© Pol J Radiol 2026; 91: e45
DOI: https://doi.org/10.5114/pjr/211474

Received: 27.09.2025
Accepted: 29.09.2025
Published: 30.01.2026

Polish Journal of Radiology

http://www.polradiol.com

Letter to the Editor

Comments on data balance, test-set accounting, and thresholded reporting in an adaptive convolution neural network model for tuberculosis detection and diagnosis using semantic segmentation

Emmanuel Pio Pastore[A,C,D,E,F]

Department of Biology, Ecology and Earth Science, University of Calabria, Rende (CS), Italy

Dear Editor,

I read with interest the article by Salkade and Rathi [1] on a segmentation plus classification pipeline for tuberculosis detection on chest radiographs. The clinical aim is timely. Two points deserve clarification because they affect the credibility of the figures: what was actually evaluated and whether the evaluation relied on a heldout test or, at any stage, on the training set.

The Methods state that the authors sampled 700 tuberculosis and 700 normal radiographs from the TB PortAL and split them 80/10/10 into training, validation, and test. Under that split, the nominal test would contain 140 images (70 tuberculosis and 70 normal). However, the confusion matrices in Figure 10 show totals that do not match this specification. For the proposed model (panel F), the matrix reads 348, 2, 2, 348, i.e., 350 per class and 700 overall. For several baselines (panels A-E and G), the normal row sums to about 700, while the tuberculosis row sums to about 140, for an overall total of around 840, implying a 5 : 1 imbalance. These discrepancies suggest that some matrices may have been computed on the training set, or datasets other than the heldout test. In machine learning, it is not customary to present confusion matrices on the training set, because this practice overstates performance and hides generalization errors.

To dispel any ambiguity, it would help to state for each model the exact sample counts used for the matrices, confirm that all results refer exclusively to the same heldout test split, and clarify whether any patient or sitelevel overlap could have leaked between training and test. Beyond that, external validation is necessary for claims approaching perfection. Without an independent cohort, the reported ROCAUC of 0.99 suggests overfitting: nearperfect confusion matrices are likely inflated by distributional alignment or inadvertent reuse of data from the training set.

For deployment on chest radiography, evaluation should also reflect class imbalance and a fixed operating point. Precision–recall AUC, predictive values at plausible prevalences, a threshold chosen on validation and then locked for test, and basic calibration would make the results easier to interpret and more transferable to clinical workflows.

References

1. Salkade SAS, Rathi SV. An adaptive convolution neural network model for tuberculosis detection and diagnosis using semantic segmentation. Pol J Radiol 2025; 90: e124-e137. DOI: 10.5114/pjr/200628.
2. Saito T, Rehmsmeier M. The precision–recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. PLoS One 2015; 10: e0118432. DOI: 10.1371/journal.pone.0118432.
3. World Health Organization. Use of computer-aided detection software for tuberculosis screening and triage: policy update. Geneva: World Health Organization; 2025.

Correspondence address:
Emmanuel Pio Pastore, Department of Biology, Ecology and Earth Science, University of Calabria, Via P. Bucci 1, 87036 Rende (CS), Italy,
e-mail: emmanuel.pastore17@gmail.com

Authors' contribution:
A Study design · **B** Data collection · **C** Statistical analysis · **D** Data interpretation · **E** Manuscript preparation · **F** Literature search · **G** Funds collection

This is an Open Access journal, all articles are distributed under the terms of the Creative Commons Attribution-Noncommercial-No Derivatives 4.0 International (CC BY-NC-ND 4.0). License (https://creativecommons.org/licenses/by-nc-nd/4.0/).