|  |
| --- |
| An Artificial Intelligence-Powered Annotation Methodology for a Sparsely Labeled Computed Tomography Lung Cancer Dataset: A Use Case for Non-Expert Observers**Eleftherios Trivizakis**1#\*, and **Kostas Marias**1,21Computational Biomedicine Laboratory (CBML), Foundation for Research and Technology Hellas (FORTH), 70013, Heraklion, Greece2Department of Electrical and Computer Engineering, Hellenic Mediterranean University, 71410, Heraklion, Greece# Presenting author: Eleftherios Trivizakis, email: trivizakis@ics.forth.gr\* Corresponding author: Eleftherios Trivizakis, email: trivizakis@ics.forth.gr |

abstract

Annotating cancer lesions is a highly challenging task [1] with a myriad of drawbacks, such as inter-observer variability, time-consuming, high monetary cost, and lack of radiologists’ availability. Tissue differentiation, either through pixel-based or bounding box annotations, is essential for image analysis in oncology [2]. Adapting machine learning models requires large, high-quality, and annotated datasets. YOLO [3] are ever-evolving detection architectures, incorporating state-of-the-art layers and advanced convergence mechanisms. Additionally, YOLO pre-trained models are widely accessible, allowing for model adaptation on size-limited oncology datasets.

This study used the publicly available NSCLC Radiogenomics dataset [4], which comprises 211 computed tomography examinations with a subset of them annotated (n=142). Moreover, the location of the non-small cell lung cancer tumors is provided as a soft label (n=211) of the lung lobe anatomy. 69 unlabeled examinations, the soft label of the tumor location, and a simple atlas of the lungs were provided to a non-expert observer. 20 examinations were rejected due to extremely large lesions, extensive metastases, or lesion uncertainties. 1238 image slices were annotated, augmenting the training set by 48.8%. In total, 3771 slices (n=141) were available as a training set, 609 (n=20) as a validation set, and 787 (n=30) as a testing set. These sets were used for adapting, assessing the fitting status, and evaluating several variations of YOLOv8 architecture, respectively. Two types of experiments were conducted: (a) using the 142 annotated examinations of the original dataset, and (b) adding annotated images from a non-clinician to the training set.

The model trained with the first cohort (n=92) (a) exhibits a detection performance of 0.65 mAP and 0.52 recall. By including the non-expert observer annotated images (b) in the training set (n=141), a mAP of 0.81 and a recall of 0.73 were achieved.

Non-expert observers' annotations can be beneficial in medical image analysis, potentially having a positive impact on the model adaptation phase or accelerating the tedious annotation process.

**REFERENCES**

[1] Trivizakis, E., Souglakos, J., Karantanas, A. and Marias, K., 2021. Deep radiotranscriptomics of non-small cell lung carcinoma for assessing molecular and histology subtypes with a data-driven analysis. *Diagnostics*, *11*(12), p.2383.

[2] Trivizakis, E., Koutroumpa, N.M., Souglakos, J., Karantanas, A., Zervakis, M. and Marias, K., 2023. Radiotranscriptomics of non-small cell lung carcinoma for assessing high-level clinical outcomes using a machine learning-derived multi-modal signature. *BioMedical Engineering OnLine*, *22*(1), p.125.

[3] Ultralytics, “ultralytics/ultralytics: YOLOv8,” Github.https://github.com/ultralytics/ultralytics (accessed Apr.27, 2023).

[4] Bakr, S., Gevaert, O., Echegaray, S., Ayers, K., Zhou, M., Shafiq, M., Zheng, H., Benson, J.A., Zhang, W., Leung, A.N. and Kadoch, M., 2018. A radiogenomic dataset of non-small cell lung cancer. *Scientific data*, *5*(1), pp.1-9.