Single and Multi-modality Approaches for Lung Cancer Classification

Joana Vale Sousa[12]
joana.v.sousa@inesctec.pt
Pedro Matos[2]
up201605912@edu.fe.up.pt
Francisco Silva[13]
francisco.c.silva@inesctec
Pedro Freitas[12]
pedro.g.freitas@inesctec
Hélder P. Oliveira[13]
helder.f.oliveira@inesctec
Tania Pereira[1]
tania.pereira@inesctec

[1] INESC TEC-Institute for Systems and Computer Engineering, Technology and Science
4200–465 Porto, Portugal

[2] FEUP–Faculty of Engineering
University of Porto,
4200–465, Porto Portugal

[3] FCUP–Faculty of Science
University of Porto,
4169–007, Porto Portugal

Abstract

When making lung cancer diagnosis, physicians usually take into account data from different modalities, and artificial-intelligence based methods could follow the same approach, in order to allow a more comprehensive analysis. Nonetheless, a great proportion of related works focus solely on imaging data. This work intended to investigate the potential of different data sources for lung cancer classification. A ResNet18 network was trained to classify 3D CT nodule regions of interest (ROI), and a random forest algorithm was used to classify clinical data. Intermediate and late fusion methodologies were also developed, that combined the information from clinical data and 3D CT nodule ROIs. The best result, an AUC of 0.8021, was achieved by an intermediate fusion model – a fully connected layer that receives deep imaging features, obtained from a ResNet18 inference model, and clinical data. Lung cancer is a complex disease, and this study shows that the combination of distinct modalities may have the potential to allow a comprehensive analysis of the pathology.

1 Introduction

Lung cancer is the leader in cancer-related deaths, mainly due to a late diagnosis that results in lower 5-year survival rates. An early detection, on the other hand, may increase these rates, and it can be achieved through screening. In the clinical practice, clinicians usually take into consideration data from different modalities, such as CT scans and clinical data, when predicting lung cancer. Artificial Intelligence (AI) methods can assist clinicians in this task, reducing the false positives and negatives and enabling a more accurate diagnosis. However, most AI methodologies focus solely on the imaging data to extract lung cancer related information, which may restrain the learning of the models [1]. Works such as [5, 7, 8] limit their analysis to CT scans, using the Lung Image Database Consortium dataset for the development of their algorithms. Over recent years, methods that combine the information from different data modalities have emerged, that often surpass the performance of approaches that rely on a single modality [6]. Lung cancer is a complex pathology, influenced by several biological factors. In that regard, multimodality may enable the possibility of developing learning models capable of delivering a proper response to that need. Therefore, this work aimed at studying and comparing lung cancer classification models that depend on a single data modality with models that mimic the clinical context by combining information from various modalities. Furthermore, the National Lung Screening Trial-(NLST) [2] dataset was used since it enables the fusion of different modalities and it contains more challenging cases (as confirmed by the results in [3]), which may enable the development of more robust models.

This work is a summary of [4] that aimed at studying lung cancer prediction models, when using a single data modality or multiple modalities.

2 Materials and Methods

2.1 National Lung Screening Trial Dataset

The NLST [2] dataset was used in this work and it includes data from different modalities, namely CT images and clinical data. Firstly, each CT scan was submitted to resampling to set the pixel spacing to 1 *mm* in axes x, y and z. Following that, the images were resized to a dimension of 128 × 128 pixels and submitted to a *min–max* normalization, in which the pixels expressed in Hounsfield Units (HU), were mapped to a range of [0, 1] with -1000 and 400 HU as lower and upper limits. 20 × 50 × 50 bounding boxes with the nodule in their center were manually created, with a total of 1079 3D nodule regions of interest (ROI) obtained, from which 655 were of the class benign and 424 of the class malignant, corresponding to a total of 1005 patients, given that for some of the patients more that one nodule was identified. Regarding the clinical data, a total of 136 features, were selected, under the following tags: demographic, smoking, work history, disease history, personal cancer history, family history, and alcohol.

2.2 Methodology

Single- and multi-modality strategies were implemented for lung cancer classification, as depicted in Figure 1, using CT scans and clinical data. In all experiments, 80% of the data were used for training and the remaining 20% were used for evaluation. 5-fold cross-validation was implemented using the 80% assigned for training. Binary Cross-Entropy was used as loss function and Area Under the ROC Curve (AUC) was used as evaluation metric.

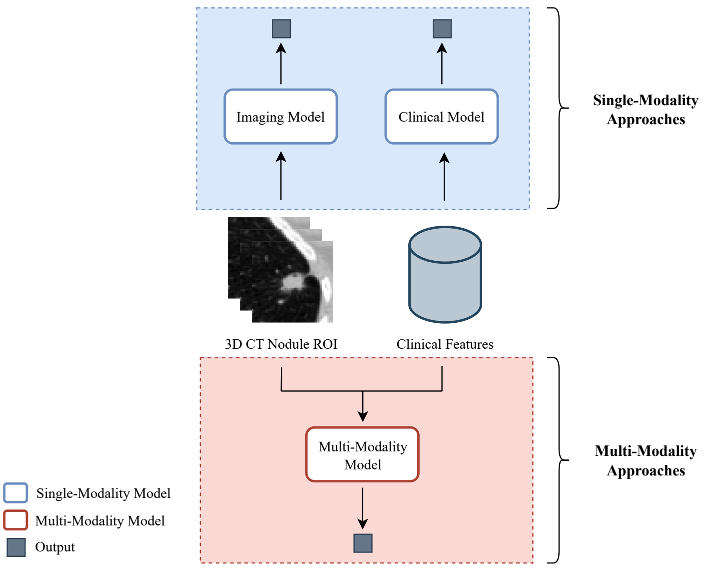

Figure 1: Pipeline for study of the single- and multimodality strategies for lung cancer classification. From [4].

Regarding the single-modality approaches, for the imaging data, a 3D ResNet-18 was implemented, and as for the clinical data, a random forest (RF) was implemented. After the evaluation of the impurity-based feature ranking of the best performance RF model, the range of features was reduced to 42.

With respect to the combination of the two data modalities in the multimodality approaches, two main types of strategies were implemented:

intermediate fusion, in which features from each modality are concatenated to be used as input of a single model; and late fusion, in which the outputs of each modality model are combined to produced the final classification result [6]. The pipeline implemented for the multimodal strategies is represented in Figure 2. For the intermediate fusion, two methods were investigated: half-intermediate fusion (HIF), in which the input corresponds to the combination of clinical data and the malignancy prediction for 3D nodule ROIs given by an inference model (the ResNet18 imaging model that achieved the highest AUC); and full intermediate fusion (FIF), in which the input is the combination of clinical data and 512 deep imaging features of 3D nodule ROIs given by the last layer prior to the classification layer of the same inference model. In both, the concatenated features are fed to one fully connected layer, followed by a sigmoid activation layer that outputs the final probability. In these experiments, two sets of clinical features were studied: one with the original 136, and another with the selected 42, as described above. For the late fusion (LF) approach the malignancy prediction corresponds to the weighted average of the outputs of the imaging and the clinical models. The weight assigned to each output ranged between 0.1 and 0.9.

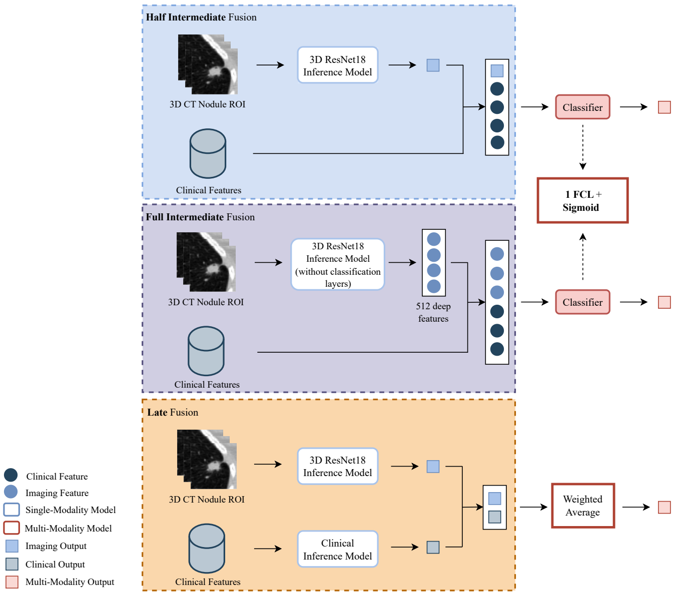

Figure 2: Pipeline of the multi-modalities strategies. From [4].

3 Results and Discussion

Table 1 presents the results obtained for all experiments.

Table 1: Results obtained for the five methodologies implemented. The Full Intermediate Fusion approach obtained the highest performance metric, highlighted in bold.

Approach		#Clinical Features	AUC
Single-Modality	Image Model	-	0.7897
	Clinical Model	136	0.5241
Multi-Modality	HIF	42	0.7934
	FIF	42	**0.8021**
	LF	136	0.7911

From the results one can see that the multimodality approaches present the highest performance metrics, which may suggest that integrating information from different modalities may enable a more comprehensive analysis, specifically when comparing to the clinical model. On the other hand, in comparison to the imaging model, these improvements are minimally significant. In fact, the results of the imaging model showed the importance of the CT volumes for the prediction of lung cancer. Given the poor results of the clinical model, perhaps the clinical features used may not have the relevance needed that would allow the model to correctly learn to distinguish between cancer and non-cancer diagnosis. This is in agreement with what one would expect in a clinical context, in which physicians usually do not consider only the characteristics of the patient, such as medical history and personal information. In a similar way, given that the output of the LF method is the result of the weighted average of the predictions of the single-modality models and considering the poor results of the clinical model, it was expected that it would exhibit the lowest performance metric among all multimodality approaches. Those perceptions are also made evident in both intermediate fusion methodologies, as the network mainly focus on the imaging inputs, generated by the imaging inference model. Furthermore, the structure of the intermediate fusion models is composed of a single fully connected layer, that is equivalent to the last layer of the imaging model, and possibly this configuration may not have been appropriate to capture to its full extent the relationship shared by the features of the CT images and the clinical data, supposing there is one.

4 Conclusion

The goal of this work was the investigation of the combination of more than one type of information for the prediction of lung cancer. The results obtained demonstrated the importance of the imaging data, essential for lung cancer diagnosis, whereas the clinical features used showed a poor predictive capability when used alone. The results obtained by the multimodality approaches, on the other hand, showed the potential of combining different data modalities. The future investigation could include the combination of other strategies and architectures, namely the use of deep learning methodologies to extract features of the clinical data, in the hope of analysing its full potential.

5 Acknowledgments

This work is financed by National Funds through the FCT - Fundação para a Ciência e a Tecnologia, I.P. (Portuguese Foundation for Science and Technology) within the project LUCCA, with reference 2022.03488.PTDC, and by National Funds through the Portuguese funding agency, FCT-Foundation for Science and Technology Portugal, two PhD Grants Numbers 2023.03408.BD and 2021.05767.BD.

References

[1] Francisco Silva *et al.* Towards machine learning-aided lung cancer clinical routines: Approaches and open challenges. *Journal of Personalized Medicine*, 12(3), 2022. ISSN 2075-4426. doi: 10.3390/jpm12030480. URL `https://www.mdpi.com/2075-4426/12/3/480`.

[2] National Lung Screening Trial Team *et al.* The National Lung Screening Trial: Overview and Study Design. *Radiology*, 258(1): 243–253, Jan 2011. doi: http://doi.org/10.1148/radiol.10091808.

[3] J. M. Uthoff *et al.* Computed Tomography Features of Lung Structure Have Utility for Differentiating Malignant and Benign Pulmonary Nodules. *Chronic Obstructive Pulmonary Diseases: Journal of the COPD Foundation*, 9(2):154–164, 2022. doi: http://doi.org/10.15326/jcopdf.2021.0271.

[4] Joana Vale Sousa *et al.* Single Modality vs. Multimodality: What Works Best for Lung Cancer Screening? *Sensors*, 23(12), 2023. ISSN 1424-8220. doi: 10.3390/s23125597. URL `https://www.mdpi.com/1424-8220/23/12/5597`.

[5] Juan Lyu *et al.* Multi-Level Cross Residual Network for Lung Nodule Classification. *Sensors*, 20(10), 2020. ISSN 1424-8220. doi: 10.3390/s20102837. URL `https://www.mdpi.com/1424-8220/20/10/2837`.

[6] Khaled Bayoudh *et al.* A survey on deep multimodal learning for computer vision: advances, trends, applications, and datasets. *The Visual Computer*, 38:2939–2970, 2022. doi: https://doi.org/10.1007/s00371-021-02166-7.

[7] Liu Dongxu *et al.* Res-trans networks for lung nodule classification. *International Journal of Computer Assisted Radiology and Surgery*, 17:1059–1068, 2022. doi: https://doi.org/10.1007/s11548-022-02576-5.

[8] Mundher Al-Shabi *et al.* 3D axial-attention for lung nodule classification. *International Journal of Computer Assisted Radiology and Surgery*, 16:1319–1324, 2021. doi: https://doi.org/10.1007/s11548-021-02415-z.