10 GB). The supplied archive is a locally verified subset of that acquisition workspace and is preserved separately from the cleaned canonical modelling directory. Project records identify all 2,990 disease-class images as CLCuD rather than as a mixture of unspecified cotton diseases. However, the identity and qualifications of the original label assessor, the visual diagnostic criteria and any laboratory confirmation are not documented; the labels are therefore treated as project-provided image labels rather than laboratory-verified diagnoses. Healthy and CLCuD photographs were collected by the same photographer under the same general location and time conditions. The collection includes a mixture of upper- and lower-leaf views, but per-image view metadata have not yet been recovered.

The use of one photographer and broadly shared acquisition conditions reduces some class-specific device variation, but it does not establish that every image represents a different physical leaf or plant. Filename timestamps, capture bursts and perceptual similarity must therefore be incorporated into group reconstruction.

B. Canonical preprocessing and quality control

Processing records were used to reconstruct the deterministic image-preparation sequence. Source images remain preserved, and all processing outputs are stored separately. The modelling workflow uses:

1. Resolution standardization. The Python Imaging Library (PIL) converts images to red–green–blue (RGB) format and resizes them to 180 x 180 pixels with Lanczos interpolation.

2. Leaf isolation. rembg, Open Neural Network Exchange (ONNX) Runtime and U²-Net estimate an alpha foreground, which is composited on uniform black.

3. Automatic quality routing. Foreground count, coverage, border proximity, aspect ratio and blur heuristics flag multiple leaves, excess padding, near-border foreground, unusual shape or low-content images for review. These are image quality flags and do not change disease labels.

4. Foreground centring. The non-black leaf bounding box is cropped, placed at the centre of a square canvas and resized to 180 x 180 pixels.

5. Canonical freeze. The resulting 472 healthy-labelled and 2,990 CLCuD-labelled images are copied to an immutable canonical source directory.

6. Reconstructed-group allocation. Exact/perceptual hashes, appearance and capture timing reconstruct related-image groups before the locked 70/15/15 split.

7. Training-time transformations. Geometric and photometric transformations are generated in memory while sampling training images. No generated image is added to the canonical, validation, test or external sets.

Images were processed in batches so that foreground-isolation outputs and automatically routed cases could be inspected without altering the preserved source files. Automated rules only created

review queues. Manual verification then considered leaf presence, single-leaf suitability, foreground extraction, crop completeness, background content and class consistency before the canonical set was frozen. Diagnostic classes were not changed by an image-quality heuristic alone.

The retained preparation records do not preserve a class-stratified final exclusion ledger for unusual-shape, border-crop and other review queues. The study therefore cannot quantify whether diagnostically atypical but technically usable leaves were excluded disproportionately; all performance claims are conditional on the frozen canonical cohort.

Fig. 1 illustrates the canonical preprocessing stages and the automatic leaf, boundary and pseudo-vein proxies used in this study.

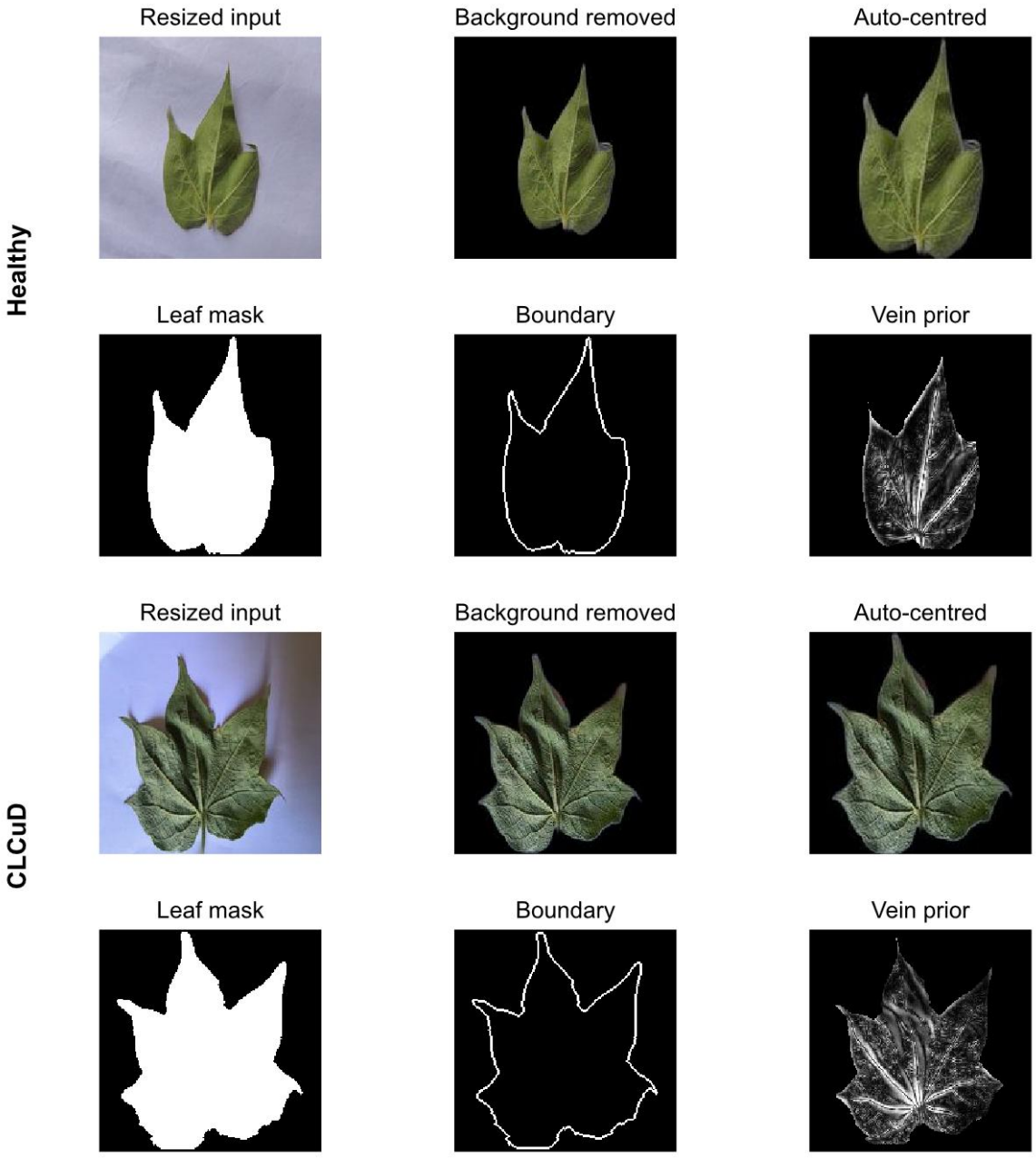

Fig. 1. Representative healthy and CLCuD images across canonical preprocessing. The first three columns show the resized input, foreground isolation and centring. The remaining columns show the

automatic leaf mask, boundary and soft pseudo-vein prior. The vein response is a morphology proxy, not expert-traced anatomical ground truth

Fig. 2 summarizes the internal development and validation sequence from field acquisition to matched-control evaluation.

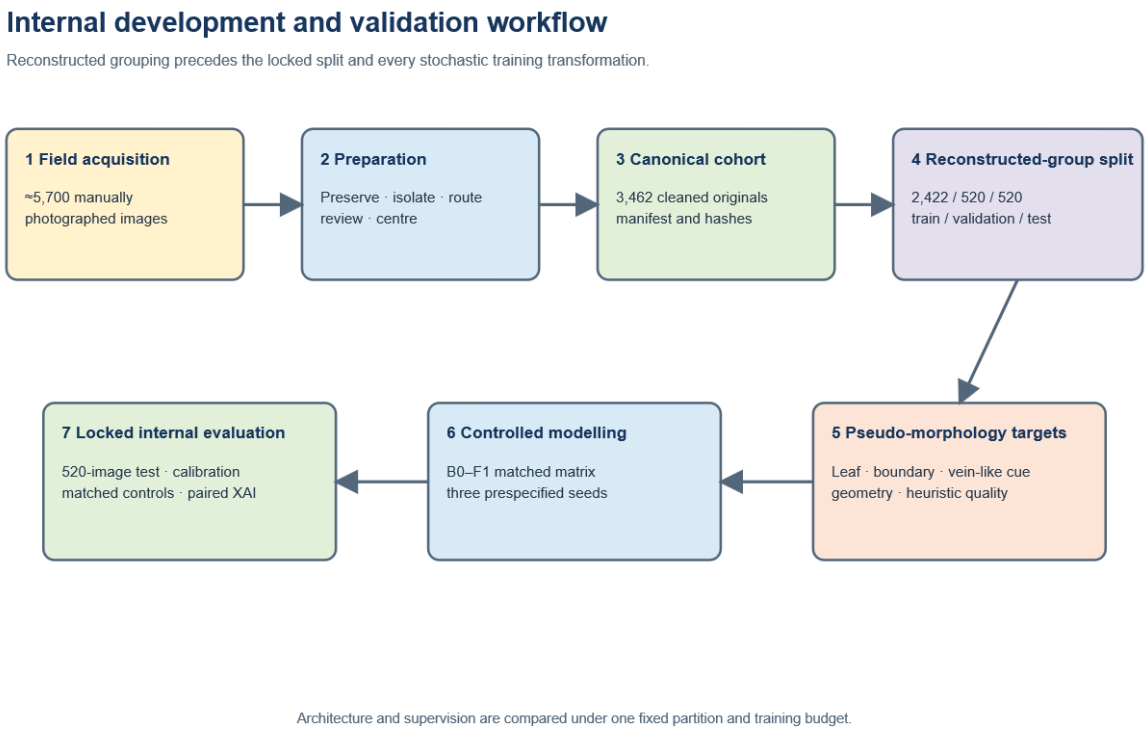

Fig. 2. Internal development and validation workflow. Field acquisition and source preservation lead to the canonical cohort, reconstructed-group locked split, automatic morphology targets, controlled B0–F1 modelling and internal evaluation. Reconstructed grouping precedes the locked split and every stochastic training transformation

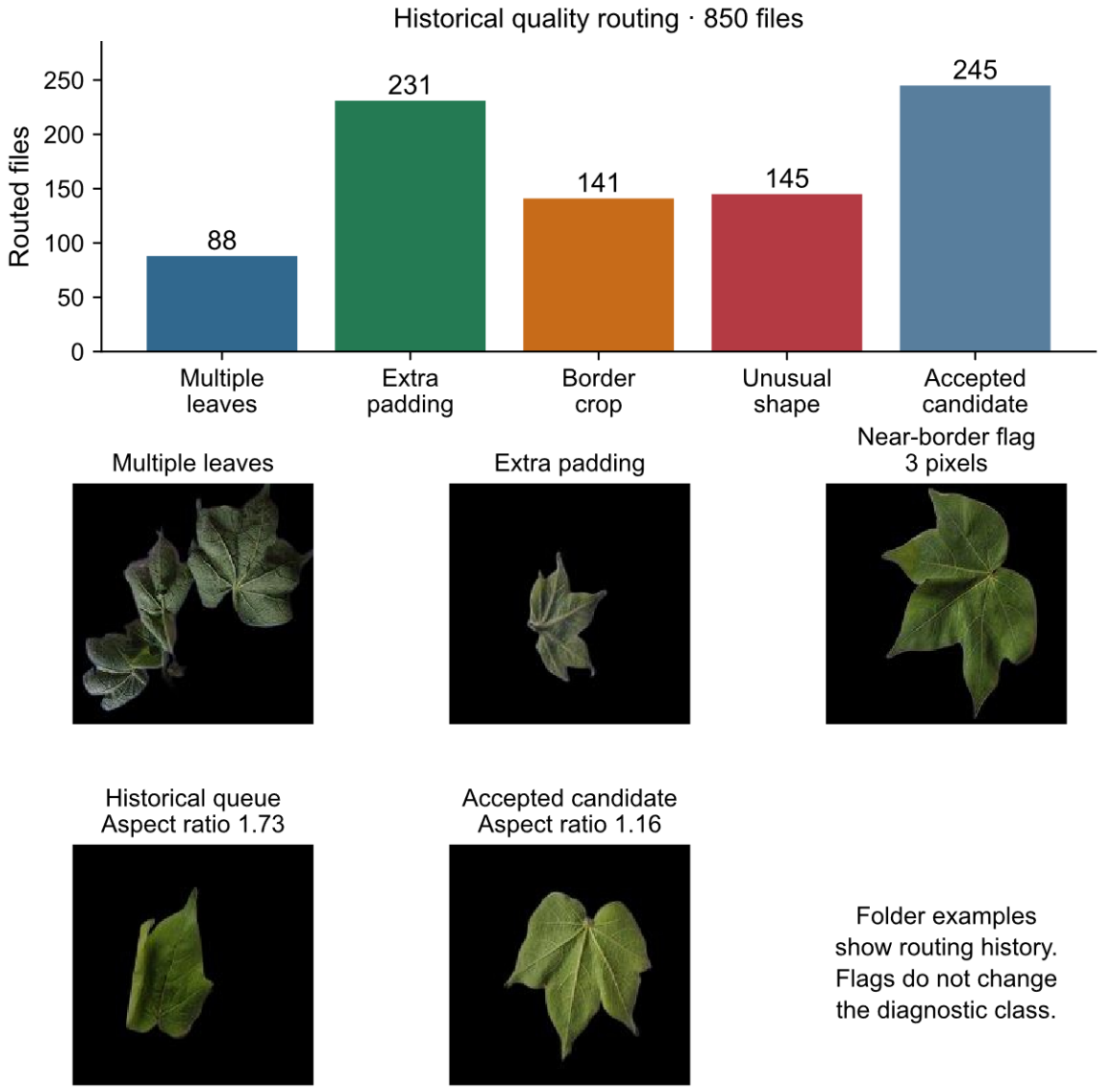

Fig. 3. Recovered healthy-class quality-routing evidence. The upper panel reports the 850 terminal routing outcomes recorded during the quality-control process: 88 multiple-leaf flags, 231 extra-padding flags, 141 border-crop flags, 145 unusual-shape flags and 245 accepted candidates. The lower panel shows one representative image from each routing category. The displayed near-border example has a measured minimum border distance of 3 pixels. The displayed accepted candidate has aspect ratio 1.162 and minimum border distance 17 pixels. The unusual-shape example has measured aspect ratio 1.727 and is therefore shown only as historical queue evidence; it is not claimed to satisfy the reported extreme-aspect threshold. These are routing counts rather than counts removed from the canonical cohort, and no routing rule changes the diagnostic class

The final canonical directory contains 3,462 decoded images and supplies the only internal images used by the evaluated models.

C. Canonical dataset audit

Table II reconciles the researcher-reported field totals with the verified canonical inventory.

Table II. Field-collection and canonical cleaned dataset inventory

Property	Healthy	CLCuD	Total
Initial field collection (reported, approximate)	~1,200	~4,500	~5,700
Initial field-collection storage (reported, approximate)	—	—	~10 GB
Locally supplied high-resolution acquisition subset	—	—	1,044 images (2.15 GB)
Canonical cleaned source images	472	2,990	3,462
Successfully decoded	472	2,990	3,462
Resolution	180 x 180	180 x 180	180 x 180
Exact duplicate files within canonical set	0	0	0

Note. The initial acquisition values are researcher-reported approximations; the supplied archive and canonical counts were verified against the study records.

Secure Hash Algorithm 256 (SHA-256) inventory found 3,462 unique byte hashes. This is an exact-duplicate result, not proof that the dataset lacks visually near-identical captures or re-encoded copies. Perceptual-hash, appearance-neighbour and timestamp grouping was therefore performed before split allocation.

Fig. 4 shows representative near-neighbour pairs retained within common reconstructed source groups.

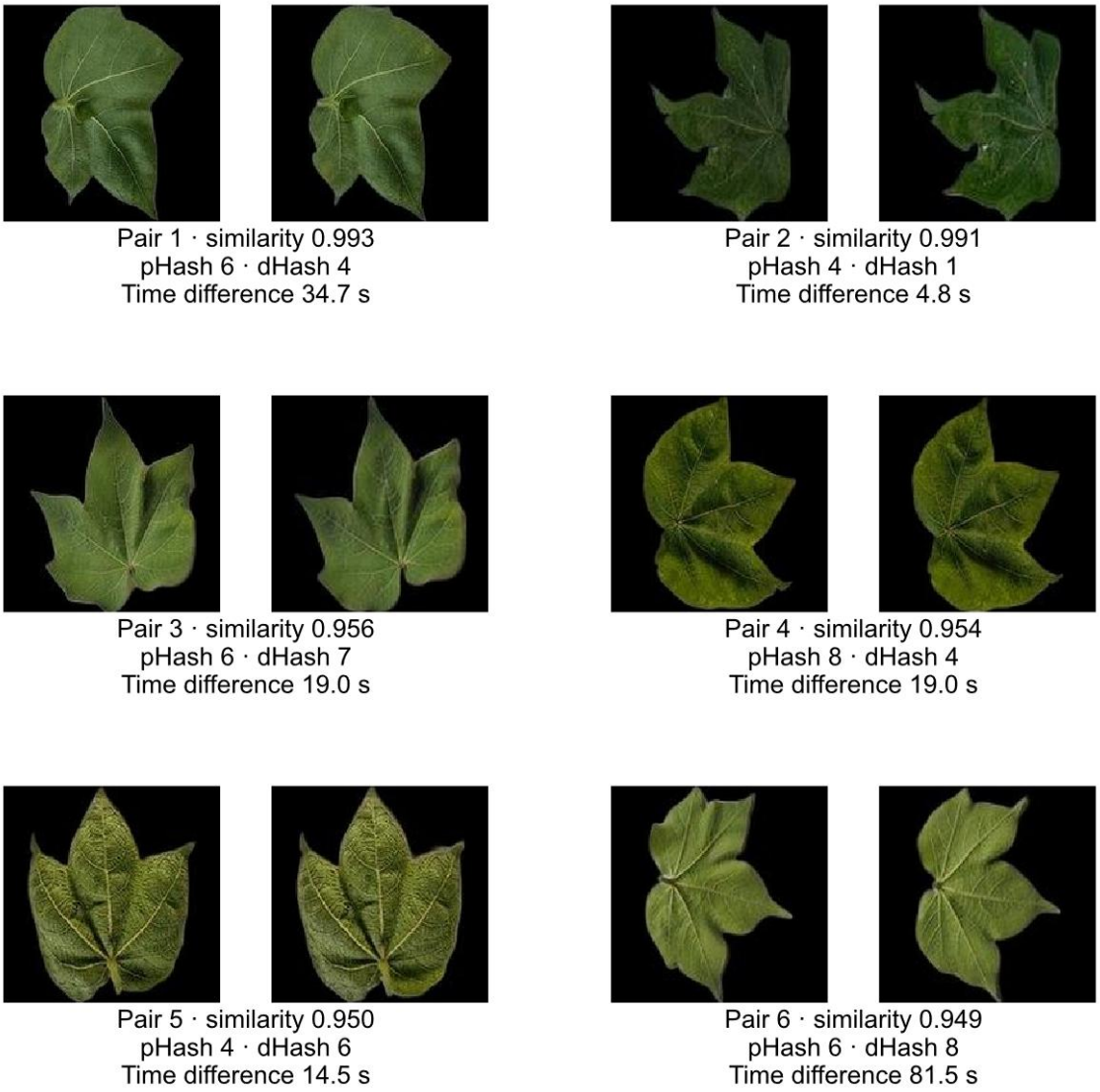

Fig. 4. Six representative pairs selected from the longer near-neighbour contact sheet supplied during the data audit. Similarity, perceptual-hash distances and capture-time separation were used to keep likely repeated views in a common source group. Only six of the 19 grouped pairs are displayed to avoid an excessively long figure

The class ratio is approximately 6.33:1 in favour of CLCuD. Accuracy is not an appropriate sole endpoint; macro-F1, balanced accuracy and per-class sensitivity and specificity are reported. Class weighting was restricted to training.

Fig. 5 summarizes the duplicate audit, reconstructed source groups and locked allocation.

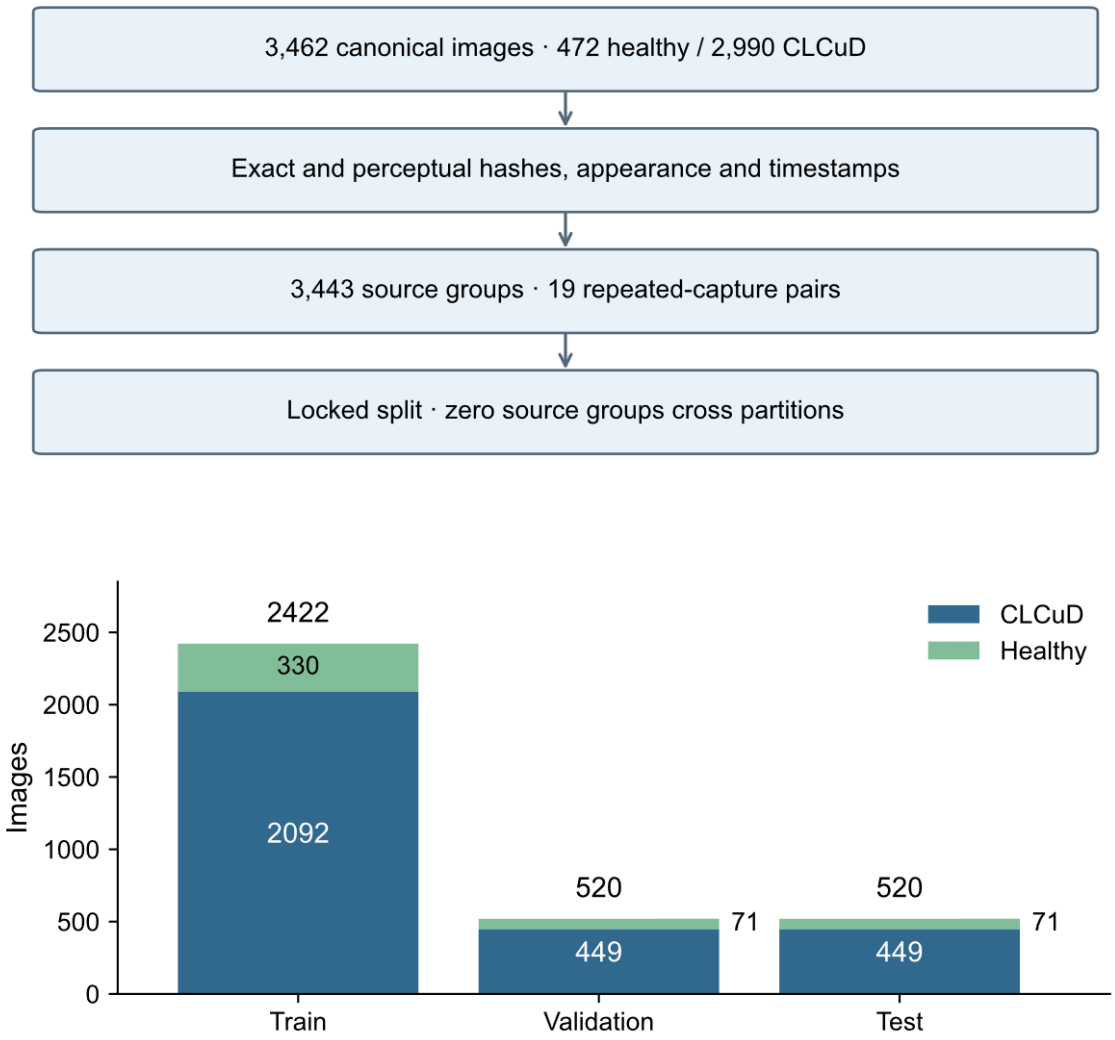

Fig. 5. Dataset audit and locked split. Exact and perceptual hashes, foreground-normalized appearance and capture timestamps produced 3,443 reconstructed source groups. Nineteen same-label likely repeated-capture pairs were kept together. No reconstructed group crosses partitions; the retrospective label-blind audit later identified one unresolved cross-label, cross-partition near-neighbour outside those groups

D. Resolution and scope

All images in the locked experimental set are already 180 x 180 pixels. A separate 2.15-gigabyte acquisition archive contains 1,044 high-resolution files, with sample dimensions of 3,264 x 2,448 and 4,160 x 3,120. Only 27 filenames overlap the canonical healthy set, and a complete mapping to the 3,462-image modelling set is unavailable; those files are therefore documented but not substituted into the evaluated experiment. Inputs may be resampled to 224 x 224 for compatibility with pretrained encoders, but such resampling does not recreate fine venation that was lost during the earlier resize. The vein branch therefore uses an automatic visibility-confidence proxy, and all claims concern ridge-based pseudo-venation rather than traced anatomy. Claims about fine secondary-vein recovery remain limited.

The primary outcome is binary healthy-versus-CLCuD diagnosis. Disease severity is not evaluated because the source provides no expert-reviewed ordered severity labels. Likewise, diffuse structural symptoms will not be described as lesion segmentation without lesion annotations.

IV. Proposed Method

Let a red–green–blue image be $x \in [0,1]^{H\times W\times 3}$, where H and W are image height and width. The binary diagnostic label is $y\in\{0,1\}$, with 1 denoting CLCuD. Automatic structural pseudo-targets are the leaf mask m_l, boundary map m_b and vein map m_v. Their confidence indicators down-weight low-quality proposals in the corresponding auxiliary losses.

Fig. 6 presents the proposed architecture and its inspectable branch outputs.

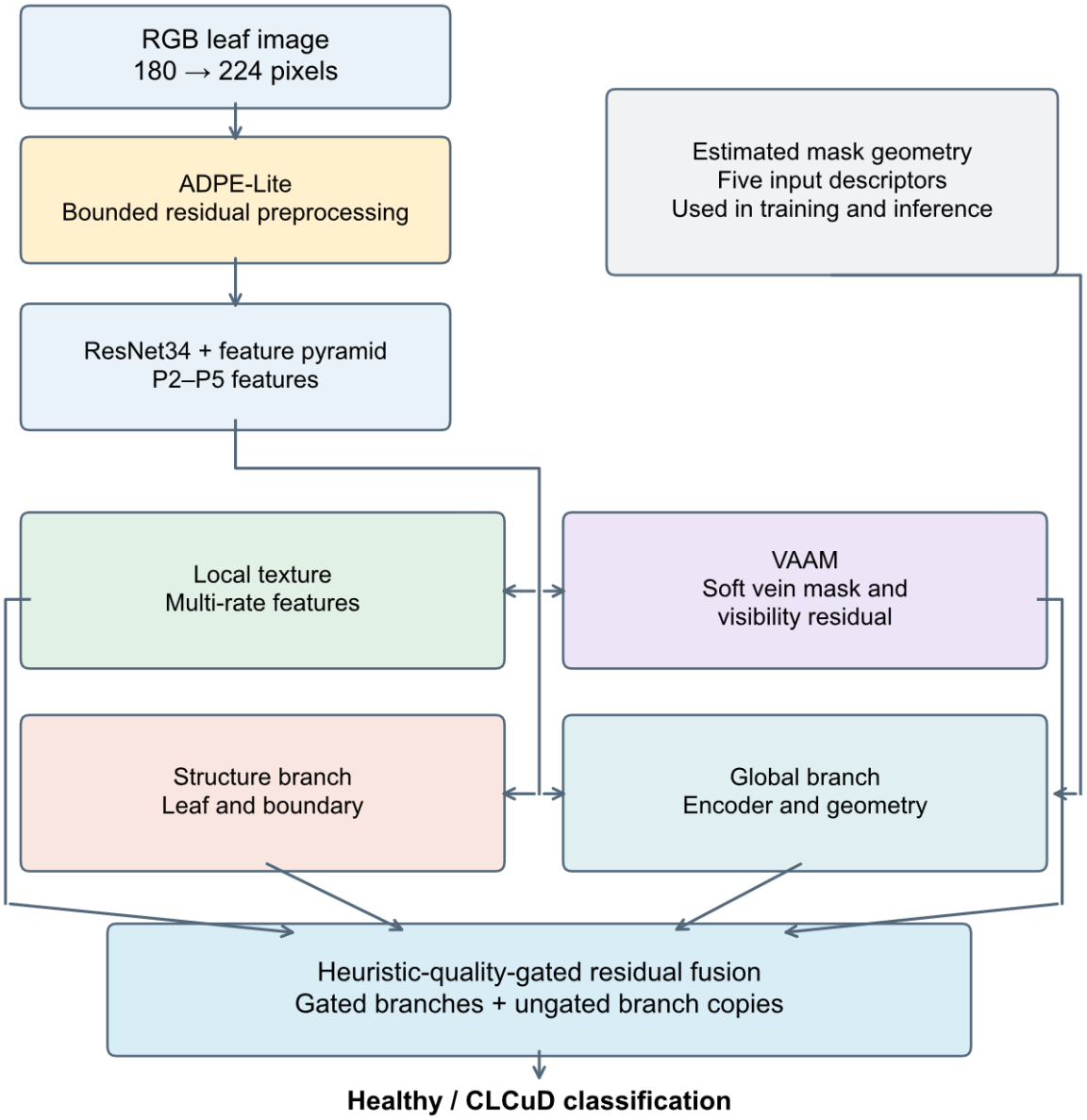

Fig. 6. Proposed MorphoReli-CLCuD architecture. ADPE-Lite applies bounded residual preprocessing before a shared ResNet34 feature pyramid. Local texture, VAAM pseudo-venation, leaf-boundary and global-geometry branches retain separate representations until gated residual fusion. The evaluated output is binary healthy/CLCuD classification with quantitative XAI; the architecture exposes all branch maps and gate values for audit

A. ADPE-Lite: bounded residual preprocessing

ADPE-Lite is a shallow residual module with two learned spatial gates. Their architectural intent is to represent illumination and structure; their semantic roles are evaluated through bounded outputs and cue-level analysis. The module predicts a bounded residual *r(x)* and combined spatial gate *a(x)*, as defined in Equation (1):

$$\tilde{x} = clip(x + a(x) \odot r(x), 0, 1) \quad (1)$$

In the absence of paired clean/enhanced targets, ADPE-Lite is not trained to imitate a subjectively enhanced reference. The implementation uses RGB identity and image-edge preservation, combined in Equation (2):

$$L_ADPE = 0.15L_identity + 0.10L_edge. \quad (2)$$

Here, *clip* bounds pixel values, $\odot$ denotes element-wise multiplication, and the residual is limited to magnitude 0.15 per channel. $L_{identity}$ penalizes unnecessary RGB change and L_{edge} penalizes edge disagreement between input and output.

B. Shared encoder and feature pyramid

The enhanced image is processed by a pretrained ResNet34 encoder with a feature pyramid network [26]. Multi-resolution maps $P_2,\ldots,P_5$, ordered from higher spatial resolution to deeper semantic features, share computation across the structural and diagnostic branches. ResNet34 is chosen as a practical capacity/compute compromise for the small source collection; the backbone is not itself a novelty claim.

C. VAAM: vein-aware attention module

VAAM fuses the high-resolution P_2 feature map with an upsampled P_3 map and predicts a vein logit map ℓ_v, its probability map $m_v=\sigma(\ell_v)$, vein embedding z_v and vein-visibility estimate q_v. Automatic ridge/vesselness pseudo-maps supervise the vein head only when their validity indicator is active; valid targets are weighted by confidence clipped to [0,1]. All images remain active for diagnostic classification. No manual or manually verified vein mask was used, so expert-labelled masks would be required to measure anatomical accuracy.

Let F be the fused VAAM feature map. The implemented interaction is given in Equation (3):

$$H_v = D(P_2) + q_v\, V(F \odot m_v), \quad (3)$$

where D is the ungated disease-feature path, V is the vein-conditioned path and $m_v=\sigma(\ell_v)$ is the soft predicted vein probability. The residual disease path prevents an uncertain vein estimate from removing all appearance evidence. No distance transform of the predicted vein map is used inside the evaluated VAAM implementation.

D. MorphoReli proxy branches

Four branch embeddings are retained until fusion:

- z_t: local texture and surface evidence from multi-rate convolutions;
- z_v: pseudo-venation evidence from VAAM;
- z_m: leaf-mask and boundary evidence; and
- z_g: global-average-pooled encoder appearance together with five descriptors recomputed from the transformed leaf and boundary masks: bounding-box aspect ratio, compactness, normalized horizontal and vertical centroid coordinates, and horizontal reflection overlap.

Each branch supplies a scalar heuristic quality estimate q_k to its gate. The confirmatory implementation supervises all four estimates using the prespecified automatic targets in Section 5.6. For shared context representation z_c, branch index $k\in\{t,v,m,g\}$, learned projections W_k, W_f, logistic sigmoid σ, and semicolon denoting concatenation, the branch gates are defined in Equation (4) and the fused representation in Equation (5):

$$g_k = \sigma(W_k\,[z_k\,;\, z_c\,;\, q_k]) \quad (4)$$

$$z_f = LayerNorm(W_f\,[g_t\,z_t\,;\ g_v\,z_v\,;\ g_m\,z_m\,;\ g_g\,z_g\,;\ z_t\,;\ z_v\,;\ z_m\,;\ z_g]) \quad (5)$$

LayerNorm denotes layer normalization. Each branch embedding and gate has 128 components; gate multiplication is element-wise, and the context is the mean of the four branch embeddings. The gate input has 257 components: the branch, context and scalar quality. Concatenating gated and ungated embeddings produces 1,024 components, projected to 256 and normalized as shown above. The classifier then applies the sigmoid linear unit (SiLU), dropout with probability 0.25 and a two-logit linear head followed by softmax. Reported scalar gate summaries are means of their 128 components. The ungated copy preserves branch information.

E. Joint objective

Structural targets are automatically generated pseudo-labels rather than expert anatomical annotations. Let I_A, I_V and I_M indicate that ADPE-Lite, the vein branch and the full morphology branches are active, and let I_{aux} indicate that auxiliary supervision is enabled. The implemented configuration-dependent objective is given in Equation (6):

$$L = L_cls + I_A\,L_ADPE + I_aux\,[I_V\,L_vein + I_V(1 - I_M)\,0.05\,L_visibility + I_M\,L_morphology] \quad (6)$$

The auxiliary loss terms are expanded in Equation (7).

$$L_vein = 0.20(L_vein,BCE + L_vein,Dice);\ L_morphology = 0.15(L_leaf,BCE + L_leaf,Dice) + 0.10\,L_boundary,BCE + 0.05\,L_quality \quad (7)$$

where L_{cls} is class-weighted, two-class cross-entropy; identity and edge terms regularize ADPE-Lite; leaf and vein maps use binary cross-entropy (BCE) plus soft Dice; the boundary map uses BCE; and $L_{quality}$ is mean-squared error for the four gate-quality outputs. In the vein-only V1 configuration, visibility is supervised directly. In C1, C2, P1 and F1, the four quality outputs replace that separate visibility term. B0 uses classification alone, A1 adds the ADPE-Lite terms, and C0 retains the four-branch architecture while setting the auxiliary switch I_{aux} to zero. P1 uses the same losses as C2 with structural targets permuted across source groups.

For image i, the vein and leaf/boundary weights are $w_{v,i}=clip(c_{v,i},0,1)v_{v,i}$ and $w_{l,i}=clip(c_{l,i},0,1)v_{l,i}$, where c is the automatic confidence and $v\in\{0,1\}$ is the validity indicator. Weighted map losses are divided by the sum of their batch weights, with a denominator floor of 10^{-6}. Consequently, an invalid structural proposal contributes zero to its map loss while its image still contributes to classification. All coefficients and validity rules were fixed for the matrix and were not tuned on the test set.

V. Experimental Protocol

A. Reconstructed-group split construction

The canonical cleaned images were grouped using filename timestamps, capture bursts, perceptual hashes and a compact foreground-normalized appearance embedding. With seed 20260908, the automated process produced 3,443 source groups, including 19 two-image groups; every group occurs in only one of train, validation or test. The locked allocation contains 2,422 training images (330 healthy and 2,092 CLCuD), 520 validation images (71 healthy and 449 CLCuD), and 520 test images with the same class counts as validation. Random geometric and photometric transformations were sampled only during training; validation and test images remained the unchanged canonical files assigned by the manifest.

Grouping uses the ten nearest neighbours by cosine similarity after a 64-component principal-component analysis and unit-length normalization of foreground-cropped appearance features. Features concatenate 16 × 16 grayscale intensity and gradient magnitude, 8 × 8 RGB values, foreground coverage, border contact and normalized bounding-box dimensions. Same-label pairs are joined if pHash/dHash Hamming distances are at most 5/6 with similarity at least 0.97; or at most 10/10 with similarity at least 0.985; or at most 12/14 with similarity at least 0.925 and capture times within 120 seconds. Transitive links form source groups. These rules do not establish that only 19 related pairs exist, and their same-label restriction is distinct from a cross-label conflict audit.

A retrospective label-blind audit applied the retained similarity, hash and timestamp logic without using class labels. It flagged 20 candidate relationships. One candidate crossed both class and partition: a CLCuD-labelled image in validation and a healthy-labelled image in training had cosine similarity 0.9555, pHash distance 12, dHash distance 7 and a capture-time separation of 14.8 seconds. Side-by-side inspection indicates that they are likely repeated views of the same physical leaf. Because the locked allocation and trained checkpoints were not reconstructed after this finding, the pair remains a residual label and dependence risk. The evaluation is therefore described as reconstructed-group and image-level, not as proven leaf-, plant- or capture-independent validation.

B. Automatic morphology-prior generation

No manual mask correction or expert morphology adjudication is used in the evaluated experiment. On 8-bit RGB inputs, the initial foreground test is max(R,G,B) > 12. Binary closing with radius 2, hole filling, opening with radius 1, largest-component retention, a second closing with radius 2 and final hole filling produce the leaf mask. Let n be the component count, a the leaf-area ratio and b the border-contact ratio. The leaf-quality components are u_c=1 for one component and max(0.35,1/n) otherwise, u_a=clip(1-|a-0.42|/0.42,0,1), and u_b=clip(1-2.5b,0,1). Leaf confidence is c_l=0.40u_c+0.35u_a+0.25u_b.

The pseudo-vein signal combines normalized green intensity and luminance as 0.65G+0.35Y. Hessian ridge responses are evaluated at Gaussian scales 0.8, 1.4 and 2.2. The response is limited to a leaf interior eroded by four pixels, combined with normalized local contrast as 0.78 vesselness + 0.22 contrast, normalized by its 99th percentile, and binarized at the larger of the within-leaf 82nd percentile and 0.12. Connected components smaller than six pixels are removed. If r is 95th-percentile ridge contrast, s is vein sharpness, c is binary coverage and t is multiscale stability, the coverage score is u_{cov}=clip(1-|c-0.12|/0.12,0,1) and vein confidence is c_v=0.35clip(r,0,1)+0.25clip(s/0.20,0,1)+0.20u_{cov}+0.20clip(t,0,1).

Confidence values are clipped to [0,1]. Leaf/boundary supervision is valid when c_l is at least 0.80 and pseudo-vein supervision when c_v is at least 0.70; otherwise that map contributes zero auxiliary weight while the image remains in the diagnosis loss. Of 3,462 images, 3,361 leaf proposals and 2,879 pseudo-vein proposals are valid. Leaf validity is similar by class (healthy 456/472, 96.6%; CLCuD 2,905/2,990, 97.2%), whereas pseudo-vein validity differs (healthy 318/472, 67.4%; CLCuD 2,561/2,990, 85.7%). Mean c_v is 0.707 for healthy and 0.718 for CLCuD. These class differences make confidence and validity potential auxiliary class cues and motivate the restricted interpretation of P1 below. All quantities are engineering heuristics, not anatomically calibrated probabilities.

C. Model configurations and training protocol

Table III defines the controlled configuration matrix and the purpose of each comparison.

Table III. Controlled model matrix

ID	Configuration	Controlled purpose
B0	ResNet34 classifier	Standard diagnosis baseline
A1	B0 + ADPE-Lite	Adaptive preprocessing contribution
V1	B0 + VAAM and pseudo-vein loss	Pseudo-venation contribution
C0	Four-branch gated architecture with diagnosis loss only	Parameter-matched architecture control
C1	Four morphology branches with concatenation and structural losses	Fusion reference
C2	C1 + learned quality-gated fusion	Correctly paired pseudo-morphology supervision
P1	C2 architecture with cross-source-group mismatched structural targets	Supervision-specificity control
F1	ADPE-Lite + VAAM + MorphoReli fusion	Integrated proposed configuration

Note. Every identifier is evaluated with the same locked partitions, optimization budget and checkpoint-selection rule.

All eight configurations use the same locked groups, data loading, optimizer, training-transformation policy and validation-based checkpoint rule. Inputs are resized to 224 x 224 pixels. ADPE-Lite operates in [0,1] before ImageNet normalization, and each geometric transformation is applied identically to RGB images and structural pseudo-labels. AdamW optimization uses an initial learning rate of 3 x 10^-4, weight decay of 10^-4, batch size 32 and ImageNet-pretrained encoders. Each configuration is trained for 12 epochs at seeds 20260908, 20260909 and 20260910; the epoch with maximum validation macro-F1 supplies the test prediction for that run. The diagnosis loss uses inverse-frequency class weights $N/(2N_c)$, where N_c is the number of training images in class c. A cosine-annealing schedule spans the 12 epochs.

D. Metrics

The primary internal endpoint is macro-F1 at threshold 0.5. Supporting classification endpoints are balanced accuracy, class-wise precision, recall and specificity, AUROC, AP and the confusion matrix. Calibration is summarized with top-label ECE, Brier score and negative log-likelihood. Ninety-five percent CIs and paired differences use 2,000 resamples of reconstructed source groups. Holm's procedure controls multiplicity for the prespecified comparisons. Active parameter counts are calculated from modules traversed by each configuration.

E. Quantitative XAI and shortcut controls

Grad-CAM++ and Integrated Gradients were generated for B0 and F1 on the same 520 test images at seed 20260908. F1 was selected because it instantiates all three proposed components, while B0 supplies the standard diagnosis baseline; this analysis does not directly explain C2. Each model's predicted class was the attribution target, so 15 images with different B0 and F1 predictions have different attribution targets. Grad-CAM++ uses the second convolution of the final ResNet34 residual block. Integrated Gradients uses a zero-RGB baseline, 16 equally spaced interpolation points including endpoints and mean sampled gradients; absolute channel contributions are summed. The zero baseline makes exact-black background pixels receive zero Integrated Gradients attribution by construction. Both maps are min-max normalized per image with a denominator floor of 10^-8; a constant map becomes zero. Grad-CAM++ maps with less than 0.99 combined leaf/background attribution mass are

treated as degenerate and excluded from fraction comparisons, with their frequency reported separately.

The binary leaf region is leaf mask >=0.5. The pseudo-vein intervention region is (vein probability >=0.5) within the leaf. The margin region is formed by thresholding the boundary proxy at 0.25, expanding it with 9 x 9 max pooling at stride 1 with four-pixel padding, and intersecting it with the leaf. Foreground-only input is RGB multiplied by the leaf mask; background-only input is RGB multiplied by one minus the leaf mask. For pseudo-vein and margin removal, pixels inside the selected region are replaced by the mean RGB colour of the retained leaf. The five-component geometry vector is held fixed in every intervention. Paired XAI differences use 2,000 reconstructed-source-group bootstrap resamples.

Because canonical images already place the segmented leaf on a uniform black canvas, foreground-only input is nearly identical to the original and is not an independent foreground-reliance test. No area-matched random or spatially shifted within-leaf masking controls were executed. The structural-region results therefore measure sensitivity to these particular occlusions, not selective or causal use of venation or margin anatomy. Background-only results also include the unchanged geometry pathway. These constraints are part of the interpretation rather than evidence that the interventions isolate biological reasoning.

F. Matched controls and corrected uncertainty analysis

The pipeline recomputes all five geometry descriptors from the transformed leaf and boundary masks after each geometric transformation. Explicit validity masks activate leaf/boundary supervision when leaf confidence is at least 0.80 and pseudo-vein supervision when vein confidence is at least 0.70. The four gate inputs receive automatic quality targets through mean-squared error: $q_t=s_v$, $q_v=c_v$, $q_m=c_l$, and $q_g=0.7c_l+0.3$ clip$(1-|a-0.42|/0.42,0,1)$, where s_v is clipped vein sharpness and a is leaf-area ratio. The thresholds and constants are fixed engineering heuristics and were not fitted to test data.

C0 is parameter-matched to C2 but uses diagnosis loss alone. P1 retains the C2 architecture and losses while a deterministic within-batch derangement assigns the vein, leaf, boundary, confidence, validity and quality targets to an image from another reconstructed source group. RGB images, diagnosis labels and geometry inputs remain aligned. Across the 29,064 recorded mappings per seed, 93.74%–93.78% are same-label and 6.22%–6.26% are cross-label. P1 therefore mainly disrupts image-to-target spatial correspondence, but it does not cleanly remove class-level auxiliary statistics. A same-label-only P1 retraining control was not executed; consequently, C2-P1 cannot by itself separate spatial correspondence from every class-conditioned confidence cue.

The prespecified effect contrasts compare C2 with C0 and P1 on macro-F1 at threshold 0.5, average seed-paired differences across the three seeds, and adjust the two one-sided p-values by Holm's procedure at family-wise alpha 0.05. The original implementation independently resampled the shared test groups within each sampled seed. The corrected crossed bootstrap instead draws one set of 517 reconstructed test groups per replicate and applies that same group multiplicity to every sampled seed and both members of every model comparison; it independently draws three seed identifiers with replacement. All images in a selected group are retained. Ten thousand replicates use random seed 20260909. Percentile intervals use the 2.5th and 97.5th percentiles, and the one-sided tail probability is (1 + number of differences at or below zero)/(10,001). This preserves shared-case dependence across seeds and model pairs. The corrected analysis is computed from the archived per-image probabilities and requires no model retraining.

VI. Results

A. Verified data and pipeline findings

The canonical dataset contains 3,462 successfully decoded and byte-unique cleaned source images: 472 healthy-labelled and 2,990 CLCuD-labelled. These original-image records, together with the locked source-group manifest, are the sole internal dataset used for model development and evaluation.

Automated morphology processing generated a leaf mask, boundary, normalized distance map, soft vein prior and binary high-response vein map for every canonical image. Mean leaf-mask confidence was 0.940 (median 0.954), and mean algorithmic vein-visibility confidence was 0.717 (median 0.717). These confidence values are internal quality heuristics and have not been calibrated against expert anatomy. Audit panels showed coherent foreground boundaries but also confirmed that some vein priors include lamina texture; consequently, all vein agreement results are interpreted as morphology-proxy results.

B. Repeated-seed results and corrected uncertainty analysis

All eight configurations completed the equal-budget 12-epoch protocol at the three prespecified seeds. Each run produced probabilities for the same 520 test images and 517 reconstructed source groups. Recalculation from the archived probabilities reproduced the reported scores. Table IV and Fig. 7 summarize variation across training seeds.

Table IV. Locked-test performance across seeds 20260908, 20260909 and 20260910

Model	Macro-F1, mean (SD)	Balanced accuracy, mean (SD)	AUROC, mean (SD)	Brier, mean (SD)
B0	0.907 (0.005)	0.898 (0.007)	0.965 (0.004)	0.037 (0.001)
A1	0.912 (0.004)	0.909 (0.004)	**0.974 (0.002)**	**0.032 (0.002)**
V1	0.902 (0.008)	**0.911 (0.008)**	0.968 (0.005)	0.038 (0.002)
C0	0.894 (0.021)	0.892 (0.013)	0.964 (0.010)	0.040 (0.008)
C1	0.903 (0.022)	0.889 (0.015)	0.957 (0.007)	0.037 (0.005)
C2	**0.916 (0.006)**	0.903 (0.018)	0.968 (0.009)	0.032 (0.002)
P1	0.895 (0.021)	0.898 (0.020)	0.963 (0.010)	0.041 (0.009)
F1	0.901 (0.006)	0.894 (0.030)	0.960 (0.023)	0.038 (0.002)

Note. Bold values mark the largest mean for macro-F1, balanced accuracy and AUROC and the smallest mean for Brier score. A1 and C2 Brier scores differ before rounding. Standard deviations describe three trained models and are not deployment confidence intervals.

Table V. Supporting locked-test metrics at the fixed 0.5 threshold, averaged across the three seeds

Model	Healthy precision	Healthy recall	CLCuD precision	CLCuD recall	AP
B0	0.861	0.817	0.971	0.979	0.993
A1	0.857	0.840	0.975	0.978	0.995
V1	0.809	0.854	0.977	0.968	0.993
C0	0.826	0.812	0.970	0.972	0.992
C1	0.873	0.798	0.969	0.981	0.991
C2	0.892	0.822	0.972	0.984	0.994
P1	0.812	0.826	0.972	0.970	0.993
F1	0.856	0.812	0.971	0.976	0.991

Note. AP is average precision. Healthy and CLCuD recall are the class-specific sensitivities; for a two-class problem each is the other class's specificity.

Table VI. Calibration and mean confusion counts at the fixed 0.5 threshold across the three seeds

Model	ECE	NLL	TN / FP	FN / TP
B0	0.030	0.152	58.0 / 13.0	9.3 / 439.7
A1	0.025	0.133	59.7 / 11.3	10.0 / 439.0
V1	0.034	0.152	60.7 / 10.3	14.3 / 434.7
C0	0.031	0.156	57.7 / 13.3	12.7 / 436.3
C1	0.034	0.169	56.7 / 14.3	8.7 / 440.3
C2	0.023	0.132	58.3 / 12.7	7.3 / 441.7
P1	0.044	0.157	58.7 / 12.3	13.7 / 435.3
F1	0.036	0.149	57.7 / 13.3	10.7 / 438.3

Note. ECE is 10-bin expected calibration error and NLL is negative log-likelihood. TN, FP, FN and TP treat CLCuD as the positive class and are means of integer counts from three seeds.

Table VII. Validation-only threshold sensitivity across the three seeds

Model	Validation-selected threshold range	Test macro-F1, mean	Test balanced accuracy, mean
B0	0.19–0.50	0.905	0.890
A1	0.11–0.35	0.919	0.896
V1	0.10–0.43	0.919	0.905
C0	0.08–0.42	0.912	0.892
C1	0.25–0.51	0.914	0.889
C2	0.08–0.40	0.914	0.889
P1	0.17–0.47	0.909	0.889
F1	0.10–0.58	0.911	0.886

Note. Each threshold was selected only on that seed's validation probabilities by maximum macro-F1 and then applied once to its locked test probabilities. This sensitivity analysis is exploratory and is separate from the prespecified fixed-threshold endpoint.

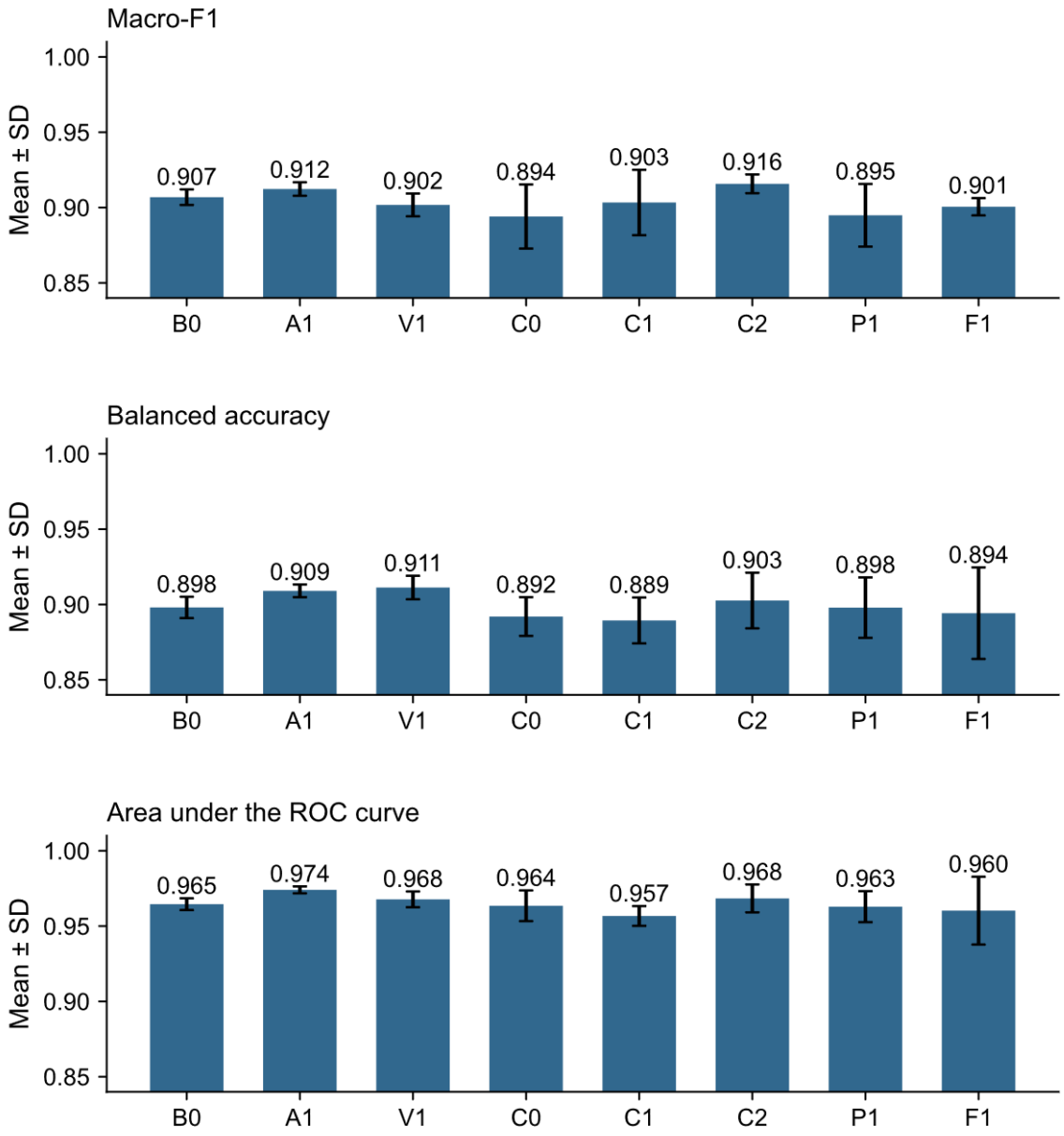

Fig. 7. Mean locked-test performance across the three prespecified runs; error bars show one standard deviation across seeds. These bars do not represent confidence intervals for deployment performance

C2 had the largest mean macro-F1, 0.916 (SD 0.006). Its seed-paired differences were positive at all three seeds: 0.0382, 0.0231 and 0.0038 versus C0, and 0.0115, 0.0384 and 0.0127 versus P1. Under the corrected crossed bootstrap, the mean C2-C0 difference was 0.0217 (95% CI -0.0027 to 0.0497; one-sided p=0.0450) and the mean C2-P1 difference was 0.0209 (-0.0048 to 0.0483; p=0.0527). Holm adjustment gave p=0.0900 for both. Because both intervals include zero and adjusted p-values exceed 0.05, the prespecified confirmatory criterion was not met. The repeated-seed direction remains a positive descriptive trend rather than confirmatory evidence.

Exploratory crossed comparisons with stronger simple baselines were also inconclusive. C2-B0 macro-F1 differed by 0.0089 (95% CI -0.0154 to 0.0346; one-sided p=0.2189), and C2-A1 by 0.0034 (-0.0166 to 0.0240; p=0.3512). A1 retained the highest mean AUROC and the smallest mean Brier

score, while V1 retained the highest mean balanced accuracy. Validation-selected thresholds varied substantially across seeds and did not preserve a unique C2 advantage; Table VII keeps this exploratory operating-point analysis separate from the fixed-threshold primary endpoint.

C. Full-cohort quantitative XAI

The B0 and F1 checkpoints at seed 20260908 were evaluated on the same 520 test images; their predicted classes differed for 15 images. Grad-CAM++ was non-degenerate for 517 B0 images and all 520 F1 images. Across images with valid paired maps, F1 increased the Grad-CAM++ leaf-attribution fraction by 0.2014 (95% reconstructed-group bootstrap CI 0.1869–0.2155) and reduced background attribution by the same amount. Pseudo-vein overlap increased by 0.0409 (0.0380–0.0441), while margin overlap decreased by 0.0153 (-0.0174 to -0.0132). Integrated Gradients changed leaf attribution by 0.0008 (0.0006–0.0009), decreased pseudo-vein overlap by 0.0038 (-0.0058 to -0.0020), and increased margin overlap by 0.0180 (0.0155–0.0206). The attribution families disagree about which internal pseudo-regions gain evidence. The Integrated Gradients foreground shift is also partly constrained by the zero baseline and exact-black canvas, so it is not treated as independent corroboration of the larger Grad-CAM++ shift.

Original AUROC was 0.9651 for B0 and 0.9769 for F1. Foreground-only AUROC was 0.9662 and 0.9765, as expected for an intervention that is almost identical to the canonical black-background input. Background-only AUROC was 0.5097 and 0.6063, but the unchanged geometry vector remains available. Pseudo-vein removal changed AUROC by +0.0079 for B0 and -0.0090 for F1; margin removal changed it by -0.0069 and -0.0369. Without area-matched random or shifted controls, the larger F1 reductions show sensitivity to these occlusions but do not establish selective use of venation or margin anatomy.

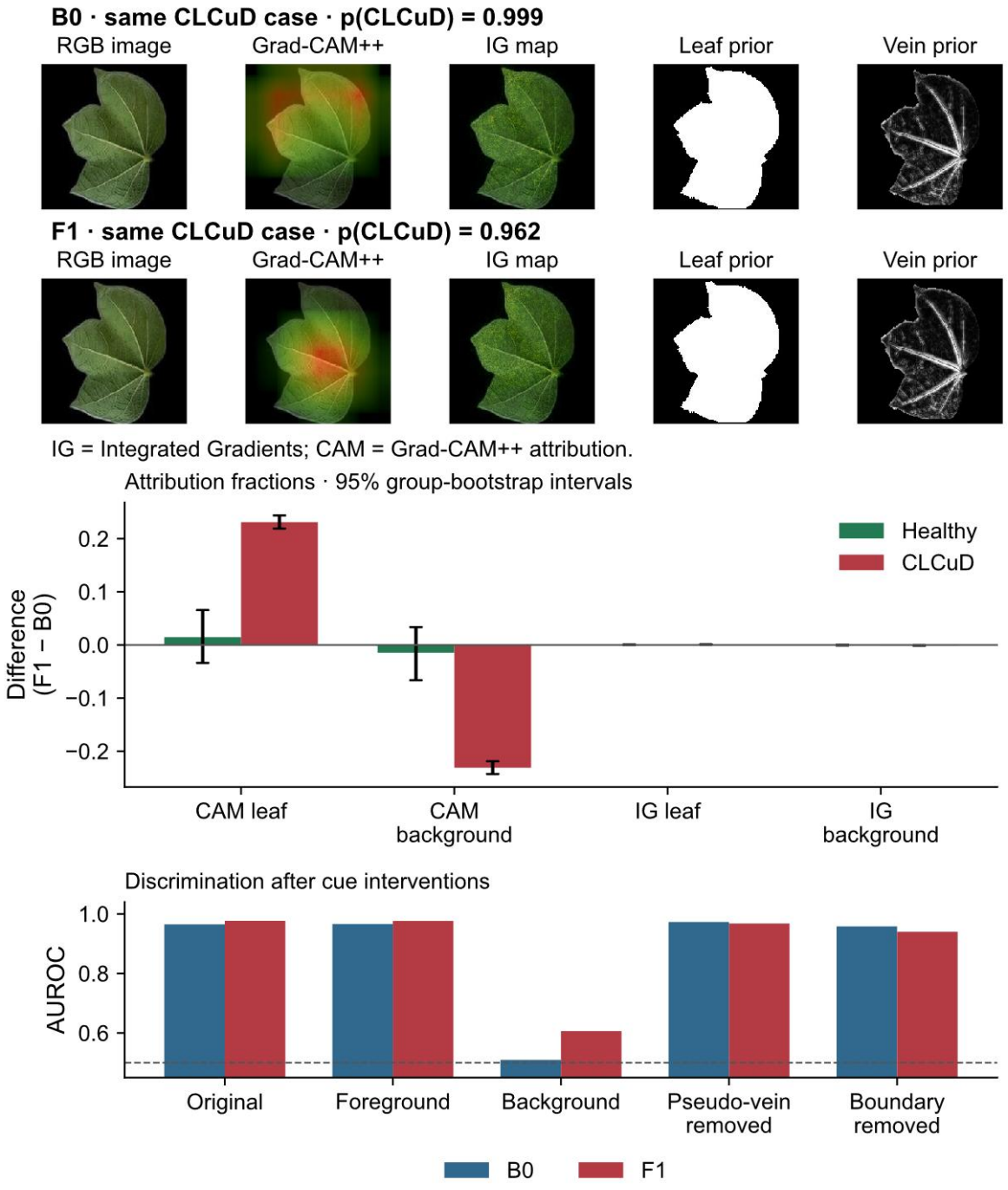

Fig. 8. Full-cohort paired B0-F1 attribution and intervention summary for seed 20260908. Intervals resample reconstructed source groups. Grad-CAM++ fractions exclude invalid paired maps. Pseudo-vein and margin regions are automatically generated proxies. The interventions retain geometry and do not include area-matched random-mask controls

VII. Discussion

C2 produced the highest mean macro-F1 and exceeded C0 and P1 at each of the three seeds. The corrected crossed bootstrap, however, gives intervals that include zero and Holm-adjusted p-values of 0.090. The data therefore support a consistent positive internal trend, not the original confirmatory significance claim. Capacity alone is addressed by C0, while P1 weakens image-to-target

correspondence; because P1 is not restricted to same-label pairs and auxiliary confidence distributions differ by class, it does not isolate spatial correspondence perfectly.

The configuration matrix shows different operating strengths. A1 produced the highest mean AUROC and lowest probability error, V1 produced the highest mean balanced accuracy, and C2 produced the highest fixed-threshold macro-F1. Exploratory crossed comparisons did not establish C2 superiority over B0 or A1. The architecture should therefore be presented as an auditable morphology-guided design with promising component behaviour, rather than as a universally superior classifier.

The XAI evidence is behavioural rather than anatomical proof. Grad-CAM++ showed a substantial shift from the black background toward the estimated leaf, whereas Integrated Gradients showed a much smaller shift that is partly determined by its zero baseline. Structural occlusions affected F1 more than B0 in some conditions, but the lack of area-matched controls and retention of geometry prevent selective or causal anatomical interpretation.

The contribution remains specific. ResNet34, feature pyramids, attention, segmentation losses, residual connections and attribution methods are established. MorphoReli-CLCuD contributes an explicit CLCuD-oriented decomposition, automatic structural targets, heuristic reliability outputs, inspectable branch maps and a matched-control protocol. The current evidence defines what should be confirmed with label-adjudicated grouping, same-label permutation controls, expert morphology annotations and independent field cohorts.

VIII. Scope and Future Validation

The present evidence concerns image-level healthy-versus-CLCuD classification within one recovered, cleaned collection. Physical plant and leaf identifiers were unavailable. Reconstructed capture groups reduce some related-image leakage, but the label-blind audit identified one likely repeated cross-label, cross-partition pair, so plant-, leaf- and capture-level independence is not established. Automatic leaf, boundary, pseudo-vein and quality targets are engineering proxies rather than expert-reviewed anatomy. The 180 x 180 canonical resolution supports coarse structural guidance but cannot recover fine secondary venation that is absent from the stored image.

Future validation should adjudicate the identified cross-label pair, rebuild the grouped allocation, and then use expert structural annotations, higher-resolution images and true identifiers for plant, leaf, capture session, location, device, season and cultivar. A same-label-only P1 control and area-matched XAI masks should accompany any confirmatory rerun. Ordered severity should be added only when expert-reviewed grades are genuinely ordered, and whole-plant severity requires plant-level observations. The related manuscript, “Cross-Domain Robustness of Cotton Leaf Curl Disease Classifiers,” reports frozen external and hard-negative evaluation of the resulting checkpoints; the present paper reports architecture development and internal validation only.

IX. Conclusion

MorphoReli-CLCuD provides an explicit and testable way to combine bounded adaptive preprocessing, pseudo-venation, leaf-boundary and global-geometry evidence for healthy-versus-CLCuD image classification. Across three prespecified seeds, C2 achieved the highest mean macro-F1 and showed positive differences from the same-capacity diagnosis-only and mismatched-target controls. After preserving shared test-group dependence across seeds, those differences did not satisfy the prespecified confirmatory criterion. A1 led AUROC and Brier score, while V1 led balanced

accuracy. The architecture, retained probabilities and corrected analyses provide a reproducible basis for a future adjudicated, group-independent confirmation; current conclusions are limited to positive internal trends within the frozen canonical cohort.

X. Data Availability

The cotton-leaf images analysed in this study are original field photographs collected by the authors and are not publicly deposited. The data supporting the findings, including the split manifest, source-group identifiers (IDs), checksums and derived analysis records, are available from the corresponding author on reasonable request, subject to institutional and data-use conditions.

XI. Code and Reproducibility

The source code is not publicly released. The experimental records comprise locked split manifests, source-group identifiers (IDs), source and bundle checksums, configuration definitions, prespecified seeds, training histories, per-image probabilities, bootstrap analyses and explainability (XAI) outputs. These records can be shared with reviewers or on reasonable request.

One archival limitation applies. For the F1 seed-20260909 run, the complete 12-epoch history, validation predictions, test predictions and metrics were retained, but the model checkpoint was not. Its reported numerical outcome therefore remains auditable, while exact future re-inference for that one run is unavailable. This limitation does not affect the C2–C0 or C2–P1 primary comparisons because all selected checkpoints and predictions for those comparisons are present.

Source-code hashes identify the inspected implementation but do not independently certify the exact source revision executed for every checkpoint. The original dependency snapshot for the initial runs was not retained; the final F1 run used Python 3.12.13, PyTorch 2.10.0+cu128 and a Tesla T4 GPU. Reproduction would require the original source images, saved manifests, retained outputs, available checkpoints and matching software dependencies.

References

[1]. R. W. Briddon and P. G. Markham, “Cotton leaf curl virus disease,” Virus Research, vol. 71, nos. 1–2, pp. 151–159, 2000, doi: 10.1016/S0168-1702(00)00195-7.

[2]. M. S. Rehman et al., “Genetics and genomics of Cotton Leaf Curl Disease, its viral causal agents and whitefly vector,” Frontiers in Plant Science, vol. 8, Art. no. 1157, 2017, doi: 10.3389/fpls.2017.01157.

[3]. J. Kumar et al., “An insight into Cotton Leaf Curl Multan betasatellite,” Viruses, vol. 9, no. 10, Art. no. 280, 2017, doi: 10.3390/v9100280.

[4]. A. Magsi et al., “A novel framework for disease severity level identification of cotton plant using machine learning techniques,” International Journal of Scientific & Technology Research, vol. 10, no. 5, pp. 367–373, 2021.

[5]. R. Nazeer et al., “Detection of cotton leaf curl disease’s susceptibility scale level based on deep learning,” Journal of Cloud Computing, vol. 13, Art. no. 50, 2024, doi: 10.1186/s13677-023-00582-9.

[6]. N. Ahmad, S. Qadri, and N. Akhtar, “A multimodal fusion framework to diagnose cotton leaf curl virus using machine vision techniques,” Cogent Food & Agriculture, vol. 10, Art. no. 2339572, 2024, doi: 10.1080/23311932.2024.2339572.

[7]. S. K. Noon et al., “Computationally light deep learning framework to recognize cotton leaf diseases,” Journal of Intelligent & Fuzzy Systems, vol. 40, no. 6, pp. 12383–12398, 2021, doi: 10.3233/JIFS-210516.

[8]. S. Otiya, P. Faldu, and P. Goel, "Cotton leaf disease classification using deep convolution neural network with explainable AI," in Proc. International Conference on Sustainable Communication Networks and Application (ICSCNA), 2023, doi: 10.1109/ICSCNA58489.2023.10370214.

[9]. G. Biju et al., "Identification of Cotton Leaf Curl Disease Using CNN and Vision Transformer," in Intelligent Systems and Applications, pp. 670–688, Springer, 2024, doi: 10.1007/978-3-031-47724-9_44.

[10]. N. Sabari et al., "MultiCotNet: a novel multispatial attention-based deep learning architecture for cotton leaf disease classification," Journal of Cotton Research, vol. 9, Art. no. 29, 2026, doi: 10.1186/s42397-026-00276-y.

[11]. N. Ullah, I. De Falco, and G. Sannino, "Cotton leaf disease detection via dual-backbone CNN–Transformer fusion with quantitative XAI comparison," Electronics, vol. 15, no. 16, Art. no. 3650, 2026, doi: 10.3390/electronics15163650.

[12]. T. Chen et al., "The intelligent detection method for similar cotton diseases based on high- and low-frequency feature enhancement and attention-guided fusion," Engineering Applications of Artificial Intelligence, vol. 168, Art. no. 114099, 2026, doi: 10.1016/j.engappai.2026.114099.

[13]. S. Naveena and K. Kavitha, "EnsembleNet model: A hybrid approach for classifying Gossypium herbaceum leaf diseases using image feature analysis," Engineering Applications of Artificial Intelligence, vol. 164, Art. no. 113225, 2026, doi: 10.1016/j.engappai.2025.113225.

[14]. S. Das, G. Prethija, and C. Sudha, "LeafFusionNet: a hybrid deep learning approach for robust plant disease detection," Frontiers in Artificial Intelligence, vol. 9, Art. no. 1773329, 2026, doi: 10.3389/frai.2026.1773329.

[15]. V. V. Waykule and D. S. Bormane, "LeafMorph-X: A Morphology-Aware End-to-End Deep Learning Framework for Accurate Tomato Leaf Disease Detection," ES Food and Agroforestry, vol. 24, Art. no. 2188, 2026, doi: 10.30919/faf2188.

[16]. M. E. Haque et al., "An attention enhanced CNN ensemble for interpretable and accurate cotton leaf disease classification," Scientific Reports, vol. 16, Art. no. 4476, 2026, doi: 10.1038/s41598-025-34713-w.

[17]. P. Bishshash, M. A. S. Nirob, M. H. Shikder, A. Sarower, T. Bhuiyan, and S. R. H. Noori, "A comprehensive cotton leaf disease dataset for enhanced detection and classification," Data in Brief, vol. 57, Art. no. 110913, 2024, doi: 10.1016/j.dib.2024.110913. Dataset: SAR-CLD-2024, doi: 10.17632/b3jy2p6k8w.2.

[18]. S. Saha, M. J. Israel, and M. Bhattacharya, "Deep Inception: Cotton Leaf Disease Classification with Convolutional Precision Neural Network," in Data Science and Exploration in Artificial Intelligence, CCIS 2690, pp. 309–314, Springer, 2026, doi: 10.1007/978-3-032-19321-6_28.

[19]. S. M. M. R. Swapno et al., "Explainable transformer framework for fast cotton leaf diagnostics and fabric defect detection," iScience, vol. 29, no. 2, Art. no. 114411, 2026, doi: 10.1016/j.isci.2025.114411.

[20]. M. Naeem et al., "Automatic prediction of cotton leaf's diseases using deep learning techniques," Scientific Reports, online ahead of print, 2026, doi: 10.1038/s41598-026-60523-9.

[21]. S. Kaufman et al., "Leakage in data mining: formulation, detection, and avoidance," ACM Transactions on Knowledge Discovery from Data, vol. 6, no. 4, Art. no. 15, 2012, doi: 10.1145/2382577.2382579.

[22]. A. Ahmad, A. El Gamal, and D. Saraswat, "Toward generalization of deep learning-based plant disease identification under controlled and field conditions," IEEE Access, vol. 11, pp. 9042–9057, 2023, doi: 10.1109/ACCESS.2023.3240100.

[23]. A. Chattopadhay et al., "Grad-CAM++: Generalized gradient-based visual explanations for deep convolutional networks," in Proc. IEEE Winter Conference on Applications of Computer Vision (WACV), 2018, doi: 10.1109/WACV.2018.00097.

[24]. M. Sundararajan, A. Taly, and Q. Yan, "Axiomatic attribution for deep networks," in Proc. 34th International Conference on Machine Learning (ICML), PMLR, vol. 70, pp. 3319–3328, 2017.

[25]. V. Petsiuk, A. Das, and K. Saenko, “RISE: Randomized input sampling for explanation of black-box models,” in Proc. British Machine Vision Conference (BMVC), 2018.

[26]. T.-Y. Lin et al., “Feature Pyramid Networks for object detection,” in Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017, doi: 10.1109/CVPR.2017.106.