Semantic-Guided Robust Imaging Biomarker for Early Detection of Lung Cancer

Luoting Zhuang[1], Seyed Mohammad Hossein Tabatabaei[1], Ramin Salehi-Rad[2], Linh M. Tran[2], Denise Aberle[1], Ashley Prosper[1], William Hsu[1]

[1]Medical & Imaging Informatics, Department of Radiological Sciences, David Geffen School of Medicine at UCLA

[2]Department of Medicine, Division of Pulmonology and Critical Care, David Geffen School of Medicine at UCLA, Los Angeles, CA 90095, USA

Deep learning has emerged as a powerful tool in lung cancer risk assessment, utilizing its capability to extract intricate features from Computed Tomography (CT). While imaging-based models have demonstrated impressive performance, the lack of interpretability, sensitivity to variations in imaging, as well as shortcut learning, limits generalizability and real-world clinical application. Therefore, in this study, we utilize the semantic features gathered from radiologists' assessments of nodules to guide the model in learning clinically meaningful and robust features for predicting lung cancer.

We obtained 938 low-dose CT scans containing at least one nodule from the National Lung Screening Trial (NLST). Radiologists annotated the location of the nodule and semantic features, including nodule shape, texture, internal, and perinodular features, for a total of 1246 nodules. Additionally, the Lung Image Database Consortium (LIDC) dataset contains diagnostic and screening CT scans from 1018 cases, with 2625 lesions annotated for subjective nodule characteristics, including sphericity, lobulation, consistency, internal structure, margin, and spiculation. Two external datasets were obtained, including 56 cases from UCLA Health and 83 from the LUNGx Challenge.

In order to learn the clinically meaningful features, we trained a Contrastive Language-Image Pretraining (CLIP) model, which pulls images and text pairs close together. The CLIP model was initialized by OpenAI CLIP (ViT-B32) weights and finetuned to align imaging and semantic features and predict the one-year lung cancer diagnosis. For imaging input, we placed a 50*50*50 bounding box around the nodule and obtained 2D nodule slices from nine directions, all passing through the nodule centroid. This method meets the model input criteria while preserving as much of the 3D structure information as possible. Furthermore, we converted semantic features into text using Gemini, enabling flexibility in combining datasets with missing values and mismatched semantic features. The model was fine-tuned using Low-Rank Adaptation (LoRA), a parameter-efficient technique that required fine-tuning only 0.3% of the trainable parameters, effectively preventing overfitting. We selected 20% of nodules from NLST as the held-out test set. The rest of the nodules from NLST were combined with LIDC data for training.

We performed a 5-fold cross-validation and evaluated the performance of the one-year diagnosis of lung cancer with AUROC and AUPRC. The result is compared to both Sybil and the cancer imaging biomarker foundation model. In comparison to the foundation model (AUROC: 0.83, AUPRC: 0.59), our model demonstrated higher performance in the NLST test set, with an AUROC of 0.88 and AUPRC of 0.75. In addition, our model demonstrates superior generalizability, outperforming the foundation model on the UCLA dataset, with a 24.8% increase in AUROC and a 21.1% increase in AUPRC, and on LUNGx, with 12.8% and 11.3% increases, respectively. It also surpasses the Sybil model in both datasets, particularly in LUNGx, with a 14.6% increase in AUROC and 15% in AUPRC. Most importantly, using CLIP, we are also able to obtain predictions on semantic features, such as nodule margin (AUROC: 0.81), nodule consistency (0.81), nodule margin conspicuity (0.87), eccentric calcification (0.79), pleural attachment (0.84), paracicatricial emphysema (0.73), etc. This capability provides interpretability that could aid clinicians in understanding the underlying meaning of model predictions.

Our study has yielded promising results in accurately predicting risk and providing explainable outputs. This is achieved by using semantic features, which help the model to learn clinically relevant information. This approach prevents the model from taking shortcut learning and improves its ability to generalize.