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.

MoonScan: An AI-Based Clinical Decision-Support System for Chest X-Ray Analysis in a Low-Resource African Setting

Maïmouna Dicko[1] and Ousmane Ly[2,3,4]

[1] Université Internationale Privée d'Excellence (UIPE), Bamako, Mali
maimounaardo@gmail.com

[2] Faculté de Médecine et d'Odonto-Stomatologie (FMOS), Université des Sciences, Techniques et Technologies de Bamako (USTTB), Bamako, Mali
ousmane.ly@usttb.edu.ml

[3] Unité de Télémédecine Militaire, Centre Médico-Chirurgical des Armées (CMCA), Bamako, Mali

[4] CIDIS – Centre interdisciplinaire de développement international en santé, Faculté de Médecine et des Sciences de la Santé, Université de Sherbrooke – Campus de Longueuil, Canada
Ousmane.Ly@usherbrooke.ca

Abstract. Chest X-ray interpretation remains challenging in many low-resource healthcare settings, where access to radiologists and locally annotated medical data is limited. This paper presents MoonScan, an AI-based clinical decision-support pipeline developed and internally evaluated using 1,365 posteroanterior chest radiographs from 1,365 patients. The images were acquired during routine care at the Medical Imaging Department of Polyclinique Pasteur, a private clinic in Bamako, Mali, between June 2024 and February 2025. MoonScan segments the heart and thorax to automatically calculate the cardiothoracic index (CTI). The same workflow also includes binary classification, multi-label pathology classification, and anomaly detection. The segmentation model achieved Dice scores of 0.89 for the heart and 0.94 for the thorax. CTI-based screening identified 113 adult cases of cardiomegaly among examinations initially labelled as normal, after reassessment by the supervising radiologist. The binary classifier achieved 87.0% accuracy but only 0.47 recall for pathological examinations. The multi-label classification and anomaly detection modules were limited by severe class imbalance and the small number of annotated images. Overall, this work highlights both the potential and the challenges of developing clinically useful AI systems from locally collected data in low-resource healthcare settings.

Keywords: Chest X-ray, Medical AI, Low-resource healthcare, Cardio-Thoracic Index, Clinical decision support, Mali.

1 Introduction

Chest radiography is widely used because it is fast, affordable, and available in many healthcare facilities[1]. However, in low-resource settings, limited access to radiologists can delay interpretation and affect patient management[2, 3].
Artificial intelligence has shown promising results in chest X-ray analysis [4–6], but many existing models are trained on large datasets collected predominantly in high-income countries. Their performance may not generalize directly to African clinical settings, where locally annotated data are scarce and patient populations, disease prevalence, and acquisition conditions may differ [7, 8].
This paper presents MoonScan, an AI-based clinical decision-support pipeline developed and internally evaluated using chest radiographs acquired at Polyclinique Pasteur in Bamako, Mali. MoonScan combines heart and thorax segmentation for automated cardiothoracic index (CTI) computation with binary classification, multi-label pathology classification, and anomaly detection. It is intended to support, rather than replace, physician interpretation.
This work contributes a multi-model pipeline developed from locally collected data, a CTI module that flagged 113 adult examinations initially labelled as normal for radiologist reassessment, and an evaluation of the challenges posed by limited annotations and severe class imbalance.

2 Related Work

Large public datasets such as ChestX-ray14 [4], CheXpert [5], and PadChest [9] have supported advances in chest radiograph classification and abnormality detection. Models such as CheXNet [5] and Swin Transformer [10] have shown strong performance, but their generalizability across populations and clinical settings remains a concern [7].
Automated heart and thorax segmentation has also been used to estimate the cardiothoracic index (CTI)[11]. MoonScan combines automated CTI computation, binary classification, multi-label classification, and anomaly detection in a single pipeline developed from routinely acquired chest radiographs in Mali.

3 Dataset

3.1 Dataset Collection

The retrospective MoonScan dataset consists of 1,365 posteroanterior (PA) chest radiographs acquired during routine clinical care at the Medical Imaging Department of Polyclinique Pasteur, a private clinic in Bamako, Mali, between June 2024 and February 2025. From approximately 1,500 radiographs, 135 were excluded after review by the supervising radiologist because they were lateral views, duplicate examinations, or of insufficient quality. Only one radiograph per patient was retained, ensuring

patient-level separation across the training, validation, and test sets. Table 1 summarizes the composition of the collected dataset.

Table 1. Dataset composition.

Category	Count	%
Normal	1,197	87.7%
Pathological	152	11.1%
Medical devices	16	1.2%
Total retained	1,365	100%

3.2 Annotation

A subset of 144 radiographs, comprising 74 pathological and 70 normal examinations, was manually annotated in CVAT [12] by the researcher under the direct supervision of one radiologist. The annotations included heart and lung masks, localized radiological signs, medical devices, and image-level attributes such as sex, age range, and proposed diagnosis. Task-specific annotations were used for anatomical segmentation, multi-label classification, and anomaly detection. Figure 1 shows an example of the annotation process.

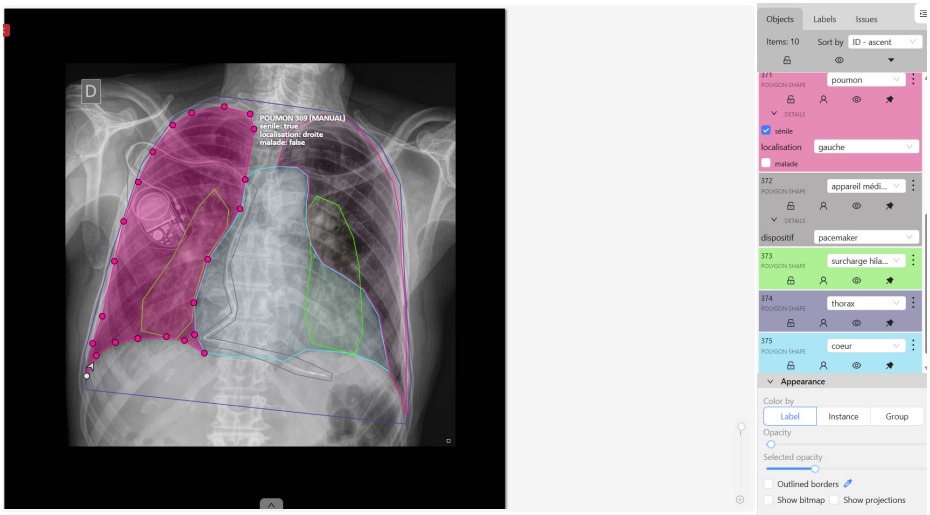

Fig. 1. Example of manual annotation in CVAT showing anatomical masks, localized radiological signs, and a medical device. Identifying information was removed.

3.3 Ethics and Data Protection

This retrospective Master's study was conducted at Polyclinique Pasteur with institutional authorization and radiologist supervision. At the time of its conduct, it was formally submitted to the institution's internal Scientific Committee for compliance with patient anonymity and medical-record confidentiality requirements. It used pre-existing routine-care radiographs and involved no patient recruitment, additional procedures, or modification of care. Before analysis, identifying image filenames were replaced with coded study identifiers (Patient_XXX_xrayY). The correspondence table was stored separately and was not shared. Individual informed consent was not collected because the analysis was retrospective, used pre-existing de-identified data, involved no patient intervention, and was conducted under institutional authorization and oversight.

4 Methodology

MoonScan is a modular research prototype composed of four AI modules (Fig. 2). For each chest radiograph, the modules operate independently, and their outputs are presented in a single interface. All results require physician review before report generation; the system was not evaluated for autonomous clinical decision-making.
Owing to the limited size of the task-specific datasets, fixed training/validation splits were used for the segmentation, binary classification, and anomaly detection modules, without separate test sets. The multi-label classification module used separate training, validation, and test sets. All subsets were mutually exclusive, and no image used for validation or testing appeared in the corresponding training set.

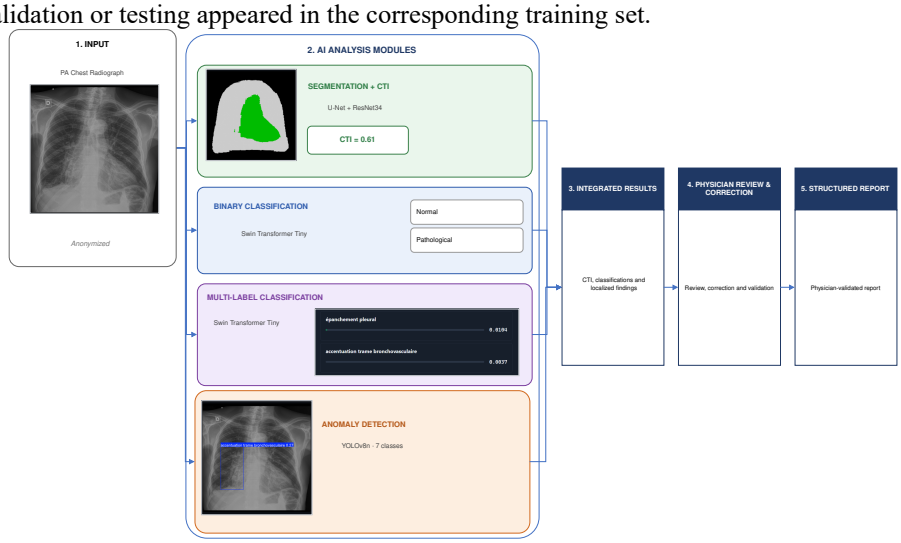

Fig. 2. MoonScan pipeline. The system combines heart and thorax segmentation for CTI computation, binary classification, multi-label pathology classification, and anomaly detection before physician review.

4.1 Heart and Thorax Segmentation

Heart and thorax segmentation was performed using a U-Net [13] with a ResNet34 encoder pre-trained on ImageNet [14]. The 117 manually annotated radiographs were divided into 93 training and 24 validation images. The predicted masks were used for automated Cardio-Thoracic Index (CTI) computation.

4.2 Automated CTI Computation

The Cardio-Thoracic Index (CTI) was computed from the predicted heart and thorax masks as the ratio of the transverse cardiac diameter to the internal thoracic diameter. Examinations with CTI > 0.5 were flagged as potential cardiomegaly. In the implemented decision rule, automatic cardiomegaly labelling was restricted to patients older than 15 years.

4.3 Binary Classification

Binary classification of normal and pathological examinations was performed using a pretrained Swin Transformer Tiny [10]. The dataset comprised 1,349 radiographs divided into 1,079 training and 270 validation images. Two variants, with and without data augmentation, were trained and compared using the same split.

4.4 Multi-label Classification

A pretrained Swin Transformer Tiny model with eight independent outputs (one for each targeted radiological sign) was trained on 187 pathological radiographs, including 74 manually annotated images and 113 CTI-reclassified cardiomegaly cases (see 5.1). A 70/15/15 training/validation/test split, class weighting, and early stopping were used. Table 2 shows the class distribution.

Table 2. Distribution of radiological signs in the multi-label dataset.

Radiological sign	Positive radiographs
Cardiomegaly	142 (29 annotated + 113 CTI-reclassified)
Bronchovascular accentuation	22
Pleural effusion	17
Interstitial opacity	13
Hilar congestion	13
Alveolar opacity	9
Alveolo-interstitial opacity	6

Mediastinal opacity	1

4.5 Anomaly Detection

Radiological abnormalities were localized using YOLOv8n [15], trained on 64 annotated radiographs using an 80/20 training/validation split. The model used seven output classes; cardiomegaly was excluded because it was assessed through CTI computation rather than localized lesion detection.

4.6 Clinical Interface

The models were integrated into a Flask-based web interface displaying patient information, CTI measurements, classification outputs, localized findings, and confidence scores. The physician could review and correct the results before generating a structured report. Figure 3 illustrates the clinical interface used to present the integrated model outputs for physician review.

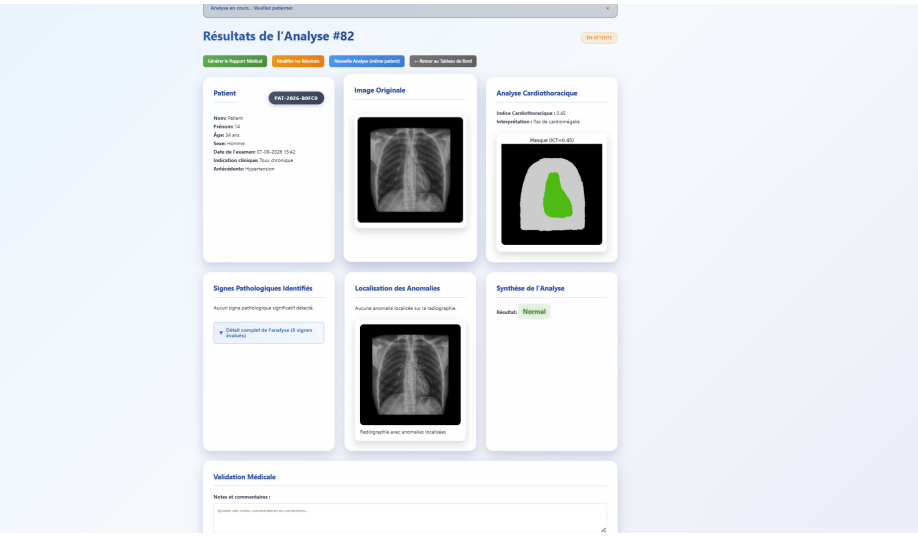

Fig. 3. MoonScan clinical interface. Results from all modules are presented to the physician for review and validation before report export.

5 Results

5.1 Segmentation and CTI

Table 3 reports the segmentation performance on the 24-image validation set.

Table 3. Segmentation performance (validation set, n=24).

Structure	Dice	IoU
Heart	0.89	0.81
Thorax	0.94	0.89
Background	0.97	0.94

The CTI module flagged 209 of the 1,197 examinations initially labelled as normal. Following reassessment by the supervising radiologist, 113 adult examinations were reclassified as cardiomegaly. The remaining 96 examinations involved patients aged 15 years or younger and were not automatically reclassified (Fig. 4).

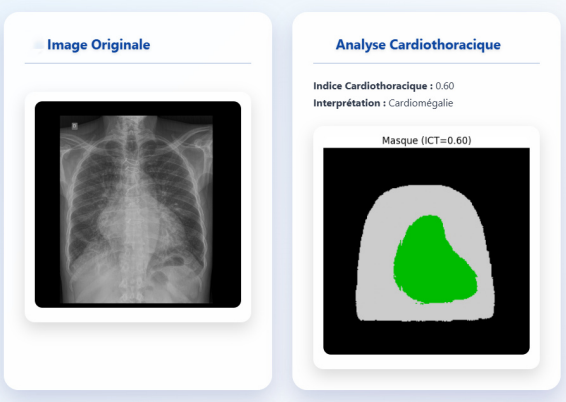

Fig. 4. Example of CTI-assisted reassessment. A PA chest radiograph initially labelled normal was flagged by the automated CTI module (CTI = 0.60) and subsequently reclassified as cardiomegaly during radiologist reassessment.

5.2 Binary Classification

Following CTI-based reclassification, the dataset comprised 1,084 normal and 265 pathological radiographs. On the 270-image validation set, the non-augmented model achieved 87.0% accuracy but missed 33 of 62 pathological examinations (pathological recall: 0.47; Table 4). The augmented model reduced false negatives to 19 and increased recall to 0.69 but produced 17 false positives compared with 2 for the non-augmented model. Neither variant achieved sufficient sensitivity for autonomous triage.

Table 4. Confusion matrix for the non-augmented binary classifier on the validation set (n = 270).

	Predicted Normal	Predicted Pathological
True Normal (n=208)	206	2
True Pathological (n=62)	33	29

5.3 Multi-label Classification

The multi-label model achieved an F1-micro of 0.71 and an F1-macro of 0.23. Performance was strongly influenced by class imbalance: cardiomegaly achieved the highest class-level F1-score (0.86), whereas signs represented by very few positive examples approached an F1-score of zero. The module therefore serves only as a proof of feasibility and requires substantially more balanced annotated data before clinical evaluation.

5.4 Anomaly Detection

YOLOv8n achieved a mAP50 of 21.8% and a mAP50–95 of 8.91%. Performance remained limited because of the small and severely imbalanced annotated dataset. The module was retained only as a proof-of-feasibility experiment for approximate localization and was not considered suitable for clinical use.

6 Discussion

MoonScan's main proof-of-concept result was the segmentation-based CTI module, which supported the reassessment of examinations initially labelled as normal and the reclassification of 113 adult cardiomegaly cases. However, this retrospective result does not establish prospective clinical effectiveness, and all outputs require physician confirmation.

The study was limited to one centre and small, severely imbalanced task-specific datasets. Segmentation, binary classification, and detection relied on fixed validation splits without independent test sets or cross-validation, making their performance estimates sensitive to the selected partition. Binary pathological recall remained insufficient, while multi-label classification and detection were strongly affected by rare classes. CTI reassessment was conducted by one non-blinded radiologist without inter-reader agreement analysis. External, multi-reader, and prospective evaluation is therefore required before clinical deployment.

7 Conclusion

MoonScan demonstrates the feasibility of developing a multi-model chest radiograph analysis pipeline using locally collected data from Mali. Segmentation and CTI-assisted reassessment showed promising internal results, whereas binary classification, multi-label classification, and anomaly detection remained exploratory. Multi-centre, multi-reader, and prospective validation is required before clinical use.

Acknowledgments. The authors thank Polyclinique Pasteur for hosting the master's research internship during which this study was conducted and for facilitating access to the clinical environment and data. The authors also thank Dr Aboubacar S. N'Diaye for his clinical supervision and radiological review. This research received no external funding.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Çallı, E., Sogancioglu, E., van Ginneken, B., van Leeuwen, K.G., Murphy, K.: Deep learning for chest X-ray analysis: A survey. Medical Image Analysis. 72, 102125 (2021). https://doi.org/10.1016/j.media.2021.102125.
2. Karera, A., Davidson, F., Engel-Hills, P.: Operational challenges and collaborative solutions in radiology image interpretation: perspectives from imaging departments in a low-resource setting. J Med Radiat Sci. 71, 564–572 (2024). https://doi.org/10.1002/jmrs.815.
3. Ofori-Manteaw, B., Ofori, E.K., Dzefi-Tettey, K., Ago, J.L., Angmorterh, S.K., Acheampong, F.: Radiologist workforce challenges and the burden of image interpretation in Ghana: Perspectives of frontline doctors and implications for healthcare delivery. PLOS Glob Public Health. 6, e0006456 (2026). https://doi.org/10.1371/journal.pgph.0006456.
4. Wang, X., Peng, Y., Lu, L., Lu, Z., Bagheri, M., Summers, R.M.: ChestX-ray8: Hospital-scale chest X-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 3462–3471 (2017). https://doi.org/10.1109/CVPR.2017.369.
5. Rajpurkar, P., Irvin, J., Zhu, K., Yang, B., Mehta, H., Duan, T., Ding, D., Bagul, A., Langlotz, C., Shpanskaya, K., others: CheXNet: Radiologist-Level Pneumonia Detection on Chest X-Rays with Deep Learning. arXiv preprint arXiv:1711.05225. (2017).
6. Irvin, J., Rajpurkar, P., Ko, M., Yu, Y., Ciurea-Ilcus, S., Chute, C., Marklund, H., Haghgoo, B., Ball, R., Shpanskaya, K., Seekins, J., Mong, D.A., Halabi, S.S., Sandberg, J.K., Jones, R., Larson, D.B., Langlotz, C.P., Patel, B.N., Lungren, M.P., Ng, A.Y.: CheXpert: A Large Chest Radiograph Dataset with Uncertainty Labels and Expert Comparison. Proceedings of the AAAI Conference on Artificial Intelligence. 33, 590–597 (2019). https://doi.org/10.1609/aaai.v33i01.3301590.
7. Mungai, B.N., Joekes, E., Masini, E., Obasi, A., Manduku, V., Mugi, B., Ong'angò, J., Kirathe, D., Kiplimo, R., Sitienei, J., Oronje, R., Morton, B., Squire, S.B., MacPherson, P., IMPALA Consortium: "If not TB, what could it be?" Chest X-ray findings from the 2016 Kenya Tuberculosis Prevalence Survey. Thorax. 76, 607–614 (2021). https://doi.org/10.1136/thoraxjnl-2020-216123.
8. (PDF) Limitations of Public Chest Radiography Datasets for Artificial Intelligence: Label Quality, Domain Shift, Bias and Evaluation Challenges,

https://www.researchgate.net/publication/395650024_Limitations_of_Public_Chest_Radiography_Datasets_for_Artificial_Intelligence_Label_Quality_Domain_Shift_Bias_and_Evaluation_Challenges, last accessed 2026/08/11. https://doi.org/10.48550/arXiv.2509.15107.

9. Bustos, A., Pertusa, A., Salinas, J.-M., de la Iglesia-Vayá, M.: PadChest: A large chest x-ray image dataset with multi-label annotated reports. Med Image Anal. 66, 101797 (2020). https://doi.org/10.1016/j.media.2020.101797.
10. Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., Lin, S., Guo, B.: Swin Transformer: Hierarchical Vision Transformer using Shifted Windows. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). pp. 10012–10022 (2021).
11. Yoshida, K., Takamatsu, A., Matsubara, T., Kitagawa, T., Toshima, F., Tanaka, R., Gabata, T.: Deep learning-based cardiothoracic ratio measurement on chest radiograph: accuracy improvement without self-annotation. Quant Imaging Med Surg. 13, 6546–6554 (2023). https://doi.org/10.21037/qims-23-187.
12. Leading Image & Video Data Annotation Platform | CVAT, https://www.cvat.ai, last accessed 2025/06/26.
13. Ronneberger, O., Fischer, P., Brox, T.: U-Net: Convolutional Networks for Biomedical Image Segmentation, https://arxiv.org/abs/1505.04597v1, last accessed 2026/08/14.
14. He, K., Zhang, X., Ren, S., Sun, J.: Deep Residual Learning for Image Recognition. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 770–778 (2016). https://doi.org/10.1109/CVPR.2016.90.
15. GitHub - ultralytics/yolov8: Ultralytics YOLOv8 discovery and quickstart for detection, segmentation, classification, pose, OBB, and tracking. · GitHub, https://github.com/ultralytics/yolov8, last accessed 2026/08/12.