

This is the the official repository for the paper:

"Foundation Models for Quantitative Biomarker Discovery in Cancer Imaging "

Suraj Pai, Dennis Bontempi, Vasco Prudente, Ibrahim Hadzic, Mateo Sokač, Tafadzwa L. Chaunzwa, Simon Bernatz, Ahmed Hosny, Raymond H Mak, Nicolai J Birkbak, Hugo JWL Aerts

Quick Start (in under 15 mins!)

We provide two notebooks that can be run in google colab to promote the use of our model by the community,

1. A reproducibility notebook that allows users to recreate our analysis for NSCLC prognostication, from downloading the data to obtaining results matching our manuscript.



2. A bring your own use case notebook that shows you how you can use the foundation model for your data.



You are not limited to using Colab and you can run our model on your data locally in only 3 steps!

Step 1: Install all our dependencies:

```
pip install foundation-cancer-image-biomarker --pre
```

Step 2: Generate a CSV file with the path to your images and seed points (in physical coordinates),

image_path	coordX	coordY	coordZ
/content/data/dummy_0.nii.gz	55.0	119.0	27.0

Step 3: Run this in your code environment,

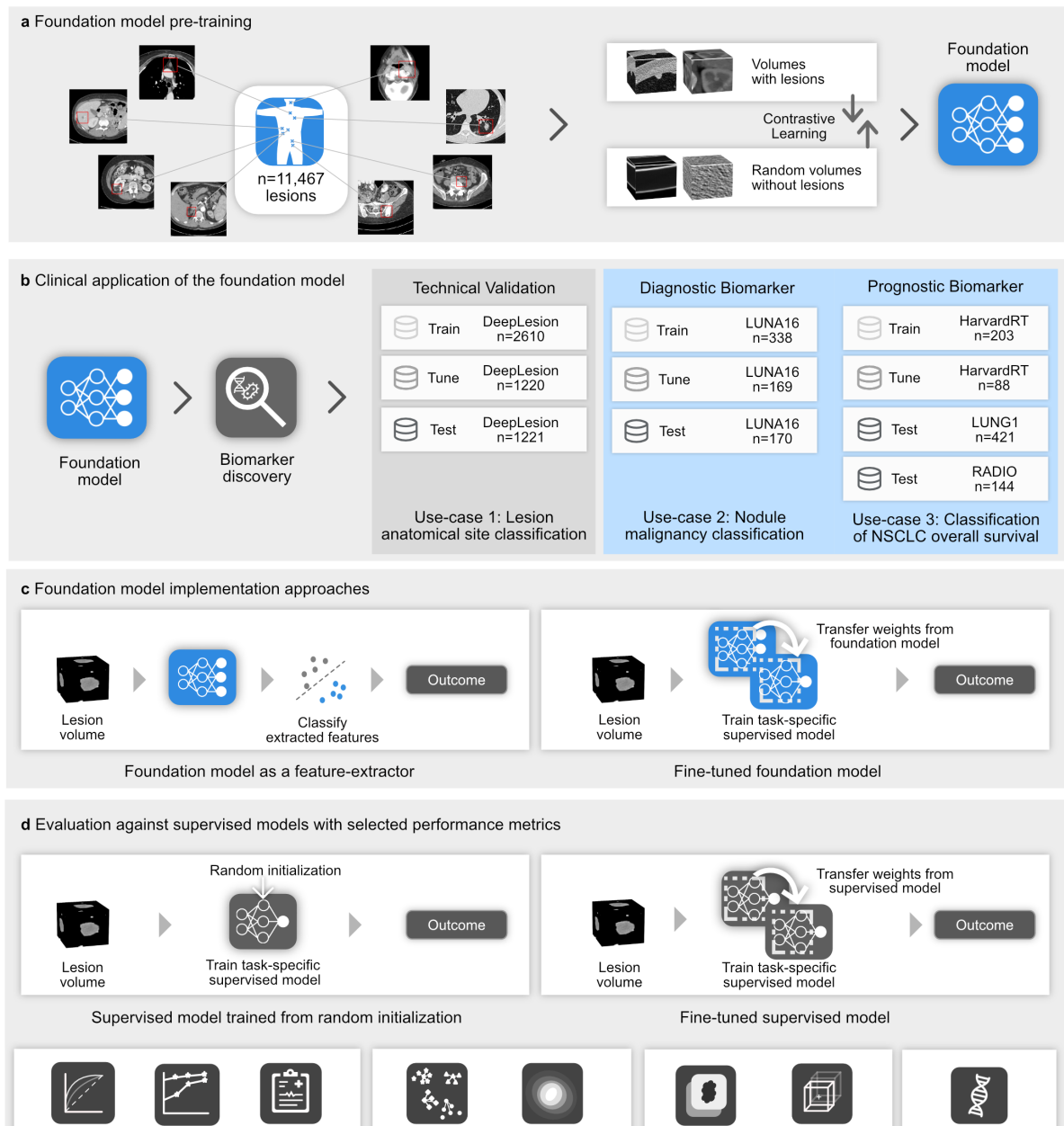
```
from fmcib.run import get_features
```

```
feature_df = get_features("csv_path_here")
```

This will preprocess your data, download our pre-trained model and execute inference on whatever device you have (CPU/GPU) and return a dataframe with the features.

Overview

A foundation model, based on a deep convolutional encoder architecture, was pre-trained by contrasting volumes with and without lesions (see **a.** in Figure). The foundation model was then used to extract biomarkers and evaluated for three classification tasks on diverse datasets. (see **b.**). We adapted foundation model to specific use cases by extracting features or through fine-tuning (see **c.**). Performance of the foundation models was compared against conventional supervised implementations, trained from scratch (left) and fine-tuned from a different task (right) (in **d.**). The comparison was made through several criteria for different use cases, including quantitative performance, stability, and biological analysis. Biological, clinical, and stability analyses are limited to use case 2 due to the availability of associated data.



ROC Curves	Limited Data Evaluation	Clinical Performance	Feature Visualization	Saliency Maps	Test-Retest	Input Perturbations	Genomic Associations
Performance Analysis			Qualitative Analysis		Stability Analysis		Biological Analysis

License

This project is licensed under the terms of the [MIT](#) license.

See [LICENSE](#) for more details.

Acknowledgements

Code development, testing, and documentation: [Suraj Pai](#), [Ibrahim Hadzic](#)

Framework used for building the code: [project-lighter](#)

project-lighter was developed internally within our lab by Ibrahim Hadzic and Suraj Pai.

Disclaimer

The code and data of this repository are provided to promote reproducible research. They are not intended for clinical care or commercial use. The software is provided "as is", without warranty of any kind, express or implied, including but not limited to the warranties of merchantability, fitness for a particular purpose and noninfringement. In no event shall the authors or copyright holders be liable for any claim, damages or other liability, whether in an action of contract, tort or otherwise, arising from, out of or in connection with the software or the use or other dealings in the software.

Code Citation

Will be updated to reflect a Zenodo DOI upon publication

```
@misc{foundation-cancer-image-biomarker,
  author = {AIM-Harvard},
  title = {Official repo for "Foundation Models for Quantitative Biomarker Discovery in Cancer Imaging"},
  year = {2023},
  publisher = {GitHub},
  journal = {GitHub repository},
  howpublished = {\url{https://github.com/AIM-Harvard/foundation-cancer-image-biomarker}}
}
```