

Software Pipelines for AI-Based Radiologic Image Analysis

Briana Malik¹, Kenny H. Cha², Berkman Sahiner², Alexej Gossmann²

¹ University of Pittsburgh Bioengineering, Pittsburgh, PA

² CDRH/OSEL/DIDSR, U.S. Food and Drug Administration, Silver Spring, MD



FDA

Abstract

Research on segmentation, classification, and analysis of medical images in DICOM format is often hindered by the time-consuming and computationally intensive groundwork (preprocessing, exploratory analyses, feature extraction, etc.) that must be laid prior to further steps. To combat this, we are developing a Python package for the analysis of DICOM datasets, consisting of preprocessing and visualization methods, as well as deep learning and radiomics approaches to extract image features. We demonstrate the functionality of our software by preprocessing, visualizing, and analyzing the Lung1 dataset [1], which consists of CT DICOM scans from 422 non-small cell lung cancer (NSCLC) patients. In particular, our package first loads and displays the DICOM image files for interactive viewing, then performs several preprocessing and normalization steps, and finally extracts image features using different approaches as described below. We use deep learning convolutional neural networks (CNNs), pre-trained on natural images from the ImageNet dataset, to extract features. We also extract various radiomics image features such as voxel intensity distribution, 2D/3D size and shape of a region of interest, distance, symmetry and weight statistics, mesh volume analysis and neighborhood pixel analysis. In addition, we train variational autoencoder models as another feature extraction approach. Subsequently, we compare the predictive ability of these diversely extracted features with respect to classifying mortality, when different types of features are used as input to classification models such as logistic regression or XGBoost. To compare classification accuracy, we use the area under the receiver operating characteristic curve, which provides an aggregate measure of performance across all possible classification thresholds. We demonstrate these diverse preprocessing, visualization, feature extraction, and analysis processes to illustrate the use and adaptability of our Python package, which could be used on various other datasets and imaging modalities. In the future, our software package can aid the regulatory science research efforts at the FDA by providing a ready-to-use solution to preprocess, visualize and extract information from medical images.

[1] Aerts, Hugo JW, et al. "Tumour phenotype by noninvasive imaging using a quantitative radiomics approach." *Nature communications* 5.1 (2014): 1-9.

Introduction

- Artificial intelligence (AI) or machine learning (ML) based medical image analysis often requires numerous preprocessing and/or feature extraction steps prior to the application of AI/ML algorithms for tasks such as segmentation, classification, prediction.
- We are developing a general-purpose Python package, to be applicable across many datasets or projects, to streamline and simplify the preprocessing and feature extraction groundwork. Our package includes
 - Preprocessing of raw image data, e.g., intensity normalization, voxel spacing, handling of metadata, extraction of bounding boxes or ROIs.
 - Various visualization methods for the images and ROIs.
 - Various feature extraction methods, ranging from radiomics analysis to deep convolutional neural networks (CNN) and variational autoencoders (VAE).
- To demonstrate and test our Python package, we use it to
 - Perform preprocessing, visualizing, and feature extraction steps on a small dataset of lung cancer CT images [1].
 - Perform the downstream task of patient outcome classification based on the extracted features, and compare the predictive ability of different feature types.

Materials and Methods

Preprocessing methods:

- Load in the DICOM medical imaging data, corresponding masks/bounding boxes and create "data generators" for ML training/testing ease.
- Normalize the voxel intensity, voxel spacing (isotropic), and image dimensions.

CNNs:

- Class of deep neural networks, very popular for (medical) image analysis.
- We use architectures such as ResNet50, VGG16, InceptionV3, which are pretrained on the ImageNet dataset consisting of colored, natural images.

Radiomics analysis / Pyradiomics:

- Python package for the extraction of radiomics analysis features, including mesh volume analysis, weight statistics, symmetry, and neighborhood pixel analysis.
- Input: DICOM images and masks to extract various classes of features.
- Can be combined with other patient data for comprehensive analysis.

VAEs:

- Unsupervised neural networks: machine isn't told what to look for.
- Consists of an encoder NN, a latent space, and decoder NN.
- Via the encoder-decoder interplay, the VAE learns to generate a compressed/optimized latent representation of the input image, which can be used as the "extracted features" for downstream analysis.
- The latent space distribution is controlled by setting a prior distribution, most commonly a centered isotropic multivariate Gaussian $N(0, I)$.

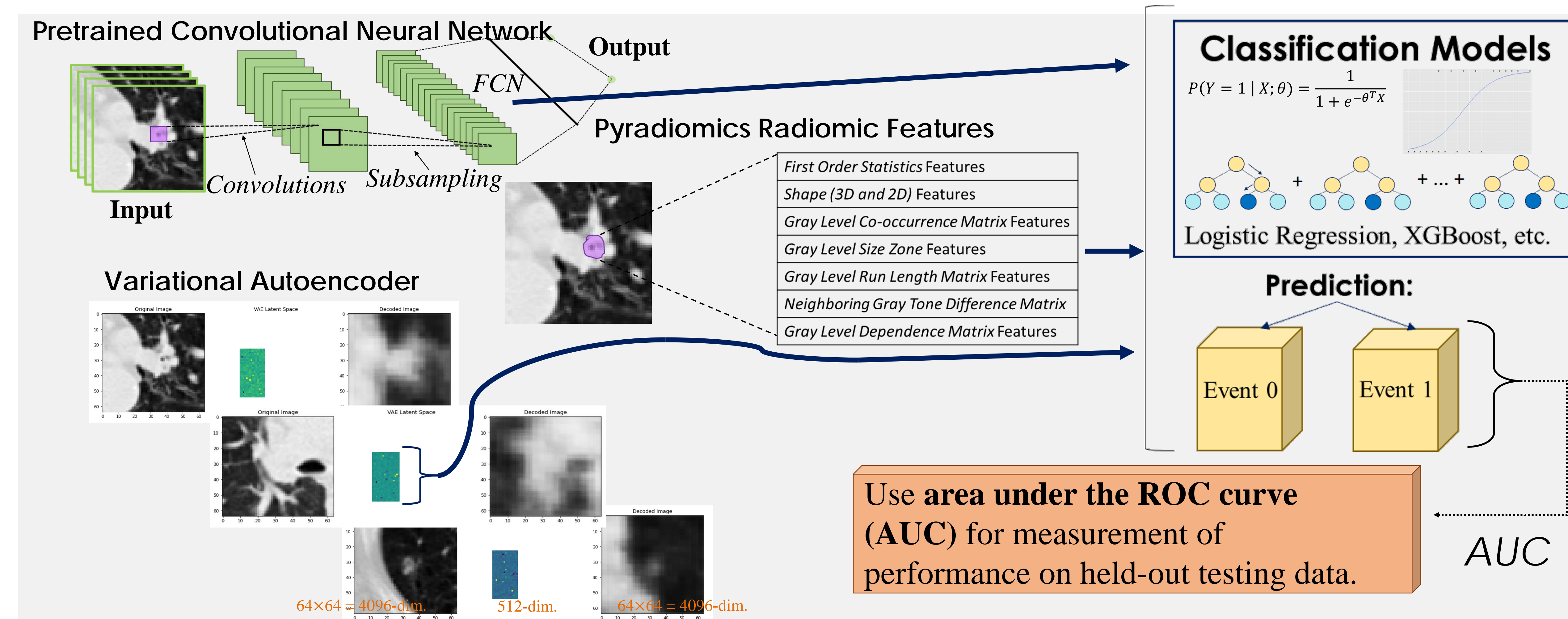
Example Dataset:

- We used LUNG 1 dataset [1] of scans from 422 non-small cell lung cancer (NSCLC) patients as input for feature extraction and classification.
- Training/testing split:** 322 and 100 patients for training and testing.

Extracted Features Classification Analysis:

- Used to predict the mortality outcome for the patient (0 or 1).
- Classification models: **XGBoost** (gradient boosted trees), and **Logistic Regression** (with L2-penalty). Performance is measured using the area under the receiver operating characteristic curve (AUC), to avoid dependence on class prevalence and decision threshold choice.

Figure 1. Our process begins with 3 feature extraction methods – CNNs, Pyradiomics Extraction, and VAEs. CNN feature maps, Pyradiomics feature values, and VAE latent representations are used as features for classification models XGBoost and Logistic Regression to predict patient outcome. Classification performance is measured using AUC, and the predictive capabilities of the different feature types are compared on the Lung1 dataset [1].



Results and Discussion

- After data preprocessing and feature extraction using the highlighted approaches, we fit classification models whereby the different types of features would be used to classify the patient mortality outcomes (0 or 1), and we calculate AUC of the resulting classifiers on the held-out testing data.
- As Table 1 illustrates, Pyradiomics features result in the greatest AUC in this comparison outperforming both DL feature extractors (CNN and VAE), albeit only for XGBoost. This is explained given that XGBoost captures non-linear relationships well, as opposed to logistic regression.
- The comparatively poor performance of classification based on CNN features can be explained given that
 - The CNN feature extractor was trained on images very different from medical images (ImageNet).
 - The XGBoost and logistic regression classifiers were trained on a relatively small dataset (322 patients) while the number of CNN features used is large (2048).
- Similarly, the VAE training suffered from the small dataset size, which we attempted to alleviate by using data augmentation techniques (transposing, rotation, flip, and noise). Nevertheless, the unsupervised learning of VAEs did not appear to learn lesion features effectively. We attempted to fix this by training the VAEs only on the bounding boxes that contain mostly the cancer lesion and very little background which slightly improved the AUC.

Feature Extraction Method	XGBoost AUC	Logistic Regression AUC
Pyradiomics	0.7411	0.5756
VAE Whole Image (40e)	0.5200	0.5499
VAE Bounding Box (300e)	0.5442	0.5205
VAE Bounding Box (400e)	0.5468	0.5405
CNN Whole Image (ResNet50)	0.5252	0.6299
CNN Bounding Box (ResNet50)	0.5383	0.5783
CNN Bounding Box (VGG16)	0.5295	0.6633

Table 1. AUC for the various feature extraction methods used to illustrate their performance with different classification models. *Note that these are preliminary results, and the full analysis is not yet complete.*

Conclusion

- Our package implements diverse image preprocessing, feature extraction, and analysis methods, with the aim of adaptability for various datasets/modalities and downstream tasks.
- Using the LUNG1 CT DICOM non-small cell lung cancer dataset, we demonstrate functionality of image preprocessing and normalization, feature extraction, and evaluate the predictive performance of the extracted features using classification models.
- We simplify the time consuming and computationally intensive tasks of image intensity normalization, voxel and bounding box normalization, and feature extraction with pretrained Convolutional Neural Networks, Pyradiomics radiomic analysis features, and Variational Autoencoders.
- To evaluate features, we use XGBoost and Logistic Regression classifiers, calculating the AUC for performance measurement.
- Feature extraction and classification is demonstrated based on full CT images, as well as for bounding boxes, with only the region of interest as input, ignoring the background/noise.
- Our package offers a streamlined, concise approach for medical image preprocessing and feature extraction for downstream analysis. In the future, when it is ready for a wider release, our package can be used for a variety of regulatory science research efforts at the FDA.

Acknowledgements

This project was supported in part by an appointment to the Research Participation Program at the U.S. Food and Drug Administration administered by the Oak Ridge Institute for Science and Education through an interagency agreement between the U.S. Department of Energy and the U.S. Food and Drug Administration.