Lung Cancer Risk Prediction with Multimodal Deep Learning

Fabienne Greier, Julian Link, Prasanga Dhungel, and Johannes Brandt

Technical University of Munich
{fabienne.greier, julian.link, prasanga.dhungel, johannes.brandt}@tum.de

Abstract. Lung cancer is the leading cause of cancer-related mortality worldwide. For its effective treatment, it is crucial to catch it at an early stage. Lung cancer screening, employing low-dose Computed Tomography (CT) scans, has demonstrated considerable efficacy in decreasing mortality rates associated with lung cancer by identifying malignant pulmonary nodules at an earlier phase. In addition, Machine Learning and Deep learning approaches have shown promise in classifying lung cancer from Lung CT. However, most of the existing approaches are fundamentally limited as they fail to capture the information provided by the metadata of the patients. Multimodal learning generally outperforms single-modality models in disease diagnosis and prediction. This is particularly true in lung cancer, which is heavily contextualized through non-imaging risk factors.
In this work, we show that combining different levels of features, including clinical metadata and imaging data at the lung and nodule levels, provides a good estimation of malignancy, surpassing the predictive capability of utilizing these data sources independently. Furthermore, our analysis reveals superior performance when employing a feature extractor pretrained specifically on lung CTs compared to one pretrained on alternative domains.

Keywords: Lung Cancer · Computed Tomography · Deep Learning · Multimodality.

1 Introduction

Lung Cancer stands preeminent as the primary contributor to cancer-associated fatalities across the globe [1]. For its effective treatment, it is crucial to catch it at an early stage. Lung cancer screening (LCS) employing yearly low-dose CT imaging has been shown to diminish lung cancer mortality by early identifying malignant lung nodules [2].

Screening for lung cancer generates large amounts of CT images that require evaluation by radiologists. Since each scan needs careful examination, the workload for radiologists can be overwhelming due to the sheer number of scans. As screening datasets become more abundant and accessible for analysis, machine learning and deep learning algorithms have significantly progressed in predicting lung cancer [3, 4]. By training algorithms with diverse examples of lung nodules, these technologies can potentially serve as objective diagnostic tools to assist radiologists in interpreting CT scans more efficiently and accurately, particularly in challenging cases.

While initial applications of deep learning methods have demonstrated potential in distinguishing lung cancer from lung CT scans, many existing approaches overlook the essential patient metadata, thus limiting their effectiveness. In the field of deep learning, models fusing data from different modalities outperform their unimodal counterparts since more information is aggregated [5]. Prediction of lung cancer with CT image would also benefit from the metadata of the patients, as it would enrich the model with a comprehensive understanding of individual health profiles. With the hypothesis that CT images and Clinical Data Elements (CDE) provide complementary information for predicting lung cancer, we propose integrating CT image features and CDE features into a unified framework in this work.

Related Works Myriads of studies have utilized deep learning-based models for lung cancer prediction. These studies typically fall into two main categories: lung-level and nodule-level predictions.

Nodule-level predictions Nodule-level predictions involve two primary stages: nodule detection and subsequent diagnosis, where an assessment of malignancy probability is allocated to identified nodules. Subsequently, the case is categorized as either cancerous or non-cancerous [6, 7]. Liao et al. [8] proposed a 3D CNN for nodule detection and classification. Although nodule-level methodologies have demonstrated heightened accuracy compared to lung-level models in forecasting malignancy risk for candidate nodules, determining malignancy risk within a nodule hinges heavily on the model's capacity to accurately detect the nodule initially. Given that many training datasets lack ample instances of large lesions, nodule detection may struggle to identify them effectively, potentially resulting in false negatives during nodule-level prediction.

Lung-level prediction Lung-level prediction uses the entire CT scan as an input [9, 10]. Jiang et al. [11] proposed a 3D CNN network to predict lung cancer on CT images. Since lung-level methods can leverage global information from the entire CT image, lung-level methods are robust to predict lung cancer when there are big lesions [12] as they have enough information to compare the lesion size to that of a CT image.

Recently, there has been a proliferation of research endeavors that combine predictions at both the nodule and lung levels. Ardila et al. [3] use a Region Proposal Network (RPN) to identify candidate nodules and a 3D Inception network to calculate the malignancy score. Gao et al. [13] combined lung CT screens with patients metadata. Aslani et al. [12] combine lung-level and nodule-level information across the time and integrate patient's metadata to predict the malignancy probability. Similarly, Li et al. [14] proposed a longitudinal multimodal transformer that integrates CT images with the latent clinical signatures for nodule classification.

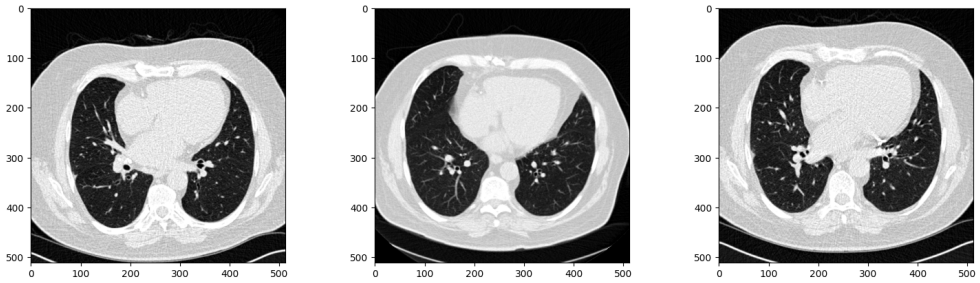
2 Experimental Setup

The National Lung Screening Trial (NLST) dataset [2] consists of patient data from $54,000$ high-risk individuals. Data was collected over the course of three years and contains CT or X-ray images as well as information about the patient's demographic.

Table 1: Demographics of the NLST dataset used for training.

Features	**Values**
Age (average)	61.4 years
Height (average)	67.87 inches
Weight (average)	183.46 lbs
Chronic Obstructive Pulmonary Disease	
Diagnosed	5.1 %
Not diagnosed	94.9 %
Ethnic	
Neither Hispanic nor Latino	97.5 %
Hispanic or Latino	1.75 %
Race	
White	90.8 %
Black	4.45 %
Asian	2.05 %
Pacific	0.36 %
Native American	0.36 %
Percentage of male/female	59 %/41 %
Number of years smoking (average)	39.84 years
Started smoking (average)	16.7 years
Pack years (average)	55.97
Cigarettes per day (average)	28.44
Smoking status at trial start	
Smoking	48.2 %
Not smoking	51.8 %
Education	
Less than high school	6.08 %
High school graduate or GED	23.78 %
Post-high school training, excluding college	13.91 %
Associate's degree	22.97 %
Bachelor's degree	16.74 %
Graduate	14.22 %
Other	2.31 %

Scans were halted after three years or on the first confirmation of malignancy through biopsy. Three exemplary lung CT scans can be seen in fig. 1. For this paper, only the CT arm of the study was analyzed.



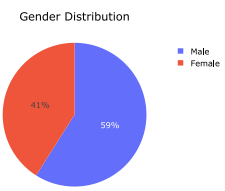
(a) Lung Segment 1 (b) Lung Segment 2 (c) Lung Segment 3

Fig. 1: Exemplary lung segments of the NLST dataset

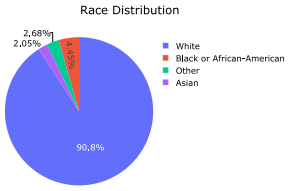
The dataset has various biases that must be considered when training a model. As shown in fig. 2a, the dataset has more male than female patients. For the NLST, only patients over the age of 55 years were considered, which can be seen in fig. 2d. In the NLST dataset, the percentage of patients with lung cancer is only 3.85% (fig. 2c). The NLST primarily includes individuals who have a history of heavy smoking, which significantly increases the risk of lung cancer. This implies that the dataset may only partially represent the general population, especially non-smokers or light smokers. fig. 2c shows the smoking history of patients in terms of pack years, where one pack year is equivalent to smoking 20 cigarettes per day for one year. Patients smoked at least 30 pack years, with an average of 56 pack years across the entire dataset.

Additionally, the dataset's geographical and racial diversity should be considered. The NLST was conducted in specific locations across the United States, and this geographical limitation can introduce biases related to environmental factors and healthcare access. Moreover, the racial and ethnic composition of the dataset may not adequately represent the diversity of the wider population. This could affect the generalizability of the findings, as certain racial and ethnic groups might have different risk factors or disease prevalence rates.

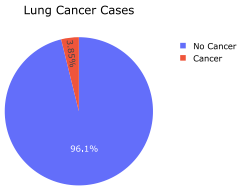
The metadata for each patient in the NLST contains a large amount of information about their ethnicity, body parameters, education, smoking history, and medical history. For our model, we selected a subset of features that are described in Table 1.



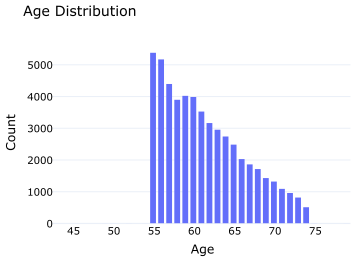
(a) Gender Distribution for the NLST Dataset



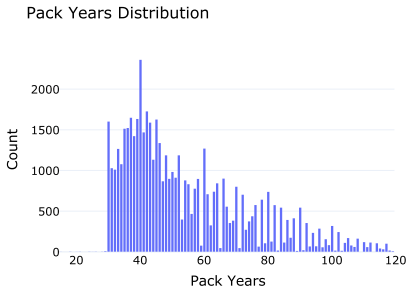
(b) Race distribution for the NLST dataset



(c) Percentage of lung cancer cases in the dataset



(d) Age Distribution for the NLST Dataset



(e) Package Year Distribution for the NLST Dataset

Fig. 2: Relevant features of the NLST dataset

3 Methods

3.1 Pre-Processing

The CT images in the dataset are initially in the dicom format. We first converted them into nifti since nifti enables faster and easier preprocessing. For each patient assigned to the CT arm, for each time point, there are multiple CT images with different slice thicknesses and reconstruction filters. We determined the best CT image for each patient and time point based on the recommendation of a medical expert. The criteria for selecting the best reconstruction filter were as follows: filters with higher alphabetical ranking or numerical value were preferred; if filters with alphanumerics were not available, the bone filter was prioritized, followed by any other available filter. If more than one CT scan remained after applying this technique, we chose the CT scan with the smallest slice thickness as the best CT. From this point onwards we only worked with these best CT images.

From these CT images, we compute the bounding box for the candidate nodules using nnDetection [15], a self-configuring framework for 3D (volumetric) medical object detection. For nnDetection, we utilized a model pre-trained on the Luna16 dataset [16] to identify the candidate nodules.

Then, we used the morphological method by Liao et al. [8] to compute the lung mask. Based on the candidate nodules computed and the lung mask generated, we selected the top five confident nodules inside the lungs and saved it for the nodule-level analysis. The pre-processing pipeline can be seen in Fig 3a.

3.2 Training

Given the significant skew in the dataset towards negative cases and its non-uniform distribution across gender and age groups, we initially curated a cohort comprising 3186 patients. This cohort was stratified into three categories: cancer-positive, no cancer-positive scans, and no cancer-negative scans, thereby ensuring a uniform distribution across both age and gender demographics. We employed a 70/20/10 train/validation/test split while maintaining consistent distributions of cancer-positive, no cancer-positive scans, and no cancer-negative scans of all three categories across all sets. Since we are only working with the timestep T0, our objective is to predict, based on the CT scan at timestep T0, whether the patient has lung cancer at that time or not, with the ground truth being the presence or absence of cancer at that timepoint.

Predictions with only the tabular data We used Xgboost [17] with *scale_pos_weight* > 10, and trained it using the feature set listed in Table 1.

Prediction with only the nodules We further process the top five most confident nodules identified in the pre-processing phase. If fewer than 5 nodules are identified in the pre-processing step, we simply repeat the least-confident nodules to reach 5. Standardizing their shapes is the first step in this process. To achieve this, we determine the 95^{th} percentile shape across each dimension, which we empirically found to be $41 \times 41 \times 16$. Subsequently, we reshape each nodule to conform to this determined shape. If a nodule's dimensions are smaller than the specified anchor shape in any dimension, we pad it with -1024. Conversely, we employ a center cropping approach if a nodule exceeds the anchor shape in any dimension. Following reshaping, we constrain pixel values within the lung window range (-1024 HU to 150 HU) and normalize them to a scale between 0 and 1 before feeding them into the encoder.

We experimented with various encoders to extract nodule features, namely an untrained 3D ResNet, a pre-trained 2D ResNet converted to 3D using ACSConv [18], and MedicalNet [19], and a pre-trained 3D ResNet trained on medical images. Initially, the encoder individually processed each nodule to compute the features, which were then concatenated into a unified vector and passed through a fully connected layer to determine malignancy probability. Given the absence of ground truth regarding the specific cancerous nodule in cases where cancer is present, a Multi-Instance Learning (MIL) problem arises [20]. To address this, various strategies were explored. One approach involved employing max-pooling across feature entries instead of concatenation. Additionally, we also experimented with the Leaky Noisy-OR method proposed in [8] to enable the model to learn to assign cancer to a dummy nodule when the cancerous nodule is not among the top five selected in the pre-processing step.

Multimodal Training In multimodal training, features were computed separately for nodules, whole lung images, and tabular data. Nodule features were derived using the aforementioned pipeline. The whole CT image is windowed similarly as described previously, and all images are reshaped to a size of $400 \times 400 \times 120$. Then, it is encoded with a Resnet-3D obtained with ACSConv to extract the global features. For the patient's metadata, categorical and numerical features were preprocessed separately. Categorical features were converted into one-hot vectors, and numerical features were normalized before being passed to a 2-layer perceptron to obtain tabular embeddings. Lung feature, nodules features, and tabular embeddings were concatenated and then passed to another 2-layer perceptron to predict the malignant probability.

3.3 Training Implementation

We employed Adam optimizer [21] ($\alpha = 1e-3, \beta_1 = 0.9, \beta_2 = 0.999$) for optimization. Training spanned across 30 epochs. To dynamically adjust the learning rate, we decreased it by a factor of 10 upon reaching a plateau. Moreover, we implemented early stopping with a patience of 5 to mitigate overfitting. For cases not encompassing the entire lung image, we employed a batch size of 8, while for other cases, the batch size was set to 4. The network was trained end-to-end with the weighted BCELoss:

$$L_{BCE}(\mathbf{y}, \mathbf{p}) = -\sum_{n=1}^{N} \beta \cdot \mathbf{y}_n \cdot \log(\mathbf{p}_n) - (1-\beta) \cdot (1-\mathbf{y}_n) \cdot \log(1-\mathbf{p}_n)$$

where $\beta \in (0.5, 1)$ is a hyperparameter. Empirically, $\beta = 0.7$ worked the best for us.

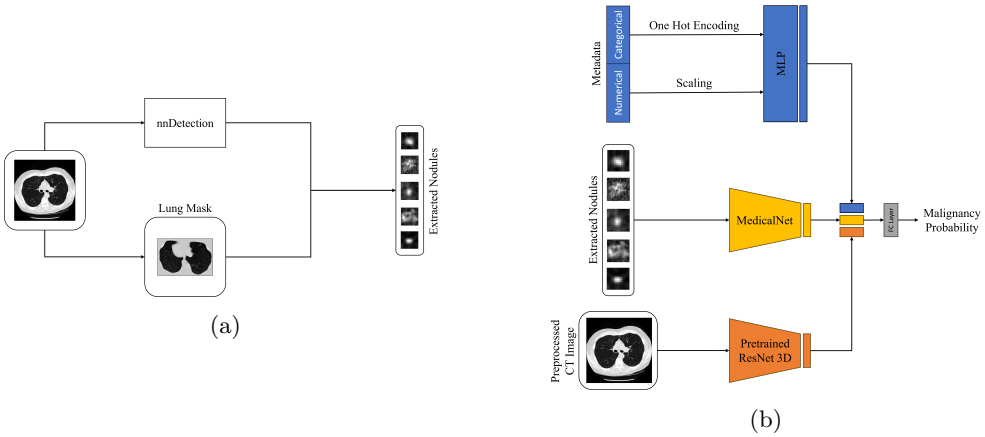


Fig. 3: **(a)** The outline of our pre-processing pipeline involves the identification of candidate nodules, followed by the selection of the top five most confident nodules situated within the lung region. **(b)** For our Multimodal Approach, we integrate this nodule level data with lung level and clinical metadata. Feature extraction from nodules and lung images is accomplished using MedicalNet and Pretrained ResNet 3D, respectively, while neural networks are employed to embed the metadata.

4 Results

We evaluated our various models using the test set outlined in Section 3.2. Table 2 presents a summary of the results. Our multimodal instance (Acc = 0.82, F1 = 0.72) demonstrates superior performance compared to unimodal instances. Moreover, the approach integrating both global lung features and localized nodule characteristics along with metadata outperformed the approach utilizing only nodule features and metadata. This indicates the advantage of incorporating lung information alongside localized nodule details for enhancing prediction. Furthermore, we observed that models employing encoders pre-trained on medical images outperform those using encoders pre-trained on general images or those without any pre-training.

Table 2: Lung Cancer Prediction results of our experimental methods on the test set compared to the state-of-the-art DeepCAD [12]. **Bold** indicates our top-performing method. Performance metrics for DeepCAD are reported as described in [12].

Approaches	**Accuracy**	**Precision**	**Recall**	**F1**
Metadata	0.61	0.59	0.65	0.61
Nodules (ACSConv [18])	0.71	0.54	0.60	0.56
Nodules (Maxpool)	0.73	0.57	0.62	0.59
Nodules (MedicalNet [19])	0.78	0.63	0.68	0.65
Nodules (Leaky Noisy-OR [8])	0.75	0.40	0.75	0.52
Nodules + Metadata	0.78	0.65	0.76	0.70
Multimodal	**0.82**	**0.69**	**0.79**	**0.72**
DeepCAD [12]	0.85	0.84	0.87	0.85

5 Discussion and Conclusion

This work proposes a pipeline to predict lung cancer from CT scans. We show that the combination of different levels of features including clinical metadata and imaging data at the lung and nodule level provides a good estimation of malignancy surpassing the predictive capability of models utilizing these data sources independently. Furthermore, our analysis reveals superior performance when employing a feature extractor pretrained specifically on lung CTs compared to one pretrained on alternative domains.

Leveraging longitudinal data from the NLST dataset could potentially enhance results. However, this aspect is not explored within the confines of this work and is left as a future prospect.

References

[1] K. C. Thandra *et al.*, "Epidemiology of lung cancer," *Contemporary Oncology/Współczesna Onkologia*, vol. 25, no. 1, pp. 45–52, 2021.

[2] N. L. S. T. R. Team, "The national lung screening trial: Overview and study design," *Radiology*, vol. 258, no. 1, pp. 243–253, 2011.

[3] D. Ardila *et al.*, "End-to-end lung cancer screening with three-dimensional deep learning on low-dose chest computed tomography," *Nature medicine*, vol. 25, no. 6, pp. 954–961, 2019.

[4] P. Huang *et al.*, "Prediction of lung cancer risk at follow-up screening with low-dose ct: A training and validation study of a deep learning method," *The Lancet Digital Health*, vol. 1, no. 7, e353–e362, 2019.

[5] Y. Huang *et al.*, "What makes multi-modal learning better than single (provably)," in *Advances in Neural Information Processing Systems*, A. Beygelzimer *et al.*, Eds., 2021. [Online]. Available: https://openreview.net/forum?id=UlSjqPEkI1V.

[6] Y. Li and Y. Fan, "Deepseed: 3d squeeze-and-excitation encoder-decoder convolutional neural networks for pulmonary nodule detection," in *2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI)*, IEEE, 2020, pp. 1866–1869.

[7] J. Ding *et al.*, "Accurate pulmonary nodule detection in computed tomography images using deep convolutional neural networks," in *Medical Image Computing and Computer Assisted Intervention- MICCAI 2017: 20th International Conference, Quebec City, QC, Canada, September 11-13, 2017, Proceedings, Part III 20*, Springer, 2017, pp. 559–567.

[8] F. Liao *et al.*, "Evaluate the malignancy of pulmonary nodules using the 3-d deep leaky noisy-or network," *IEEE transactions on neural networks and learning systems*, vol. 30, no. 11, pp. 3484–3495, 2019.

[9] J. L. Causey *et al.*, "Lung cancer screening with low-dose ct scans using a deep learning approach," *arXiv preprint arXiv:1906.00240*, 2019.

[10] J. Wang *et al.*, "Lung cancer detection using co-learning from chest ct images and clinical demographics," in *Medical imaging 2019: Image processing*, SPIE, vol. 10949, 2019, pp. 365–371.

[11] H. Jiang *et al.*, "Attentive and ensemble 3d dual path networks for pulmonary nodules classification," *Neurocomputing*, vol. 398, pp. 422–430, 2020.

[12] S. Aslani *et al.*, "Enhancing cancer prediction in challenging screen-detected incident lung nodules using time-series deep learning," 2022.

[13] R. Gao *et al.*, "Cancer risk estimation combining lung screening ct with clinical data elements," *Radiology: Artificial Intelligence*, vol. 3, no. 6, e210032, 2021.

[14] T. Z. Li *et al.*, "Longitudinal multimodal transformer integrating imaging and latent clinical signatures from routine ehrs for pulmonary nodule classification," *arXiv preprint arXiv:2304.02836*, 2023.

[15] M. Baumgartner *et al.*, "Nndetection: A self-configuring method for medical object detection," in *Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part V 24*, Springer, 2021, pp. 530–539.

[16] A. A. A. Setio *et al.*, "Validation, comparison, and combination of algorithms for automatic detection of pulmonary nodules in computed tomography images: The luna16 challenge," *Medical image analysis*, vol. 42, pp. 1–13, 2017.

[17] T. Chen and C. Guestrin, "Xgboost: A scalable tree boosting system," in *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, 2016, pp. 785–794.

[18] J. Yang *et al.*, "Reinventing 2d convolutions for 3d images," *IEEE Journal of Biomedical and Health Informatics*, vol. 25, no. 8, pp. 3009–3018, 2021.

[19] S. Chen *et al.*, "Med3d: Transfer learning for 3d medical image analysis," *arXiv preprint arXiv:1904.00625*, 2019.

[20] M. Ilse *et al.*, "Attention-based deep multiple instance learning," in *International conference on machine learning*, PMLR, 2018, pp. 2127–2136.

[21] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.