Classification of lung nodules on CT via pseudo-colour images and deep features from pre-trained convolutional networks

Francesco Bianconi[1] Mario L. Fravolini[1] Elena Caltana[1]
Muhammad U. Khan[1,2] Barbara Palumbo[3]

Computational Color Imaging Workshop (CCIW 2024)
Milan, 25–27 Sep. 2024

[1]Department of Engineering, Università degli Studi di Perugia, Italy
[2]CNIT, Perugia Research Unit, Italy
[3]Department of Medicine and Surgery, Università degli Studi di Perugia, Italy

Outline

Outline

Background and motivation

Lung cancer: facts and figures

- **Leading cause** of cancer-related death in the US (about 1 in 5 of all cancers)
- **Chance** of developing lung cancer in a lifetime:
  - 1/16 (men)
  - 1/17 (women)
- Five-year **survival** rates:
  - 65% (localised)
  - 37% (regional)
  - 9% (distant)

Source: American Cancer Society, accessed 16 Jun. 2024

Motivation (1/3)

- Survival depends a great deal on the **stage** the disease is first detected
- **Early detection and diagnosis** are critical for a better outcome
  - At an initial stage lung cancer usually presents as a small, rounded opacity, often detected on CT (**lung nodule**)
  - However, only a small fraction of lung nodules represent malignancies
- The **clinical management** of patients with suspicious lung nodules is **intrinsically difficult**

Motivation (2/3)

Computerised analysis (*radiomics*) can improve the diagnosis of indeterminate lung nodules detected on CT

- Based on the extraction of **quantitative features**
- Takes advantage of picture data **invisible to the naked eye**
- Leverages on **AI** methods and datasets of pre-classified data

Current approaches:

- **Conventional radiomics** (feature engineering & hand-crafted features)
- **Deep Learning radiomics** (CNN)

Motivation (3/3)

- Deep Learning radiomics is generally superior in accuracy, however:
  - We need **large datasets** to train the nets
  - We may easily incur in **overfitting** and **lack of generalisation**
- Alternatively, we can use and **pre-trained** networks off-the shelf, but:
  - The majority of pre-trained CNN accept **planar colour images** as input
  - CT data are **grey-scale and volumetric**

Objectives

- To investigate **pseudo-colouring** schemes to
  - Transform 3D gray-scale CT data to 2D pseudo-colour images
  - Extract feature from the pseudo-colour images by pre-trained CNN
- To evaluate the effectiveness of this strategy to discriminate **benign vs. malignant** lung nodules detected on CT

Materials

Datasets

- Two independent datasets of **solid** and **part solid** lung nodules
- Sourced form public, open access collections (**LIDC-IDRI**, **LUNGx**)
- Common inclusion criterion:
  - Nodule size[1] between 10.0 mm and 50.0 mm

[1] Defined as the length of the largest side of the axis-aligned bounding box

LIDC-IDRI

- $n = 633$ (261 benign, 372 malignant)
- Manually annotated lesion delineation and malignancy score by at least one radiologist

Inclusion criteria:

- **Annotations** by at least **two** radiologists
- **Texture score** (solidity) ≥ 3.0 ($1.0 =$ GGO, $5.0 =$ solid)
- **Malignancy score** either ≤ 2.5 ($\rightarrow$ benign) or ≥ 3.5 ($\rightarrow$ malignant)

Lesion delineation (ROI) based on the 50% consensus rule

LUNGx

- $n = 69$ (32 benign, 37 malignant)
- Manual lesion delineation by consensus (panel of two experts)

Inclusion criteria:

- Solid or sub-solid nodules determined by visual assessment (panel of two experts)

Methods

The procedure

Involves the following steps:

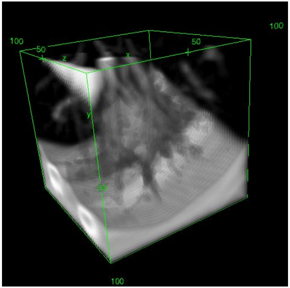
1. **Image preparation** (pre-processing)
2. Generation of the **pseudo-colour images**
3. **Feature** extraction via pre-trained CNNs

A further optional step was also considered:

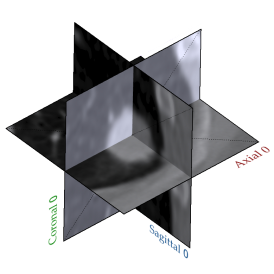
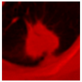
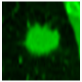
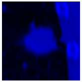
- **Background removal** (contextual information)

Image preparation

- Isotropic **spatial resampling** (0.6 mm × 0.6 mm × 0.6 mm)
- Extraction of a **cubic tensor** (91 × 91 × 91) around the centroid of the nodule
- Signal **windowing** ([-1000 HU, 500 HU])
- Signal **quantisation** (256 levels)

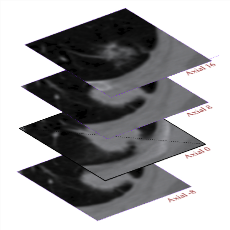
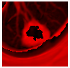
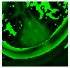
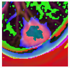


Pseudo-colouring by three orthogonal slices (PCL)

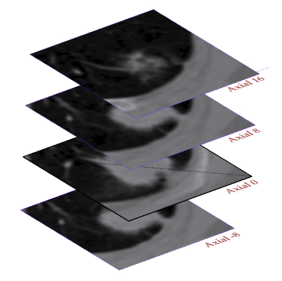
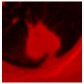
CT data		R, G, B channels		Pseudo-colour image
	$\rightarrow$	Axial (0) 		
	$\rightarrow$	Sagittal (0) 	$\Rightarrow$	
	$\rightarrow$	Coronal (0) 		

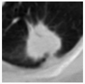
Note: '0' indicates the central slice – i.e., through the centroid of the ROI

Pseudo-colouring by principal components (PCA)

CT data		R, G, B channels		Pseudo-colour image		
	$\Rightarrow$	$\leftarrow$ Stack + PCA + L_1 norm.	$\rightarrow$	1[st] p.c. 		
			$\rightarrow$	2[nd] p.c. 	$\Rightarrow$	
			$\rightarrow$	3[rd] p.c. 		

Pseudo-colouring by central axial slice (GS)

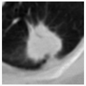
<table>
<tr><th>CT data</th><th></th><th>R, G, B
channels</th><th></th><th>Pseudo-colour
image</th></tr>
<tr><td></td><td>→
→
→</td><td>Axial (0)

Axial (0)

Axial (0)
</td><td>⇒</td><td></td></tr>
</table>

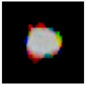
Background removal

Two options:

- Both the nodule and the contextual conformation is retained:

- Contextual information is blanked (background removal):

Feature extraction via pre-trained CNNs

- Three models pre-trained on `IMAGENET` with best available weights as provided by PyTorch's `DEFAULT` option:
  - ConvNeXT
  - ResNet50
  - Swin V2
- Feature extracted from the last `avgpool` layer and L_1 normalised
- Networks operated in frozen (`eval`) mode

Conventional radiomics features

A total of 109 IBSI-compliant features:

- 12 morphological, 19 intensity-based, 23 histogram-based, 23 from grey-level co-occurrence matrices (GLCM), 11 from grey-level run-length matrices (GLRLM), 5 from neighbourhood grey-tone difference matrices (NGTDM) ad 16 from grey-level size zone matrices (GLSZM)

Calculation based on **LIFEx v. 7.4.0**

Pre-processing (spatial resampling, signal windowing and quantisation) same as for the CNN-based features

Feature normalisation and classification

Feature normalisation methods:

- None
- Min-max
- Z-score

Classifiers:

- 1-NN
- Gaussian NB
- Linear classifier
- Logistic regression

Experiments

Experimental conditions

Four experimental conditions

- **Internal validation** (on each dataset by 4-fold):
  1. LIDC-IDRI
  2. LUNGx
- **Cross validation**:
  3. LIDC-IDRI (train), LUNGx (test)
  4. LUNGx (train), LIDC-IDRI (test)

Factorial plan

For each experimental condition we carried out a full-factorial plan with the following factors and levels:

Factor	Levels
Pseudo-colour method*	GS, PCA, PCL
Background removal	Yes, No
Feature extraction	conventional, ConvNeXT, ResNet50, Swin V2
Feature normalisation	None, Min-max, Z-score
Classifier	1-NN, Gaussian NB, Linear classifier, Logistic regression

*Does not apply to conventional radiomics features.

Results and discussion

Results (1/4)

LIDC-IDRI (internal validation):

- PCA 84.7% (ConvNeXT + background removal)
- PCL 88.5% (ConvNeXT + background removal)
- GS 59.7% (ResNet50, no background removal)
- Conventional 85.5%

Results (2/4)

LUNGx (internal validation):

- PCA 65.2% (ConvNeXT, no background removal)
- PCL 69.6% (ConvNeXT + background removal)
- GS 66.7% (Swion V2, no background removal)
- Conventional 60.9%

Results (3/4)

LUNGx (train), **LIDC-IDRI** (test):

- PCA 68.9% (Swin V2 + background removal)
- PCL 68.1% (Swin V2 + background removal)
- GS 58.8% (Swin V2/ResNet50)
- Conventional 65.2%

Results (4/4)

LIDC-IDRI (train), **LUNGx** (test):

- PCA 63.8% (ResNet50, no background removal)
- PCL 65.2% (Swin V2/ResNet50 + background removal)
- GS 62.3% (ConvNeXT/ResNet50 + background removal)
- Conventional 63.8%

Discussion

- Features from pre-trained CNNs **outperformed** conventional radiomics features in all the experimental conditions
- Pseudo-colour generation
  - **PCL** was the best option in three experimental conditions
  - **PCA** in the remaining one
- Best accuracy was always achieved with **background removal** on
  - Is contextual information a confounding factor?

Conclusions, limitations and future work

Conclusions

- We have investigated the ability of deep features from **pseudo-colour images** and **pre-trained CNN** to distinguish benign from malignant lung nodules on CT
- The method seems viable (results better than obtained with conventional radiomics features)

Limitations

- Relatively small **sample size** of one of the two datasets (LUNGx, $n = 69$)
- **Retrospective** nature of the study population
- Role of **clinical** features (e.g., gender, age, history) and other **radiological** features (e.g., spiculation, lobulation, nodule's location) not investigated in this study

Future work

- To better understand the role of **contextual information** (background) on nodule classification
- To determine whether **conventional** and **deep features** provide **complementary information** and investigate ways to **combine** them
- To explore other methods for generating pseudo-colour images (Random projections? Topological data analysis?)

Thank you for your attention

Any questions?