This MICCAI paper is the Open Access version, provided by the MICCAI Society. It is identical to the accepted version, except for the format and this watermark; the final published version is available on SpringerLink.

Supplementary Material for: Clean AUC Is Not Enough: Architecture-Specific Failure Modes of Tuberculosis Classifiers under Digitisation Artifacts

Freedmore Sidume[1 [0000−0002−4470−4814]]✉, Stewart Muchuchuti[1 [0000−0003−3890−0445]], Gibson Chengetenai[1], Rabson Tamukate[1], and Kushatha Ntwaetsile[2 [0000−0002−2424−2130]]

[1] School of Computing & Information Systems, Botswana School of Business Sciences, Gaborone, Botswana
freedmores@bac.ac.bw

[2] Computing & Information Systems, Botswana International University of Science and Technology, Palapye, Botswana

This supplementary material accompanies the camera-ready submission to the MIRASOL workshop (MICCAI 2026 satellite event). It provides the representative artifact image grid referenced in Section 3.2 of the main paper, a visual verification of severe artifacts, and an additional Grad-CAM++ case for ResNet50 under severe defocus blur, complementing the EfficientNet-B0 case reported in the main text.

S1. Representative digitisation artifacts across all severities

Fig. 1 shows a single validation radiograph rendered under all ten artifact families (defocus blur, motion blur, Gaussian noise, low contrast, under-exposure, JPEG compression, downsampling, rotation, scanner streak, and vignetting) at mild, moderate, and severe levels, alongside the clean image. These are the same deterministic, paired corruptions used throughout the robustness evaluation.

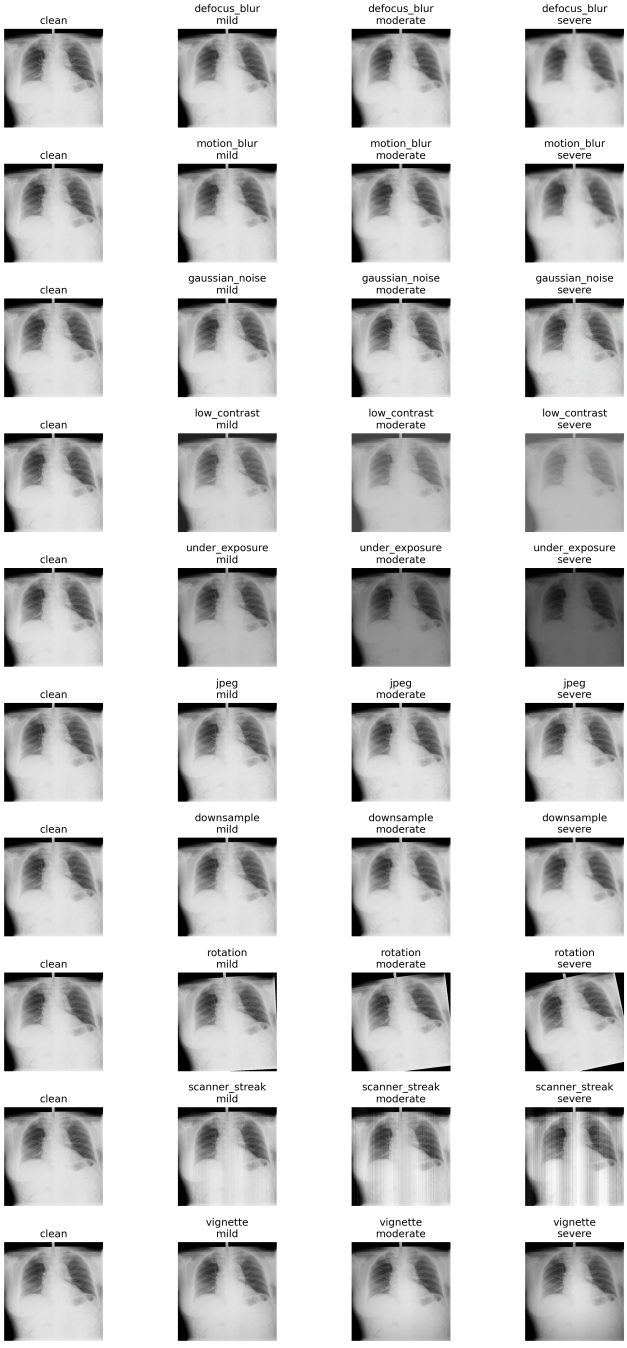

Fig. 1. Representative image grid: one validation radiograph under all ten digitisation artifact families at mild, moderate, and severe levels, plus the clean reference image.

S2. Visual verification of severe artifacts

Fig. 2 shows a second qualitative check of the severe-level corruptions used in the main evaluation (blur, noise, low contrast, under-exposure, rotation, and JPEG compression), confirming that the corruption pipeline produces visually plausible severe degradations rather than numerical artifacts alone.

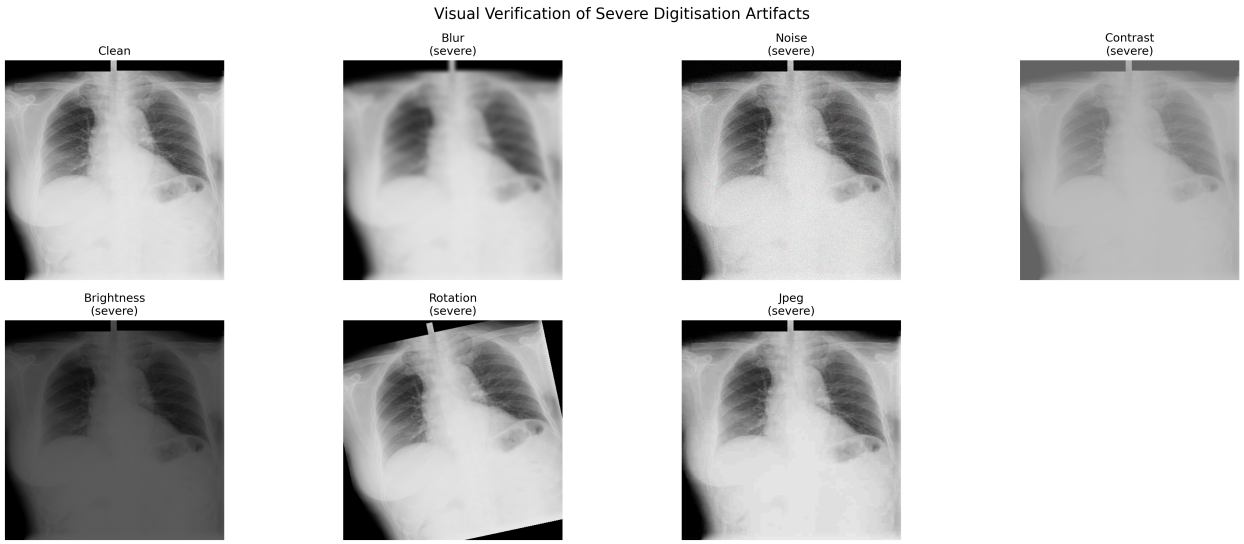

Fig. 2. Visual verification of severe digitisation artifacts on a validation radiograph.

S3. ResNet50 explanation collapse under severe defocus blur

Fig. 3 shows a representative ResNet50 case under severe defocus blur, the artifact that produced this model's largest sensitivity loss (Section 4.2 of the main paper). As with the EfficientNet-B0 case in the main text (Fig. 2), this case was selected to illustrate the failure pattern and should be read as an illustrative example rather than as representative of the full sampled cohort; aggregate explanation-stability statistics for ResNet50 under severe defocus blur are reported in Fig. 3 of the main paper.

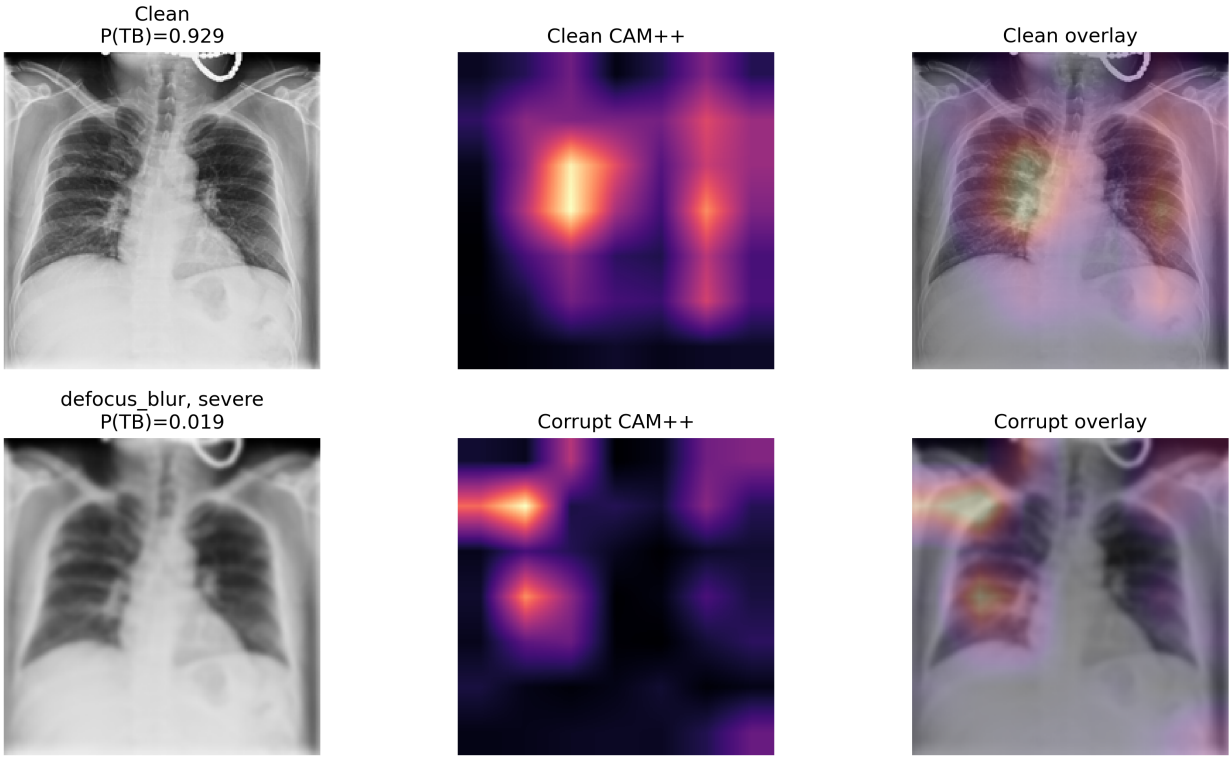

Fig. 3. ResNet50 Grad-CAM++ case under severe defocus blur. The clean prediction is correctly positive with lung-centred saliency; under severe blur the predicted probability falls below the operating threshold (a false negative) and the saliency map is displaced.