

QHCC

Generated by Doxygen 1.12.0

1 A baseline for machine-learning-based hepatocellular carcinoma diagnosis using multi-modal clinical data—QHCC	1
1.1 The example codes about how to use the QHCC dataset for machine learning tasks	1
1.2 Dependencies	1
1.3 Extracting radiomics features	1
1.4 Experiments on TNM staging	1
2 Namespace Index	3
2.1 Namespace List	3
3 Namespace Documentation	5
3.1 demo_QHCC_XGBoost Namespace Reference	5
3.1.1 Detailed Description	5
3.2 extract_QHCC_radiomics_features Namespace Reference	5
3.2.1 Detailed Description	6
3.2.2 Variable Documentation	6
3.2.2.1 radiomics_data	6

Chapter 1

A baseline for machine-learning-based hepatocellular carcinoma diagnosis using multi-modal clinical data–QHCC

1.1 The example codes about how to use the QHCC dataset for machine learning tasks

1.2 Dependencies

- Python-3.8.8
- numpy-1.21.5
- pandas-1.4.2
- sklearn-1.1.2
- xgboost-2.1.3
- pyradiomics-3.0.1
- SimpleITK-2.4.0

1.3 Extracting radiomics features

Download the data and modify the folder path in `extract_QHCC_radiomics_features.py` accordingly.
`python extract_QHCC_radiomics_features.py`

1.4 Experiments on TNM staging

`python demo_QHCC_XGBoost.py`

Cross-Validation Results on different combinations of input data:

| Data type | Redcap data || Lab data || Redcap + Lab data || Null ||

	ACC	AUC	ACC	AUC	ACC	AUC	ACC	AUC
CT	0.88 ± 0.04	0.90 ± 0.06	0.51 ± 0.08	0.68 ± 0.08	0.83 ± 0.07	0.90 ± 0.02	0.55 ± 0.08	0.64 ± 0.06
MRI	0.83 ± 0.12	0.88 ± 0.09	0.49 ± 0.06	0.65 ± 0.10	0.89 ± 0.05	0.93 ± 0.03	0.51 ± 0.10	0.57 ± 0.08
CT + MRI	0.88 ± 0.05	0.90 ± 0.07	0.55 ± 0.04	0.69 ± 0.09	0.86 ± 0.03	0.92 ± 0.02	0.54 ± 0.09	0.60 ± 0.03
Null	0.82 ± 0.12	0.88 ± 0.12	0.48 ± 0.06	0.50 ± 0.06	0.85 ± 0.07	0.90 ± 0.04	Null	Null

Chapter 2

Namespace Index

2.1 Namespace List

Here is a list of all documented namespaces with brief descriptions:

demo_QHCC_XGBoost	
The demo file for doing classification on QHCC datasets	5
extract_QHCC_radiomics_features	
Extracting radiomics features from medical images with given masks	5

Chapter 3

Namespace Documentation

3.1 demo_QHCC_XGBoost Namespace Reference

The demo file for doing classification on QHCC datasets.

3.1.1 Detailed Description

The demo file for doing classification on QHCC datasets.

Version

v1.1

3.2 extract_QHCC_radiomics_features Namespace Reference

Extracting radiomics features from medical images with given masks.

Variables

- **extractor** = featureextractor.RadiomicsFeatureExtractor()
Initialize the PyRadiomics feature extractor.
- str **mri_dir** = "./data/L1_files"
File paths: MRI image and Mask file directories.
- str **mask_dir** = "./data/L1_files_mask"
- **mri_files** = sorted([f for f in os.listdir([mri_dir](#)) if f.endswith(".nii")])
Get lists of MRI and Mask files default file format: ".nii".
- **mask_files** = sorted([f for f in os.listdir([mask_dir](#)) if f.endswith(".nii")])
- [radiomics_data](#) = pd.DataFrame()
Initialize a DataFrame to store the features.
- **mri_path** = os.path.join([mri_dir](#), [mri_file](#))
Batch processing.
- **mask_path** = os.path.join([mask_dir](#), [mask_file](#))
- **image** = sitk.ReadImage([mri_path](#))

- Load MRI image and Mask.*
- **mask** = sitk.ReadImage(mask_path)
- **temp_mri_path** = os.path.join(mri_dir, f"temp_{mri_file}")
- Save the adjusted MRI image.*
- **features** = extractor.execute(temp_mri_path, mask_path)
- Extract features.*
- dict **flat_features** = {key: value for key, value in features.items()}
- Flatten features and convert them to a DataFrame row*
- Use the filename as the ID.*
- **feature_row** = pd.DataFrame([flat_features], index=[mri_file.split('.')[0]])
- str **output_path** = "radiomics_features.csv"
- Save Radiomics features to CSV.*
- **index**

3.2.1 Detailed Description

Extracting radiomics features from medical images with given masks.

Version

v1.0

3.2.2 Variable Documentation

3.2.2.1 radiomics_data

```
extract_QHCC_radiomics_features.radiomics_data = pd.DataFrame()
```

Initialize a DataFrame to store the features.

Add to the main DataFrame.