IOCAE 2026 Conference
The 1st International Online Conference on Aerospace
16–17 April 2026 | Online
MDPI

Robust AI-Assisted HPC Borescope Inspection: Evaluating Stage-Level Decisions and Viewpoint Generalisation

Tse Yun Sang
Department Of Electronic and Computer Engineering
The Hong Kong University of Science and Technology
Contact: ystseae@connect.ust.hk

THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY

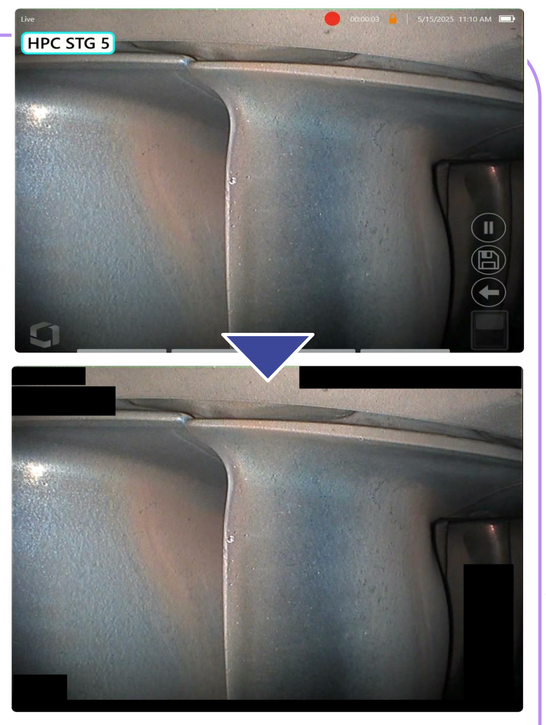
1 INTRODUCTION & AIM

Manual Inspection Bottleneck:
- Borescope inspection of aircraft HPC blades is safety-critical but relies entirely on manual expert assessment

Misalignment of Frame-Level vs Stage-Level Decisions:
- Most AI research reports frame-level accuracy, misaligned with the real inspection unit: the blade row stage. Inspectors decide per stage, not per frame

UI Shortcut Learning Risks:
- On-screen UI overlays (stage labels, timestamps, icons) cause AI shortcut learning. Engine-specific colour and sharpness variation causes poor generalisation across engines and view angles

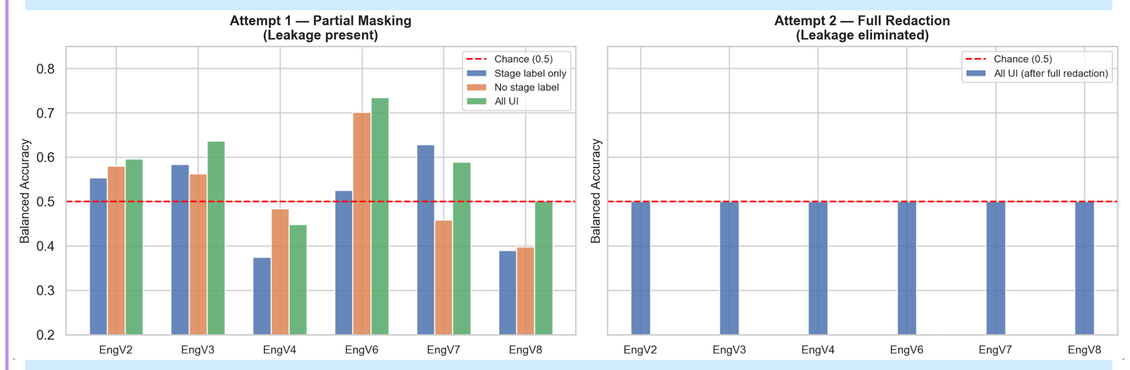
Research Aim:
- This work focuses on dent and nick defects in HPC compressor blade rows. Design and validate an audit-ready stage-level AI triage system operating on real workflow footage with provable anti-shortcut controls
- Evaluate generalisation across three independent axes: engine identity, blade stage geometry, and camera viewpoint (Leading edge to tip <> Root to Platform)



Raw borescope footage contains UI overlays that enable AI shortcuts. After black-fill redaction, only blade content remains visible

3 UI LEAKAGE CONTROL

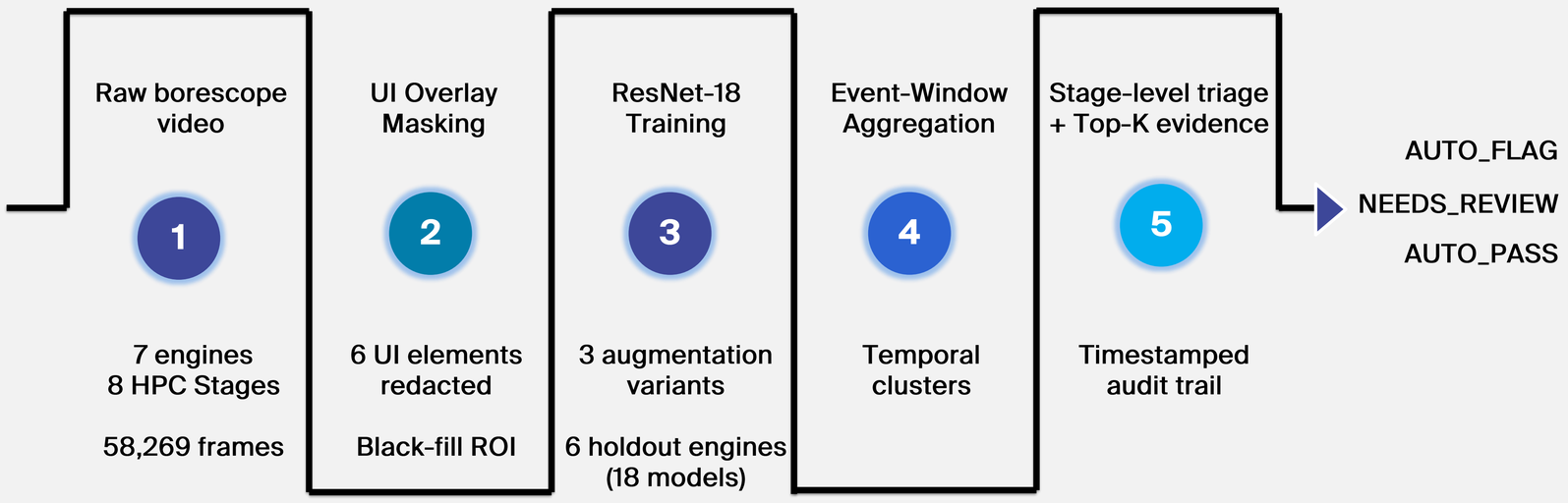
A ResNet-18 classifier trained exclusively on UI overlay regions collapses to chance (0.500) after full redaction, confirming complete elimination of shortcut leakage



Before: BA up to 0.93 **After: BA = 0.500**

Full UI redaction collapses balanced accuracy to exactly 0.500 for all six engines — stage labels, icons and timestamps carry no residual signal after masking

2 METHOD

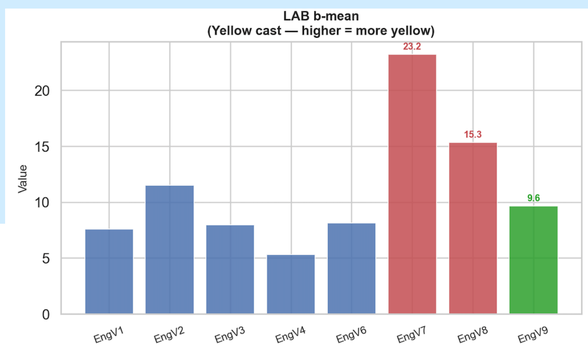


Evaluated under engine-holdout, stage-holdout, and view-holdout protocols independently

EngV1 comprises exclusively defective stages and was used for training only. EngV9 was reserved as an unseen validation engine. Primary triage evaluation covers 6 holdout engines (EngV2–EngV8 excluding EngV5), hence 6 engines appear in evaluation figures.

4 DOMAIN SHIFT

4 visual domain shift metrics measured from 200 sampled frames per engine. EngV7 and EngV8 exhibit extreme yellow cast (the degree of overall yellow tint caused by camera and lighting differences between engines), motivating grayscale augmentation



EngV9 closest to EngV2 (distance = 0.83)

EngV7 most isolated

5 MAIN RESULTS – Stage-Level Triage Results (Grayscale)

0.867 Sensitivity 0.703-0.947

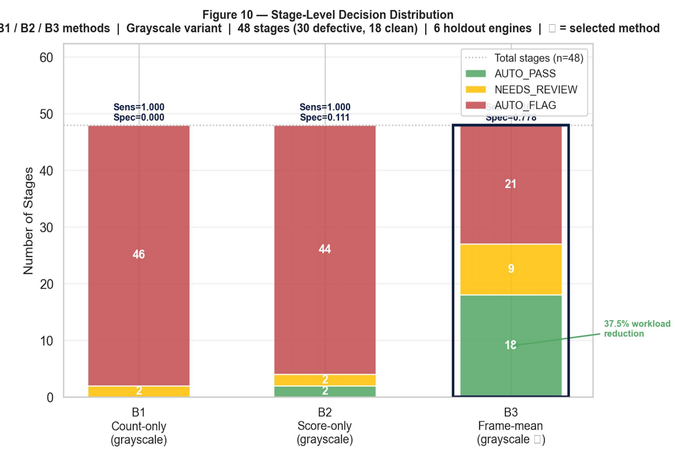
0.778 Specificity 0.548-0.910

37.5 % AUTO_PASS Rate Workload Reduction

48 stages · 6 holdout engines · 143,892 frames scored at full stage coverage · B3 frame-mean & grayscale · Wilson 95% Cis
Decision robustness: flip rate 10.0% ± 3.2% at σ=0.05

Unseen EngV9	1.000 Specificity	0.686 Sensitivity

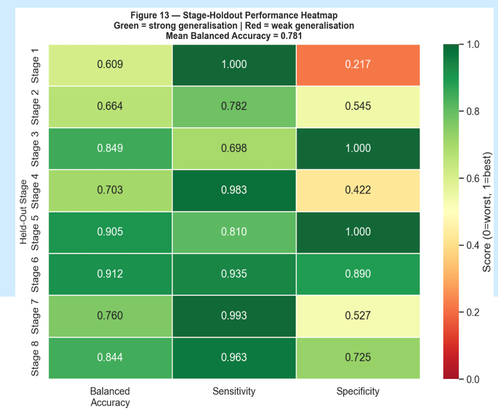
Clean stage correctly AUTO_PASS by every model: zero false alarms on a completely unseen engine



B3 frame-mean is the only method achieving both sensitivity (0.867) and workload reduction (37.5% AUTO_PASS). Count-only and score-only methods flag nearly all stages with no workload benefit

6 GENERALISATION

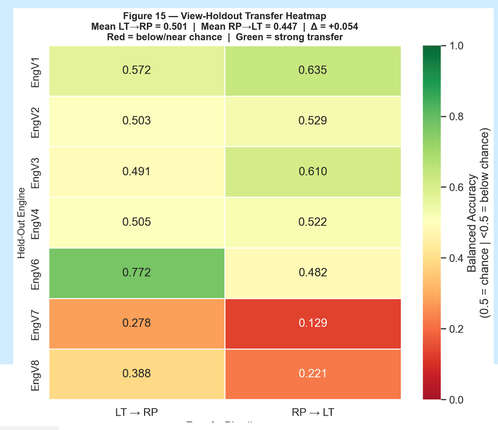
Stage-Holdout: Unseen Blade Rows



Train on 7 stages, test on 1 held-out stage. All 8 folds exceed chance.

Stages 5 and 6 generalise most strongly

Blade View-Holdout: Cross-Viewpoint Transfer



Leading Edge to Tip → Root to Platform mean BA = 0.501.

EngV6 shows strong directional transfer (0.773).

EngV7/V8 fall below chance.
Which is consistent with extreme domain shift

Conclusion:

Audit-ready triage: sensitivity 0.867, specificity 0.778, 37.5% workload reduction on 48-stage and 6 engines evaluation with timestamped Top-K evidence frames.	UI shortcut leakage eliminated: balanced accuracy collapses to exactly 0.500 after full redaction — verified by cheat-test protocol across all engines.	Stage-holdout confirms generalisation to unseen blade stages: mean BA 0.781 across all 8 HPC stages, all folds above chance.	View shift is the primary unsolved challenge: 4 of 7 engines fail cross-viewpoint transfer. Domain distance predicts unseen engine performance — EngV9 clusters nearest to EngV2 (distance = 0.83)

Future Work:

- Frame-level annotation
- View-specific models
- Additional unseen validation engines with mixed clean/defective stages