Improving accuracy in Diagnosing Disease using Radiomics

Utkarsh Rastogi
Department of Computer Engineering and Application
GLA University

ABSTRACT

Artificial intelligence is the term used to describe the use of computers and technology to simulate intelligent behavior and critical thinking comparable to a human being. In 1956, John McCarthy described the term AI as the science and engineering of making intelligent machines. However, the possibility of machines being able to simulate human behavior and think was raised earlier by Alan Turing who developed the Turing test to differentiate humans from machines and since then the computational power has grown to such extent that new data can be evaluated precisely based on the previously assessed data, in real-time. Today, AI has begun to be incorporated into medicine to improve patient care by speeding up processes and achieving greater accuracy, opening the path to providing better healthcare overall. Radiological images like CT scans, MRI, X-Rays, electronic medical records are evaluated by machine learning models which helps in the process of diagnosing and treatment of the patient. Radiomics is one such advanced technology to diagnose medical images. It extracts a large number of features from medical images via algorithms that help to diagnose disease characteristics. The collected data is in a raw form that can be used to build descriptive and predictive models related to image features. The raw data is mined with other patient data to improve accuracy. The result of these models can provide valuable diagnostics and predictive pieces of information for the tested diseases. The conclusion of radiomics is to reduce medical image diagnosis errors and improve medical facilities. The aim of this study is to present a standard-based evaluation of the disease through radiomics.

Keywords

CNN, Radiomics, ML, Neural Networks

1. INTRODUCTION

With high-performance computation power and machine learning, we can extract important and numerable features from medical images. Collection and mining this medical data into high dimensional data is known as Radiomics. It has the concept that medical images contain pieces of information that reflect key functional changes that accompany a particular disease and can be discovered by quantitative image analyses. Radiomics is based on two artificial intelligence methods. The first being, it can be a hard-coded feature that relies on expert domain knowledge, and the second being based on Deep learning methods which learn from the features itself from the extracted data. Also for designing a radiomics tool you need precisely labeled data but not in large quantity since we can extract a million features from one single image. Radiomics is similar to a Computer-aided diagnosis and Detection (CAD) system but has a significant difference, where CAD delivers only single answers i.e presence of a disease or not. Radiomics is divided into multiple processes.

Process Diagram
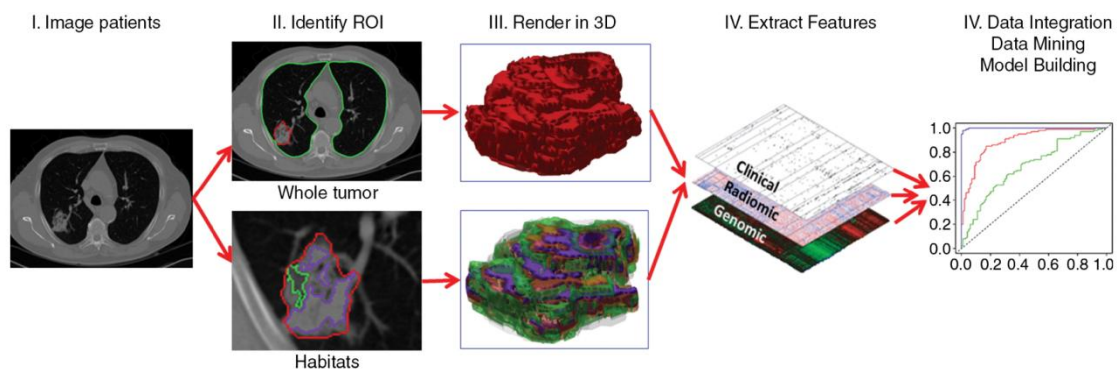
The above flowchart describes the process of radiomics and and how it helps in making decisions. The flowchart is a process model for detection of tumour. A high quality image of CT Scan of the patient is collected and region of interest (ROI) or the subregions (habitats) with the tumour is identified. After identification the region is render into a 3 Dimensional model. Extraction of features from these rendered images is done and saved in the database to design models to diagnosis and prognosis.
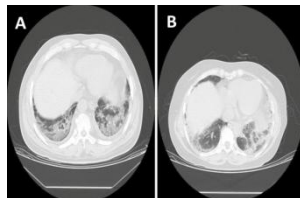
Disease Background: In 2018 UNICEF claimed 800,000 children below the age of 5 died due to pneumonia or say every 39 sec, a child died in some part of the world due to pneumonia, a disease that is a curable and mostly preventable disease. Similarly, 437000 and 272000 children die due to diarrhoea and malaria respectively. These types of diseases are easily preventable and can be easily treated by low-cost antibiotics if properly diagnosed. The UNICEF report also stated that children with severe cases of pneumonia may require oxygen treatment and many times it is ignored due to

poor medical facilities and improper judgment

Pneumonia is an infectious disease that causes inflammation in the lungs. These inflammation are often filled with fluid or purulent material which causes cough, fever, and breathing difficulty. It is caused by a variety of organism that includes but not limited to bacteria, viruses, and fungi. The different organism causes different types of pneumonia. For example, bacteria may cause bacterial Pneumonia, viral pneumonia, Mycoplasma pneumonia, etc. At present Covid-19 also causes Pneumonia with different symptoms. Each type having different symptoms but may look similar. It is detected by radiographic techniques like Chest CT scan, Chest X-ray, etc. In addition to the disease poor medical infrastructure adds up to this high mortality rate. Radiomics can help to reduce this high mortality rate by helping in making better medical judgments of the conditions.

CT Scans of Paraquat poisoning patient and pneumonia patient.

2. IMAGE ACQUISITION

Extraction of meaningful data is to be done from the precisely labeled data but variations in collecting and reconstructing parameters of these images have a vital role that is not due to the underlying disease. And these variations can cause changes. Another challenge is the diseases that look similar can confuse. For example, the CT scans of Pneumonia and acute Paraquat lung injury are so similar and overlap to such an extent that they are difficult to distinguish by eyes but radiomics can easily differentiate if it is correctly configured. Thus for radiomics, very precisely labeled data is needed. Moreover, error bars are to be added for credibility as done in measurement in labs. The dataset dimensions are 1072x712, color space- grey, and No alpha channel and are labeled a. Total images in the dataset 5232 out of which 3883 are characterized as pneumonia with the further division as 2538 as bacterial pneumonia and 1345 as viral pneumonia. The remaining 1973 x-rays are normal.
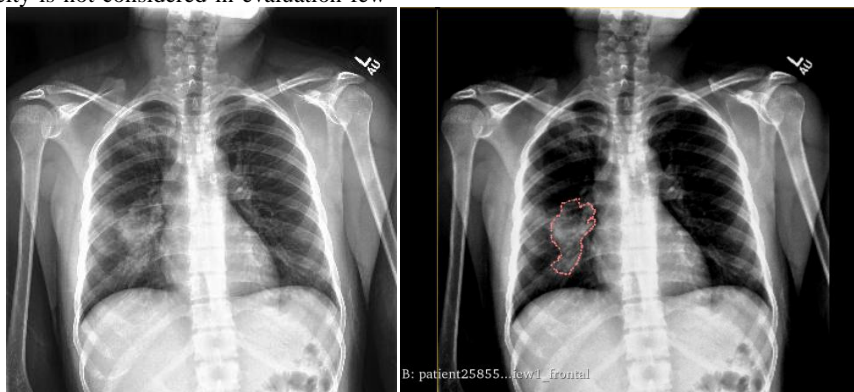
3. REGION OF INTEREST IDENTIFICATION

Identification of the region of interest is the core ideology of the practice of radiomics in oncology. Since these regions or subregions may be distributed over the site we need to identify them all as the volume of interest. However, analysis of subregions with the volumes the prognostic values are not analyzed due to the limitations in the digital images. While the generic heterogeneity is not considered in evaluation few more complex diagnostic imaging reporting and data systems such as Lung Imaging Reporting and Data System (Lung-RADS) are included. The basic idea of including them is to capture as much data as possible at the front, store it in a database, and mine to identify the features for the diagnosis. For identifying the region of interest we are using OpenCV, using the function cv2.cvtColor(src, code[, dst[, dstCn]]) and converting the image into RBG colors, and remove noise to identify the RoI precisely.

Left image is a raw chest x-ray with bacterial pneumonia. The right side image is a processed x-ray with identified region of interest. This resultant image has to be segmented for feature extraction and further quantitive analysis.

4. IMAGE SEGMENTATION

The result of image segmentation is a set of the block that covers an entire area. Each part of the segmented image consists of the same characters such as color, texture, and intensity. Image segmentation is one of the most critical and challenging components of radiomics. It is due to irregular shapes and borders. However, another challenge is that reproducible segmentation with computer aided edge has to be followed by manual curation. Since complete segmentation via computer automation is not possible, an initial examination is also required. For segmentation, we can use software like 3D Slicer, MITK, ITK-SNAP, MeVisLab, LifEx, or ImageJ. We have used 3d slicer to segment the image. We can use the segmented image to create a 3D model of it with the help of interpolation algorithms like marching

cubes.

Since computer-aided segmentation requires manual curation it is suggested that after identification of RoI or VoI, manual or semi-automation should be used irrespective of it being time-consuming. Though many deep learning algorithms have been developed, they can segment an entire organ but cannot segment a disease. Since feature extraction is completely based on the resultant image, precision must be carried seriously.

Fluorodeoxyglucose (FDG)-positron emission tomography (PET) with MRI can produce high definition radiographic images, from which region of interest can be easily identified using contrast-oriented algorithms (computer vision). Another way of identifying the region of interest is by using the deep learning technique. Support vector machines can help in detecting the location of tumors of cancers with high results. Deep learning has also been applied in atlas-based segmentation for diseases like brain and breast cancer and has resulted in extracting high definition image features that can be later used to train a classifier to predict the tumor locations. Convolutional neural networks work predominantly on this to detect diseases. The limitation of this approach is that gross tumor value (GTV) still has to be defined by a practicing oncologist since no standards are present.
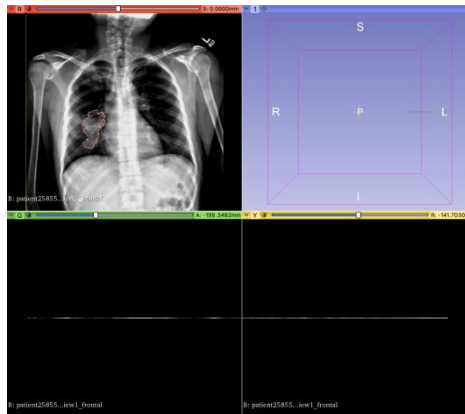
Identification of region of interest and converting it into 3d image for further analysis.

5. FEATURE EXTRACTION AND FEATURE QUALIFICATION

The core of radiomics is feature extraction and qualification. There is a difference between extracting features and selecting them for further process. For training a subtle classifier, we have to reduce the multitude of qualified features as small as possible. To avoid type 1 and type 2 errors these selections of features are important at the same time overfitting of features has to be avoided. Since overfitting of features during classifier training and testing will provide high accuracy but the performance with the new data will degrade. For the selection of these features, one way is to use principal component analysis or covariance matrices.

For extracting features we have used an open-source tool Pyradiomics. We have collected 129 features from the volume of interest out of which 3 groups have been taken for further analysis. These groups are as shape, color, and intensity features. For our second test from the same dataset for extracting and selecting features, we have used the Fisher Vector with Convolutional neural networks. Fisher Vector is are random vectors with larger dimensions, this results in higher accuracy and speed.

6. DESIGNING MODEL FOR ANALYSIS

Machine learning is an integral and import part for radiomics. Supervised and unsupervised learning models can be used for radiomics. For a simple radiomics any one of the two can be used but for superior results both have to be combined.

Supervised models requires labelled data (here we have labelled features) that will allow it search for the same features in the test images. It results in predicting the survival outcome. On the contrary unsupervised learning does not required labelled dataset but it does require a result or an outcome. The discovered features are used with classifier models like decision tress, random forest and Gaussian mixture modelling. These models help in diagnosis and prognosis of medical diseases, in our case pneumonia.

With our second test direction of using fisher vector, we have designed using clusters where the region of interest has been divided into sub regions and these regions acts as clusters of potentials for our unsupervised learning. These potential clusters can be further classified into regions such as region of high tissues and regions of necrotic tissues. These sub volumes can help in detecting high risk.

With both of the above methods we are able to detect pneumonia. Another approach is using deep learning. CNN are widely used to classify 2D and 3D images and have higher and speedy accuracy. CNN simulates the human brain in translating the visual inputs.

7. RESULT ACCURACY

The above three methods provide detection of diseases as per the requirements with higher accuracy and low loss. But for further preciseness, we can actually use the statistics approach and take mean of the results. This will provide us a conclusive result and allow us for making better predictions in diagnosis and prognosis.

Prediction of disease using supervised learning.

8. ISSUES

There are no standard present when it comes to radiomics. For extracting features, it is an individuals choice wether to accept or reject a particular feature, which has a direct impact on classifying model or result. Moreover, CNN has overfilling of features, therefore during training in the model had an accuracy of almost 100% which is impressive in the beginning but the performance reduces significantly when a raw image is fed for testing.

Another challenge is the availability of labelled medical images as patient privacy is the first concern and no open data source is available for independent researches.

```
Epoch 1/5
163/163 [==============================] - 773s 5s/step - loss: 0.1329 - accuracy: 0.9605
Epoch 2/5
163/163 [==============================] - 765s 5s/step - loss: 0.0480 - accuracy: 0.9814
Epoch 3/5
163/163 [==============================] - 762s 5s/step - loss: 0.0418 - accuracy: 0.9858
Epoch 4/5
163/163 [==============================] - 766s 5s/step - loss: 0.0465 - accuracy: 0.9849
Epoch 5/5
163/163 [==============================] - 762s 5s/step - loss: 0.0254 - accuracy: 0.9904
<keras.callbacks.callbacks.History at 0x7fd2aa937630>
```

CNN based model to predict pneumonia providing higher epoch accuracy than the traditional approach due to overfitting.

9. CONCLUSION

Imaging biomarkers constructed from quantitative image analysis have great potential to advance precision medicine and to enhance cancer biology knowledge. As radiomics cements its position in translational cancer research to attain utilization at bedside, we anticipate radiomics data being integrated and analyzed with genomics, proteomics and other -omics; providing information invaluable in personalized medicine. Radiomics will certainly progress further with the advent of more imaging data, better algorithms, and availability of other data types such as coherent datasets integrating imaging, clinical, and genomic information.

10. REFERENCES

[1] Even AJ, Reymen B, La Fontaine MD, et al. Clustering of multi-parametric functional imaging to identify high-risk subvolumes in non-small cell lung cancer. Radiother Oncol 2017;125:379-84. [Crossref] [PubMed]

[2] https://www.annalsofoncology.org/article/S0923-7534(19)32412-3/fulltext

[3] Gillies R.J.Kinahan P.E.Hricak H. Radiomics: images are more than pictures, they are data.Radiology. 2016; 278: 563-577

[4] Lambin P.Rios-Velazquez E.Leijenaar R.et al. Radiomics: extracting more information from medical images using advanced feature analysis Eur J Cancer. 2012; 48: 441-446

[5] El Naqa I, Bradley JD, Lindsay PE, et al. Predicting radiotherapy outcomes using statistical learning techniques. Phys Med Biol 2009;54:S9-30.

[6] Leijenaar RT, Nalbantov G, Carvalho S, et al. The effect of SUV discretization in quantitative FDG-PET Radiomics: the need for standardized methodology in tumor texture analysis. Sci Rep 2015;5:11075. [Crossref] [PubMed]

[7] Coroller TP, Grossmann P, Hou Y, et al. CT-based radiomic signature predicts distant metastasis in lung adenocarcinoma. Radiother Oncol 2015;114:345-50. [Crossref] [PubMed]

[8] Ypsilantis PP, Siddique M, Sohn HM, et al. Predicting Response to Neoadjuvant Chemotherapy with PET Imaging Using Convolutional Neural Networks. PLoS One 2015;10:e0137036. [Crossref] [PubMed]

[9] Kickingereder P, Burth S, Wick A, et al. Radiomic Profiling of Glioblastoma: Identifying an Imaging Predictor of Patient Survival with Improved Performance over Established Clinical and Radiologic Risk Models. Radiology 2016;280:880-9. [Crossref] [PubMed]

[10] Larue RT, Defraene G, De Ruysscher D, et al. Quantitative radiomics studies for tissue characterization: a review of technology and methodological procedures. Br J Radiol 2017;90:20160665. [Crossref] [PubMed]

[11] Parmar C, Leijenaar RT, Grossmann P, et al. Radiomic feature clusters and Prognostic Signatures specific for Lung and Head & Neck cancer. Sci Rep 2015;5:11044. [Crossref] [PubMed]

[12] Parmar C, Grossmann P, Bussink J, et al. Machine Learning methods for Quantitative Radiomic Biomarkers. Sci Rep 2015;5:13087. [Crossref] [PubMed]

[13] Balagurunathan Y, Kumar V, Gu Y, et al. Test-Retest Reproducibility Analysis of Lung CT Image Features. J Digit Imaging 2014;27:805-23. [Crossref] [PubMed]

[14] Parmar C, Grossmann P, Rietveld D, et al. Radiomic Machine-Learning Classifiers for Prognostic Biomarkers of Head and Neck Cancer. Front Oncol 2015;5:272. [Crossref] [PubMed]

[15] Balagurunathan Y, Gu Y, Wang H, et al. Reproducibility and Prognosis of Quantitative Features Extracted from CT Images. Transl Oncol 2014;7:72-87. [Crossref] [PubMed]

[16] Cunliffe A, Armato SG, Castillo R, et al. Lung Texture in Serial Thoracic Computed Tomography Scans: Correlation of Radiomics-based Features With Radiation Therapy Dose and Radiation Pneumonitis Development. Int J Radiat Oncol Biol Phys 2015;91:1048-56. [Crossref] [PubMed]

[17] Leijenaar RT, Carvalho S, Hoebers FJ, et al. External validation of a prognostic CT-based radiomic signature in oropharyngeal squamous cell carcinoma. Acta Oncol 2015;54:1423-9. [Crossref] [PubMed]

[18] Zhu Y, Li H, Guo W, et al. Deciphering Genomic Underpinnings of Quantitative MRI-based Radiomic Phenotypes of Invasive Breast Carcinoma. Sci Rep 2015;5:17787. [Crossref] [PubMed]

[19] Sollini M, Cozzi L, Antunovic L, et al. PET Radiomics in NSCLC: State of the art and a proposal for harmonization of methodology. Sci Rep 2017;7:358. [Crossref] [PubMed]

[20] Kumar V, Gu Y, Basu S, et al. Radiomics: The process and the challenges. Magn Reson Imaging 2012; 30:1234-48. [Crossref] [PubMed].

[21] Hawkins SH, Korecki JN, Balagurunathan Y, et al. Predicting Outcomes of Nonsmall Cell Lung Cancer Using CT Image Features. IEEE Access 2014; 2:1418-26.