Combining Radiomics and Deep Features for Lung Nodule Malignancy Prediction

Anthony Jatobá <aeaj@ic.ufal.br>
Igor Theotônio <icat@ic.ufal.br>
Eduardo Moraes <emmv@ic.ufal.br>
Vinícius Costa <mvsc@ic.ufal.br>

Introduction

Lung Cancer

- Lung cancer is the leading cause of cancer-related death worldwide
  - 1.8 million new lung cancer cases (2012)
  - 18.4% of the total cancer deaths
- Early detection is crucial
  - Survival rates can get up to 70-90% for stage I tumours
  - But only 15-19% for tumours detected in advanced stages

Computed Tomography

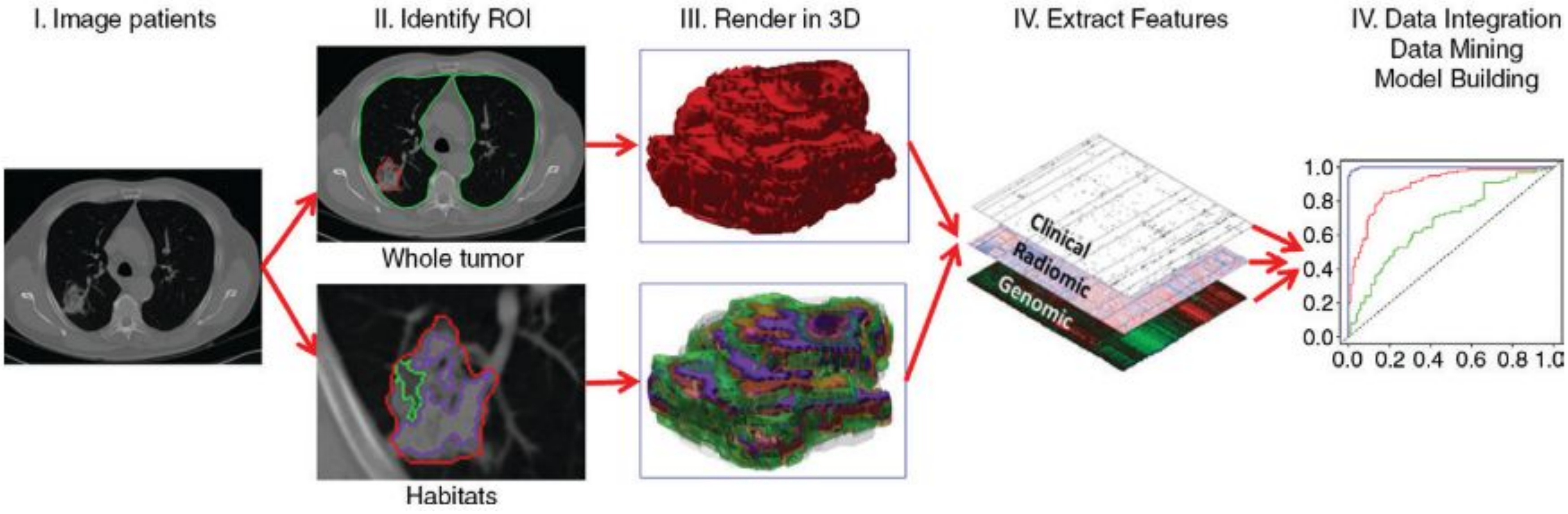
- Computed Tomography (CT) scan is the preferred method for lung cancer screening.
    - CT scans produces a high definition and contrast image that allows for a better tumour characterization;
- Radiologists have to carefully evaluate each image generated by a CT scan

Computer-aided Diagnosis (CADx)

- Computer-aided Diagnosis (CADx) systems emerge as a tool to assist the diagnosis by giving a second opinion to the radiologist;
- Two “classes” of systems:
  - Handcrafted features based approaches
  - Deep learning based approaches

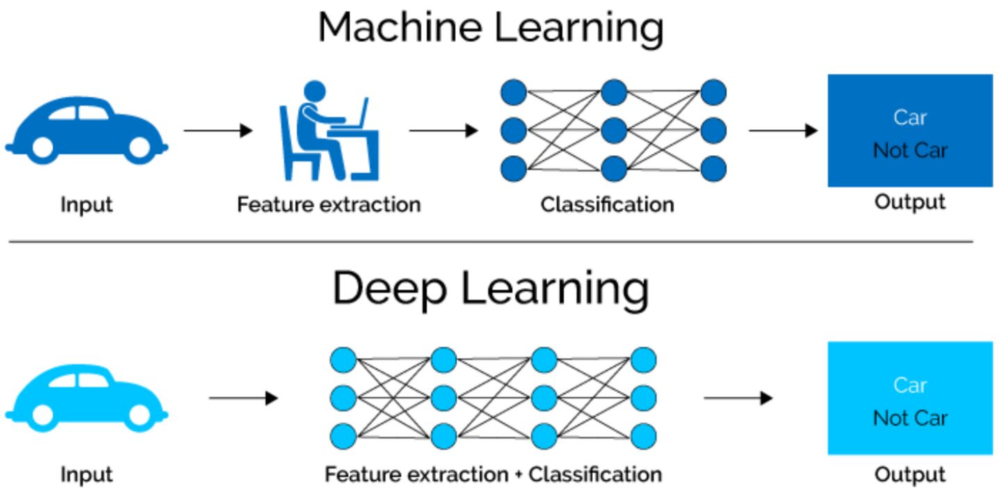
Radiomics

- The extraction of a **large** number of features from digital medical images



Deep Learning

- Deep Learning models can learn abstract representations, as well as perform a given task;



Motivation

Why combine deep features and radiomics?

- It is hard to predefine quantitative features, such as radiomics, that fully reflect the unique characteristics of a lesion
- CNNs require large scale annotated datasets to learn the representative nature of lesions
- Combining both approaches can enhance a predictive models' performance

Why multi-objective optimization?

- Optimization works on the subject optimize either accuracy or AUC
    - Accuracy is not a adequate metric for unbalanced sets of data
    - AUC summarizes performance over unused areas from the ROC curve.
- Sensitivity and specificity can be taken in account
    - Conflicting metrics!
    - Reflects on AUC
- Freedom to select more sensitive/specific models

Objectives

Objectives

- Assess multi-objective optimization performance in selecting radiomics and deep features for lung nodules malignancy prediction;
- Compare multi-objective optimization to a mono-objective optimization in regards to performance;

Experiments

Data

- LIDC-IDRI dataset
  - 897 nodules (616 benign, 281 malignant)
- 121 3D radiomics features from nodules and parenchyme
  - Intensity, shape, texture, margin
- 3D deep features

Models

- Support Vector Machines
  - Default parameters and balanced classes weight;
- 3D CNN: custom architecture found through random search
  - **Convolucional Layer**: 64 units and 3x3x3 kernel
  - **Maxpooling layer:** 2x2x2 kernel
  - **Dense layer:** 128 units and dropout rate of 0.43
  - **Dense Layer:** 64 units and dropout rate of 0.33

Experiments

For *radiomics* and *deep+radiomics* features:

- Run model for all features
- Run model for features selected by GA (AUC)
- Run model for features selected by NSGA (sens/spec)

Each experiment is performed **30 times** for statistical significance**.**

Performance Evaluation

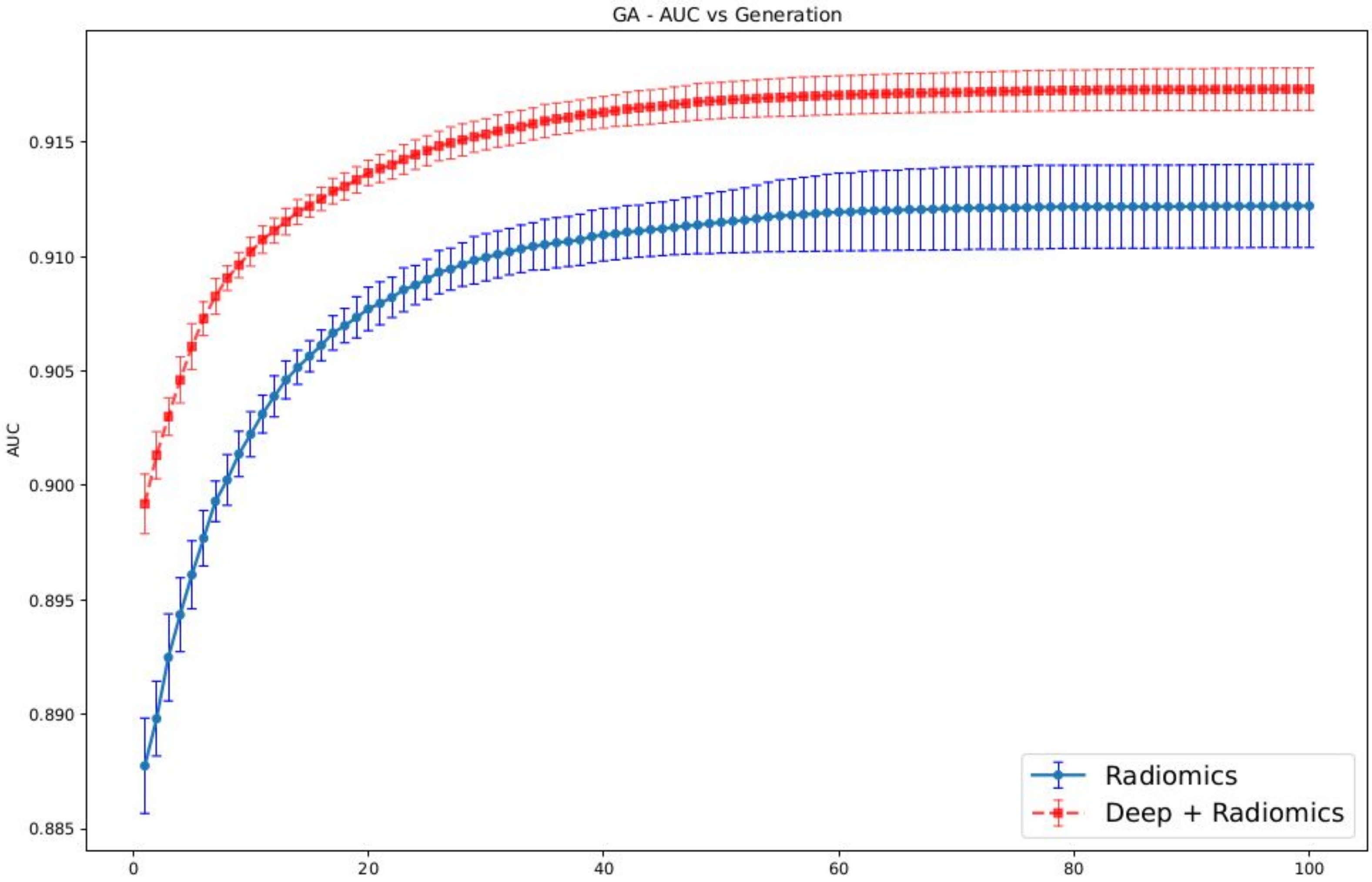
- **Metrics:** accuracy, sensitivity, specificity, AUC
- **ROC** curves
- **Pareto front**
- **Validation:** 10-fold cross validation
- **Statistical test** to ensure results significance

Results and Discussion

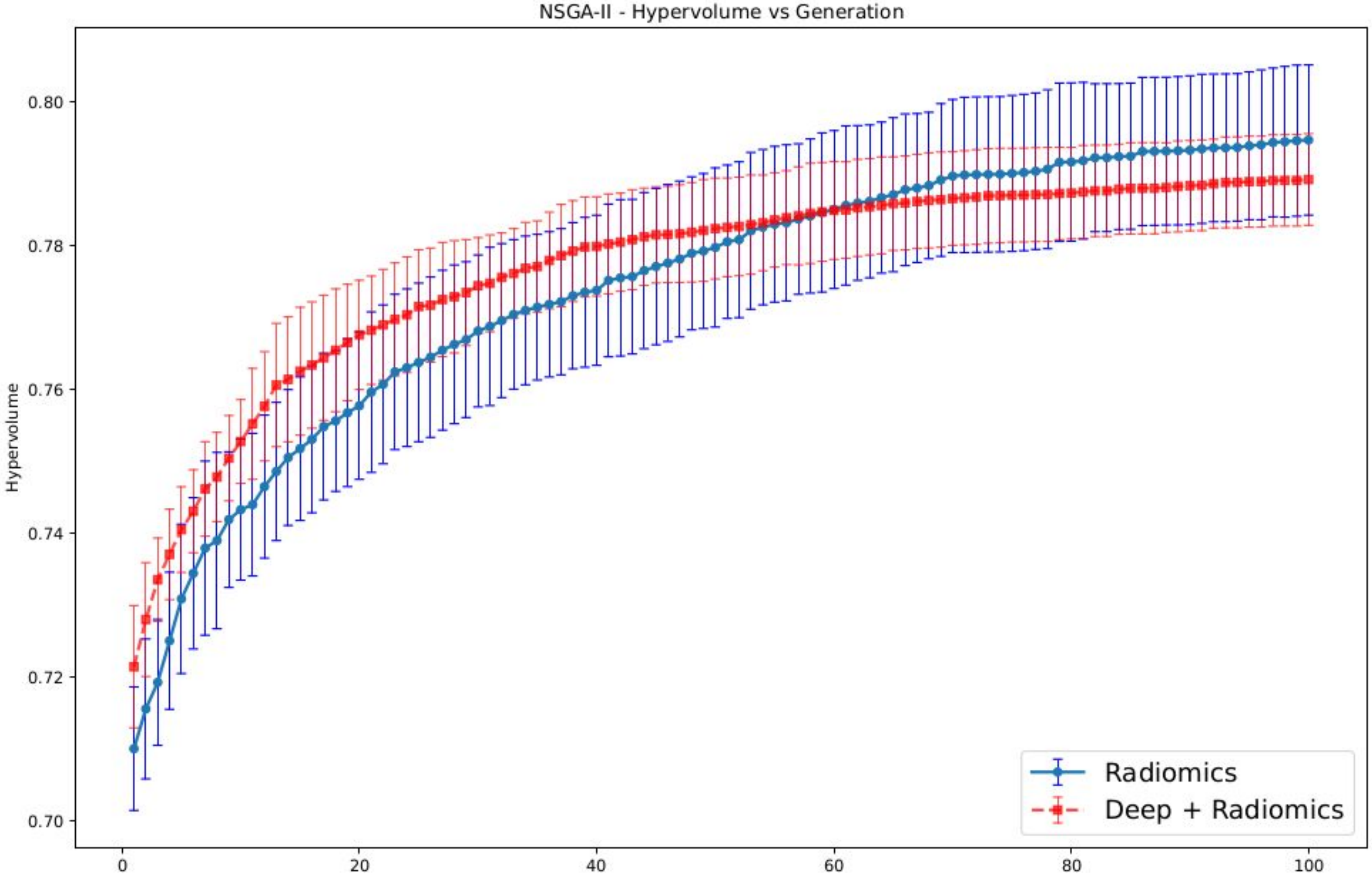
GA optimization

Overview

Base	Seleção	Acurácia	Sens.	Espec.	AUC
radiomics	-	80.5 ± 3.8	79.4 ± 7.3	81.1 ± 5.0	0.889 ± 0.033
	GA	81.8 ± 3.7	85.7 ± 6.2	80.0 ± 4.8	0.907 ± 0.029
	NSGA-II	81.9 ± 3.9	82.9 ± 7.3	81.4 ± 4.9	0.899 ± 0.033
deep + radiomics	-	81.9 ± 4.0	84.8 ± 7.2	80.6 ± 4.8	0.894 ± 0.035
	GA	83.2 ± 3.8	83.5 ± 6.9	83.0 ± 4.7	0.908 ± 0.030
	NSGA-II	83.8 ± 3.7	83.1 ± 7.1	84.1 ± 4.5	0.902 ± 0.031

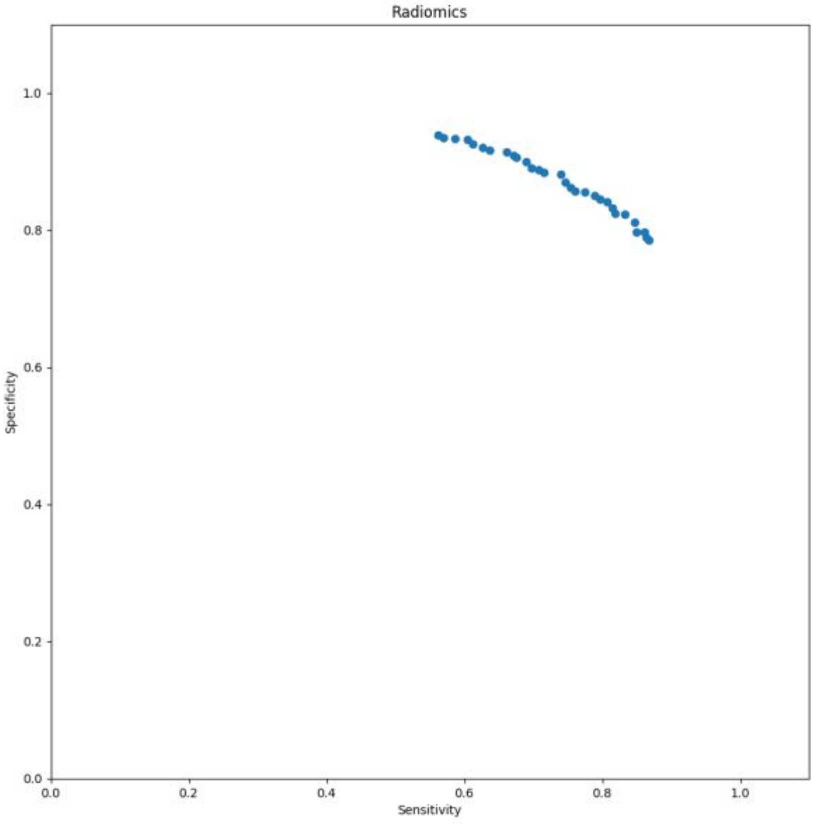

GA - AUC vs Generation
AUC
0.915
0.910
0.905
0.900
0.895
0.890
0.885
0
20
40
60
80
100
Radiomics
Deep + Radiomics

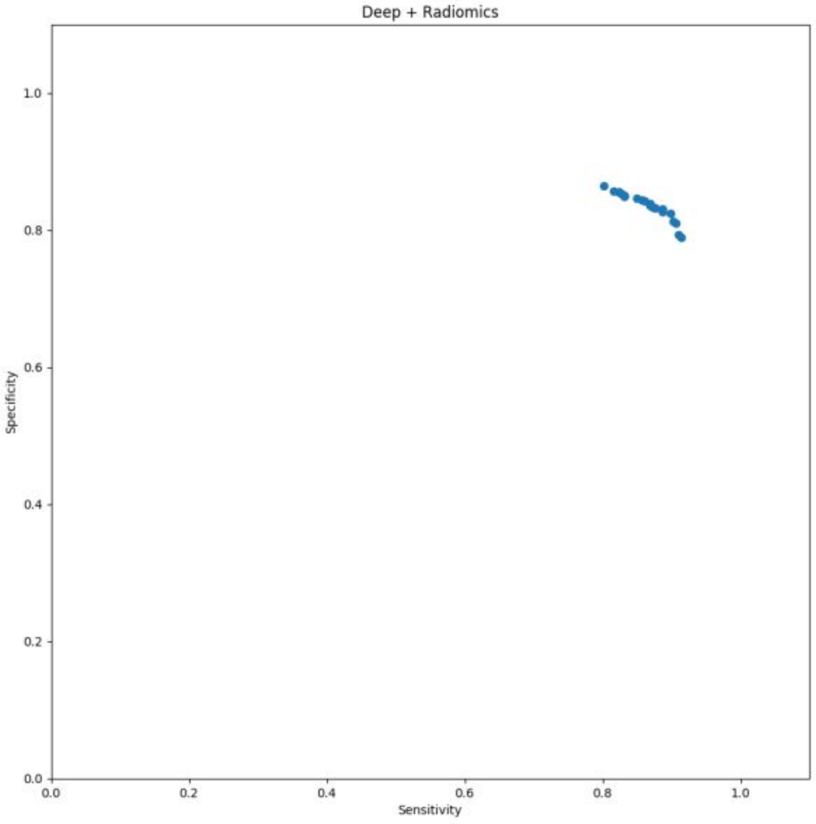
NSGA-II



NSGA-II - Hypervolume vs Generation
Hypervolume
0.70
0.72
0.74
0.76
0.78
0.80
0
20
40
60
80
100
Radiomics
Deep + Radiomics

Pareto Front


Radiomics
Specificity
Sensitivity
0.0
0.2
0.4
0.6
0.8
1.0


Deep + Radiomics
Specificity
Sensitivity
0.0
0.2
0.4
0.6
0.8
1.0

Conclusion

Conclusions

- The inclusion of deep features increased overall performance by a small amount
- GA optimization led to better results on this metric
    - NSGA-II allows the choice for models with different sensitivity/specificity values

Thank you!