Al-Nahrain Journal for Engineering Sciences **NJES** 27(2)185-192,2024
http://doi.org/10.29194/NJES.27020185

Design of MATLAB-based Radiomics Classifier Training Simulator Powered by Pyradiomics

Muhammed Selman Erel[1*], Esra Şengün Ermeydan[1], Hadeel K. Aljobouri[2], Ilyas Çankaya[1]

Authors affiliations:
1) Electrical and Electronics Engineering Department, Graduate School of Natural Science, Ankara Yıldırım Beyazıt University, 06010 Ankara, Turkey
mserel@aybu.edu.tr

2) Biomedical Engineering Department, College of Engineering, Al-Nahrain University, Baghdad, Iraq.

Paper History:

Received: 26th Oct. 2023

Revised: 14th June 2024

Accepted: 23rd Aug. 2024

Abstract

Technically, medical imaging modalities are quantitative, qualitative, and semi-quantitative. Such modalities can generate meaningful and valuable quantitative and qualitative data. Correlating predictive outcomes with quantitative and qualitative data is a difficult process. Thanks to modern computational hardware and advanced machine learning algorithms, it is not a demanding job to perform predictive analysis by cultivating quantitative and qualitative data. Radiomics is a popular topic that studies quantitative data from medical images in order to obtain biologically meaningful information for diagnosis, prognosis, theragnosis, and decision support. Handcrafted radiomics is a process including features based on shape, pixel, and texture-related knowledge from medical scans. In the pursuit of advancing the field of radiomics, we have developed a cutting-edge radiomics training simulator, powered by MATLAB. This tool has been designed for those familiar with MATLAB, making it easy for them to transition into the fascinating world of radiomics. MATLAB's user-friendly interface and strong support in the engineering community provide an ideal platform for this simulator, ensuring aspiring radiomics learners have access to the resources they need for success. Throughout the paper, purpose, design details and methodology of the simulator are described. As a case study, accuracy values for KNN, SVM, Naïve Bayes and Linear Classifiers are 0.56, 0.52, 0.25 and 0.60, respectively. Although accuracy values seem lower, recall of 0.85 for Linear Classifier gives precise results while considering medical application.

Keywords: GUI, LUNG1, MATLAB, Machine Learning, Radiomics, TCIA

تصميم نظام محاكاة للتدريب على مصنف علم الإشعاع القائم على الماتلاب بالأعتماد على البيراديوميات

محمد سلمان إيريل، إسراء شينغون إرميدان، هديل قاسم الجبوري، إلياس جانكايا

الخلاصة:

من الناحية الفنية، تعد طرق التصوير الطبي كمية, نوعية, وشبه كمية. ويمكن لهذه الطرائق أن تولد بيانات كمية ونوعية مفيدة وقيمة. يعد ربط النتائج التنبؤية بالبيانات الكمية والنوعية عملية صعبة. بفضل الأجهزة الحسابية الحديثة وخوارزميات التعلم الآلي المتقدمة، فإن إجراء التحليل التنبؤي من خلال تنمية البيانات الكمية والنوعية أصبحت ليس بمهمة صعبة. يعد علم الإشعاع موضوعًا شائعًا يدرس البيانات الكمية من الصور الطبية من أجل الحصول على معلومات ذات معنى بيولوجي للتشخيص والتنبؤ والعلاج ودعم القرار. أن ميزات علم الإشعاع المتعامل معها يدويًا هي عبارة عن عملية تعتمد على الخصائص المشتقة من الشكل والبكسل وبنية النسيج الموجودة في الصورة نفسها. في السعي للتقدم في مجال علم الإشعاع ، قمنا بتطوير نظام محاكاة متطور للتدريب الإشعاعي, بالأعتماد على برنامج الماتلاب. تم تصميم هذه الأداة لأولئك المطلعين على برنامج الماتلاب، مما يسهل عليهم الانتقال والتعامل مع عالم الاشعاع الراديوي الرائع. توفر واجهة برنامج الماتلاب المصممة سهولة للاستخدام والدعم القوي في المجال البحثي والعملي للمجتمع الهندسي عن طريق توفير منصة مثالية لهذه المحاكاة، مما يضمن حصول المتعلمين الطموحين في علم الإشعاع الراديوي على

المصادر التي يحتاجونها لتحقيق النجاح. خلال هذا البحث، تم وصف الغرض, تفاصيل التصميم ومنهجية نظام المحاكاة بالتفصيل. كدراسة حالة، تبلغ قيم الدقة لـ KNN وSVM وNaïve Bayes والمصنفات الخطية ٠,٥٦ و٠,٥٢ و٠,٢٥ و٠,٦٠ على التوالي. على الرغم من أن قيم الدقة تبدو أقل، إلا أن استدعاء ٠,٨٥ للمصنف الخطي يعطي نتائج دقيقة عند النظر في التطبيق الطبي.

1. Introduction

Medical imaging technology dates back to the discovery of X-ray, 125 years ago [1]. Usage of computerized tomography, magnetic resonance imaging, positron emission tomography, ultrasound imaging become de facto standard for modern-day clinical practice. Today's radiologists can image part of or whole human body in detail using state-of-the-art imaging modalities, Computed Tomography (CT),

Magnetic Resonance Imaging (MRI), Positron Emission Tomography (PET). Instruments of this kind have evolved into essential components for diagnosis, prognosis, theragnosis, and decision support. Their low cost, non-invasive imaging capabilities, promising technological advancements, and superior imaging precision make them stand out compared to alternative options.

Medical imaging modalities are based on analog imaging technologies generating qualitative data which urges subjective decisions based on visual or verbal inspections. With the information technologies era, door of the digital world is opened to radiologists, although they still prefer to visual inspections. Thanks to computerization, researchers focused on computerized quantitative analysis of medical imaging for clinical purposes, which is known as (Computer-aided Diagnosis) CAD systems. These systems excel at conducting intricate quantitative assessments using quantitative data. Radiomics is an emerging field that harnesses this type of data to its advantage.

1.1 Radiomics

Radiomics is one of the most popular topics nowadays which processes medical imaging data to obtain quantitative imaging metrics, called as radiomic features. It analyzes tissue and lesion characteristics like heterogeneity, shape, size, textural, filtered-based, histogram-based even combination of clinical and genomics data. Radiomics is pronounced first in the study Hugo Aerts et al. where they studied CT images of 1019 patients with lung or head-and-neck cancer. They utilized publicly available datasets in The Cancer Imaging Archive (TCIA) archive, Lung1, Lung2, Lung3, (Head & Neck) H&N1, H&N2, H&N3 and The Reference Image Database to Evaluate Therapy Response (RIDER) and decoded tumor phenotypes by using noninvasive 440 quantitative radiomics image features.

Radiomics features are categorized into five main groups: Size and shape-based features, image intensity histogram related features, relationship between image voxels/pixels, texture extracted from filtered images and fractal features. Mathematical descriptions of such features can be found in literature independent of imaging modality [2-5]. In order to establish a consensus on radiomics features, an initiative dedicated to standardization, known as the Image Biomarker Standardization Initiative (IBSI), was formed. Within this initiative, a coalition of twenty-five teams collaborated to contribute to the standardization efforts, notably, Hugo Aerts and his team were active participants in this study. They proposed standardized radiomics features in three phases which are Morphological Characteristics, local intensity, intensity-based statistics, intensity histogram, intensity volume histogram, Gray Level Co-Occurrence Matrix (GLCM), Gray Level Run Length Matrix (GLRLM), Gray Level Size Zone Matrix (GLSZM), Gray Level Distance Zone Matrix (GLDZM), Neighborhood Gray Tone Difference Matrix (NGTDM), Neighboring Gray Level Dependence Matrix (NGLDM).

To extract radiomics features, a Python software module, called PyRadiomics, is developed by, of course, Hugo Aerts et al [6]. PyRadiomics can provide four stages of radiomics analysis: Loading and preprocessing of scanning image, performing filtering processes, calculation of radiomics features. For filtering process, Laplacian of Gaussian, Wavelet, Square, Square Root, Logarithm, Exponential are filter types for filtering process. It can extract around 1500 radiomics features from a scanning image those are First Order Statistics, 2D Shape based, 3D Shape based, GLCM, GLRLM, GLSZM, NGTDM and (Gray Level Dependence Matrix) GLDM. It utilizes some third-party package like SimpleITK for preprocessing, numpy for feature calculation, PyWavelets for wavelet filtering, pykwalify for enabling yaml parameter file checking and six for Python 3 compatibility [7].

1.2 Lung Cancer

Radiomics plays an indispensable role in advancing our understanding of cancer, guiding its diagnosis, treatment, and management, heralding a new era of precision oncology. Cancer stands as one of the most prevalent causes of modern mortality, capable of originating anywhere in the body and metastasizing to distant sites. Based on the emerging part, there are many types of cancer: Breast cancer, leukemia, sarcoma, prostate cancer, melanoma, colorectal cancer, head and neck cancer, carcinoma, lung cancer. Lung cancer is the leading cancer type, which kills people most with 1,796,144 deaths reported in 2020, representing %18 of whole cancer deaths. In just a century, it has transformed from being a rare disease into one of the prominent cancer types of the twenty-

NJES is an **open access Journal** with **ISSN 2521-9154** and **eISSN 2521-9162**

first century [8-9]. According to (Global Cancer Observatory) GLOBOCAN, which is an online database about global cancer statistics for 185 countries and 36 types, incident cases of lung cancer is expected to increase by around 64% by 2040 whereas death cases is 67% [10]. Smoking is leading risk factor for lung cancer with 80% related to tobacco consumption. Excess risk among addicted smokers compared to never-smokers is in the order of 20-50-fold [11]. To classify cancer stage, there are two types of cancer staging systems: TNM staging system and numerical staging system. In TNM-staging, T represents size of primary tumor, N shows nearby lymph nodes, N is used to describe distant metastasis. In numerical staging system, progression of cancer is classified. In this staging Roman Numeral is used. Stage is numerated from 0 to IV. As the severity of progression is increased, stage is increased, too. The most common cancer staging system is TNM-Staging System. This staging system for lung cancer is organized by International Association for the Study of Lung Cancer. Currently, 9th edition of TNM staging is proposed and revision for 9th edition is about to come. Revised version is expected to be in use in January 2024 [12]. The most frequent type of lung cancer is non-small cell lung cancer with frequency of %80 of all lung cancer cases. Basically, there are three types of NSCLC (Non-Small Cell Lung Cancer): adenocarcinoma with 40%, squamous with 30% and large cell with 16% [13]. Nowadays, it is global attempt to categorize the tumor as much as possible and to cultivate as many information as possible from the tumor.

1.3 The Cancer Imaging Archive

To promote researchers to utilize and share publicly available datasets, National Institute of Health (NIH) created a imaging archive called The Cancer Imaging Archive (TCIA). It is a centralized scanning image repository containing more than 100 collections (above 3 million images). It includes Quantitative Imaging Network (QIN), Quantitative Imaging Biomarkers Alliance (QIBA), RIDER, LUNG1, LUNG2, LUNG3, Head-Neck Cetuximab and more well-known collections. The most commonly used scanning modality is Computerized Tomography (CT), followed by Magnetic Resonance Imaging (MRI), and then Positron Emission Tomography (PET), among others. The most extensively studied human body parts are, in order of prominence, the colon, followed by the brain, the chest, and finally the breast. The most popular collection of TCIA is LUNG-1 dataset. It is produced from Lung CT scans of 422 NSCLC patients from Maastricht University Medical Center, Limburg, Netherlands. Limburg, Netherlands. It consists of CT DICOM imaging scan files, manual segmentation files and clinical data including age and gender of patients, TNM staging of tumor, survival times and death status of patients. To the best of our knowledge, more than ten studies are carried out by using LUNG1 dataset [14-23]. In this study, LUNG1 dataset is used for testing the Radiomics Training Simulator.

1.4 Literature Survey

There are many simulation studies about radiomics in literature: One of them is by S. Marinov et al. They designed a radiomics software implemented in MATLAB. It is a very handy and elegant GUI app consisting of 5 parts: loading DICOM images, viewing DICOM images, adjusting parameters, creating, and modifying (Region of Interest) ROIs and image masks and calculations sections. Main (Graphical User Interface) GUI is rather neat and compact. They have IBSI-compliant features those can be viewed on the page other than main GUI. They tested their software by utilizing L1 and L2 physical phantoms. They extracted 23 different image related features [24]. Other application developed on MATLAB is by G. Pasini et al. They designed a MATLAB package named as matRadiomics. Contrary to previous study, they have developed a complete radiomics framework including importing DICOM image, segmentation of loaded images, feature extraction powered by pyRadiomics, feature selection and building machine learning models. It can show ROC curve and confusion matrix at the end of analysis. They tested their application by using (The Lung Image Database Consortium image) LIDRC-IDRI dataset. They extracted 107 features [25]. Another radiomics software study is by Z. Zhou et al. They designed a user-friendly software having GUI mostly in Python. Like the study by G. Pasini et al., it consists of complete radiomics framework. The first part is Image preparation. Multi-modality images are loaded and pre-processed, data filtration etc., then it is split in to training set and validation set. The second part is lesion labeling. Lesions or tissues are labeled. The third part is feature extraction. In this part, the software extracts quantitative image features by using pyRadiomics. Next part is feature selection part. To reduce the number of features and to avoid over-fitting, the most relevant features are selected. The part after feature selection is classifier training. In this part, seven machine learning algorithms are available to train. The last part is evaluation and visualization. Prediction performance of the model and feature-related visualizations are given at the end. Other studies, developed and benchmarked radiomics simulators, are given as below.

In the literature, there are few radiomics simulator tools available, and this inadequacy has negative implications for radiologists' training and decision support mechanisms. In this study, to make radiologists and researchers familiarize Artificial Intelligence (AI) and Radiomics and promote to become an active profile in this revolution, MATLAB-based training simulator is designed. Stand-out features of the simulator over it's rivals are simplicity, compactness and detailed classification analysis results. It can provide confusion matrix, ROC curve and classification performance parameters in once. Design details of the simulator is given in materials and methods section, testing results are given in results and discussion part and finally summary of the key findings and insights presented in the study are given in conclusion part. With this study, it is achieved that radiologists can use the designed tool for training and decision support aims.

Table (1): Radiomics Simulator Studies and Their Phantoms and Features

Radiomics Simulators	Dataset/ Phantom	Number of Features	Auth.
Pyradiomics, MITK[1], LIFEx, SERA[2], CaPTk[3],	IBSI Digital Phantom	173 features	2020, M. Lei et al., [26],
matRadiomics	LIDRC-IDRI Dataset	107 features	2022, G. Passini et al., [27]
Pyradiomics, MITK, LIFEx, SERA, CaPTk, A1, A2	IBSI Digital Phantom	173 features	2021, M. Lei et al., [28]
MIRP, S-IBEX[4], Sopiha DDM, RaCaT[5], SERA, PyRadiomics, RadiomiCRO	IBSI Digital Phantom, ImSURE	919 features	2022, A. Bettinelli et al, [29]
RaCaT	NEMA[6] image quality phantom	480 features	2019, E. Pfaehler et al., [30]

[1]**MITK=**The Medical Imaging Interaction Toolkit
[2]**SERA=**The Standardized Environment for Radio7mics Analysis Package
[3]**CAPTk=**Cancer Imaging Phenomics Toolkit
[4]**S-IBEX=**Standardized image biomarker explorer
[5]**RaCaT=**Radiomics Calculator
[6]**NEMA=** National Electrical Manufacturers Association

2. Materials and Methods

2.1 GUI (Graphical User Interface) Design

The simulator software is designed in MATLAB by using AppDesigner. It is designed for classification purposes. User interface design is given in Fig. 1.

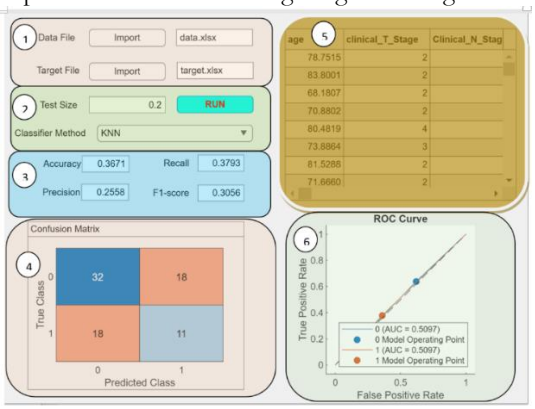

Figure (1): Graphical User Interface Design

The simulator user interface has basically six main parts: Tabular data import part, test size and classifier method selection part, classification performance metrics part, confusion matrix visualization part, tabular data visualization part and ROC curve visualization part. In tabular data part, data and target table files are imported. In test size and classifier method part, desired ratio and method is selected. And simulation is run. In tabular data visualization part, imported data can be viewed. In third part, performance results are observed. In fourth and sixth parts, confusion matrix and ROC curve can be viewed as an auxiliary property.

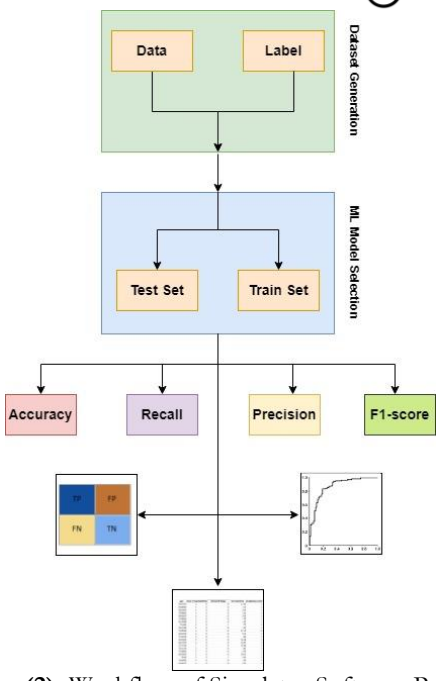

Figure (2): Workflow of Simulator Software Package

Execution of simulator workflow is depicted in detail and carefully in Fig.2. After data and label selection, test and train sets are created. Accuracy, recall, precision, F1-score, confusion matrix, ROC curve and tabular view are visual outputs of the simulator. In this software, to predict status of patients, four classifier algorithms are utilized: KNN, SVM, Naive Bayes and Linear Classifier.

2.1.1 KNN (K-Nearest Neighbour)

It is a simple non-parametric classification algorithm. Main mechanism is to assign unlabeled sample points to the class of the nearest of previously labeled points. Assignment mechanism is based on distance between data items [31]. The first step of the algorithm is to choose best "k", the number of neighbor (NN). Number of neighbor can be chosen by user or based on the impact of performance or even vs. odd cases. NN can be determined by user while considering accuracy and performance, overfitting and underfitting cases, dataset size and data distribution and computational issues. NN is set based on performance metrics like accuracy for underfitting and overfitting issues, recall and precision for the proportion of positive predictions among other cases, F1 score for balance between precision and recall, ROC Curve and AUC for better model performance to maximize AUC.

Table (2): Distance metrics and their formulas

Distance Name	Formula
Eucledian Distance	$\sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$
Cosine Distance	$(x_i \cdot y_i) / (\lvert x_i \rvert \cdot \lvert y_i \rvert)$
Mahalanobis Distance	$\sqrt{(p_1 - q_1)^T S^{-1} (p_1 - q_1)}$
Jaccard Distance	$\frac{\lvert p_1 \cap q_1 \rvert}{\lvert p_1 \cup q_1 \rvert}$

NN can be chosen based on even vs. odd cases. When NN is odd, two neighbors of one class is eventually present, possibly one neighbor of another class is present, ensuring a majority decisive vote. For

the even case, there is always possibility for an equal number of neighbors from different groups stating no majority. The second step of the algorithm is to calculate distance. For a function to be considered as a metric, non-negativity, indiscernibility, symmetry and triangle inequality should be satisfied. The third step is to find nearest neighbors. It is done by calculating smallest distances for K points with target data points. As a final stage, classification is done by majority voting by counting the number of data points in each group among K nearest neighbors. Target data point is classified by most common among its neighbors. There are many distance types: Minkowski distance, Manhattan distance, Cosine distance, Jaccard distance, hamming distance, and Euclidean distance. The most popular distance metric is Euclidean. Some of the distances and their formulas are given in Table 2.

2.1.2 SVM (Support Vector Machine)

Support Vector Machine, shortly SVM, is based on statistical learning theory [31]. How this algorithm works is to seek a hyper plane to separate the data into classes so that maximum margin of separation between classes with minimum error is obtained. The first step of the algorithm is feature selection. It is required to transform the original raw training data into a set of features, for training the classifier. There are three main types of feature selection methods: embedded methods, filter methods and wrapper methods. Embedded methods are one of most common method in SVM analysis. In embedded methods, to select features for raw input dataset, kernel functions are utilized. Linear, nonlinear, polynomial, sigmoid, gaussian, ANOVA, radial basis function are kernel functions for SVM. The most popular of them is radial basis function. Some of the kernels and their functions are given in Table 3.

Table (3): Kernels and Their Functions

Kernel	Kernel Function
Linear	$x^T y$
Polynomial	$(\gamma x^T y + c_0)^p$
RBF[1]	$\exp(\gamma x^T y + c_0)$
Tanh	$\tanh(\gamma x^T y + c_0)$

[1]**RBF**=Radial Basis Function

Kernel trick or kernel method is utilized to transform raw input data to a higher-dimensional feature space by using similarity function with kernel matrix. Kernels in models function as mapping data to higher dimensional space and decision boundary creation. Linearly seperable data is more favorable than non-seperable. To make the data seperable, kernels precisely map the input data into a higher-dimensional feature space where data is linearly seperable. Correct kernel can capture the regarding pattern in data to create a linear separation. Better separation, better performance, higher accuracy. Kernels consume higher computational resources while operating. Therefore, correct kernel can be more computationally-intelligent than others. To find a decision boundary, kernel function induces a hyperplane that separates feature set elegantly. As a second stage, SVM is trained by using already-known labelled prior data. As a final stage machine learning model is evaluated based on its sensitivity, specificity and accuracy in differentiating between classes.

2.1.3 Naive Bayes

Naive Bayes (NB) is a probabilistic classifier which is based on Bayes Theorem. It classifies the dataset by building a function assuming all features in the dataset are independent. This algorithm takes less time for training process. Contrary to KNN and SVM, there are no free parameters to set. Therefore, implementation of NB is easier. In NB, corresponding instances of n-dimensional vector of random features in domain D_x are associated with unknown target values in domain D_y. Association work is done by a probabilistic function. The aim of this method is to select unobserved random variables set which maximizes the posterior probability. Classification is performed based on a probabilistic calculations. NB is so called naïve because NB treats all instances in the domain D_x regardless of calculation order. It is naïve because it is assumed that probability of each instance is independent of any other instances. By using Bayes Theorem, overall probability is computed by calculating prior probability of a class, conditional probability of instances and probability of each instances. All in all, it is based on a simplistic hypothesis and it's practical effectiveness, simplicity and efficiency make it popular among other methods in classification work of machine learning.

2.1.4 Linear Classifier

Linear classifier (LC) is a one of the simplest and easiest method for classification. It is utilized for binary or multiclass classification tasks. Attributes of the dataset are given to the model as vector form. This classifier method works very well such as text classification and rather good for datasets with many variables. In LC, every feature or attribute is weighted in the training phase by using labeled data and each weight of the model is tuned by utilizing optimization techniques. Gradient descent, Newton's Method, Conjugate Gradient, Coordinate Descent, Bayesian Optimization, Evolutionary optimization and regularization techniques are some of the well-known optimization techniques to adjust the weight of each features. Weight is updated based on the loss function of the optimization operation.

Table (4): Linear Functions

Distance Name	Formula
Linear Function	$f(x) = w^T x + b$
Perceptron Model	$f(x) = sign(w^T x + b)$
Logistic Regression	$f(x) = \frac{1}{1 + e^{-(w^T x + b)}}$
Softmax Function	$f_j(x) = \frac{e^{w_j^T x + b_j}}{\sum_{k=1}^{K} e^{w_k^T x + b_k}}$

After operation, a linear classifier is fed with input data points with feature vector having adjusted weights. Linear classifier multiplies the weight with input data and it is classified based on the product passes threshold or not. There are many linear functions in literature but some of them are listed in the Table.4.

Working procedure of the simulator consists of 6 main steps. First, files concerning data and target is uploaded to the software. Then, data is split into test

and train based on split ratio. After that, classification method is chosen. After these pre-execution steps, classification process is started. As classification performance metrics, confusion matrix, tabular data view and ROC curve are obtained for analysis. Instruction-flow for the app user is given in Fig.3.

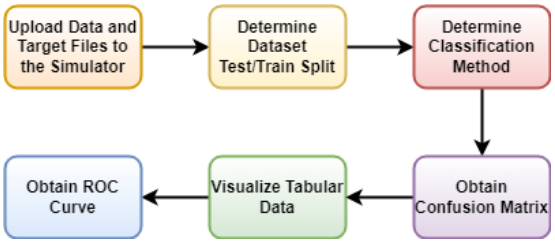

Figure (3): Instruction-flow for the App User

2.2 Dataset

In this study, a publicly available lung cancer dataset consisting of DICOM and clinical data, LUNG1 dataset, is employed. LUNG1 dataset is collected by researchers in MAASTRO, Maastricht University Medical Centre+, Department of Radiotherapy. This dataset is the collection of clinical and radiological image data of 422 NSCLC lung cancer patients.

Table (5): LUNG1 dataset Clinical Features

Age			
<65	150		
>65	250		
Mean (years)	68		
Gender			
Male	290		
Female	132		
Tumor T-Stage			
T1	**T2**	**T3**	**T4**
93	156	53	117
Tumor N-Stage			
N0	**N1**	**N2**	**N3**
170	23	141	85
Tumor Overall Stage			
I	**II**	**IIIa**	**IIIb**
93	40	112	176
Histology			
Large cell	**Squamous cell carcinoma**	**adenocarcinoma**	
114	152	51	

This dataset is announced first time in literature in the study by Hugo Aerts et al [32]. In their study, they employed 7 datasets to cultivate decision-making supportive radiomics features. LUNG1 dataset consists of 52073 CT-scan images with size of 33 GB. It includes RTSTRUCT (radiotherapy structure set) and DICOM SEG files having manual delineation by a professional radiation oncologist. Detailed information about the dataset is given in the Table.5.

3. Results and Discussion

To test the classifier simulator, 2-year Survival classification of LUNG1 data is performed. In this dataset, there are both clinical and radiological data for 422 NSCLC patients. Survival-status of the patients are modified whether they are alive or not up to 2 years. This tabular data is defined as "Target File" and it is imported. Radiomics features of DICOM images of 422 patients are extracted by PyRadiomics. Resultant tabular data is defined as "Data File" and it is imported. For the sake of rule of thumb, test data split of size of input data is defined as 0.2. Classification performance metrics are calculated based on (1-4). Accuracy is a performance metric which correctly measure the proportion of correctly classified instances among whole instances. Although, it is used as a general measure of model prediction performance, it fails in imbalanced classes where one class dwarfs other class. Also, it is rather risky to just use it in precision-matter applications like medical, military or financial applications. Precision is another metric to show model performance. It measures the proportion of true positive predictions among all predictions. Just like Accuracy, it is not recommended to just use it in cost-aware applications. If cost of false positive is high and it is desired to minimize false alarms, other metrics than precision should be concerned. Recall or sensitivity is another metric to measure true positive predictions among all actual positive instances. If false negative is a concern in sensitive applications, recall should not be used individually. In cases of incorrect flag-up, it can mislead the prediction system. F1 score is a hybrid metric of precision and recall. It is harmonic mean of them. It is more reliable than accuracy alone in case of uneven class distribution or false positive or false negative are concerned.

$$Accuracy = \frac{TP^1 + TN^2}{TP + FN^3 + TN + FP^4} \quad (1)$$

$$Recall = \frac{TP}{TP + FN} \quad (2)$$

$$F1 = 2 \times Precision \times \frac{Recall}{Precision + Recall} \quad (3)$$

$$Precision = \frac{TP}{TP + FN} \quad (4)$$

[1]True Positive, [2]True Negative,[3]False Negative, [4]False Positive

For four different classifiers of 2-year survival status, results of the metrics are given in Table 6 and Fig.4. It can be seen that LC has dramatically better performance.

Table (6): Performance Metrics for Four Classifier Methods

Classifier Method	**Accur.**	**Precis.**	**Recall**	**F1**
KNN	0.56	0.43	0.59	0.49
SVM	0.25	0.20	0.34	0.25
NaiveBayes	0.52	0.43	0.56	0.49
Linear	**0.60**	**0.45**	**0.85**	**0.59**

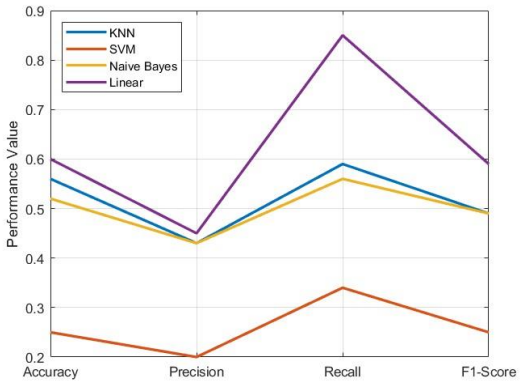

Figure (4): Performance Results for ML Models

Among four classifier methods, the most powerful classifier method is Linear Classifier Method with the recall of 0.85 where SVM has initial settings, kernel as rbf, kernel polynomial function degree as 3, regularization parameter as 1.0 and tolerance for stopping criterion as 0.001. Although performance metrics like accuracy and precision seem low, feature selection could be a way to improve the performance metrics because there are 1224 features, most of them do not have effect on performance metrics. To improve the classification performance, elegant feature selection methods can be applied. Close values of F1-score and accuracy for LC show that dataset class distribution is balanced for the test case and accuracy can be used as a performance metric alone.

4. Conclusion

Using MATLAB AppDesigner, a training simulator is designed, dedicated to training of radiomics applications. The software package was tested successfully for LUNG1 dataset of NSCLC patients. The tool was designed for applications for NSCLC disease, but it is not limited to this application. Although it contains radiomics features of CT lung scanning DICOM images of NSCLC patients, other modalities can be used. This study is conducted to accustom radiologists to artificial intelligence and radiomics revolution and make them participate and contribute to this field. Case test/train split study shows that although accuracy values of KNN, SVM, NB and LC, those are 0.56, 0.25, 0.52 and 0.60, respectively, seem lower, LC shows highly promising performance especially recall of 0.85 in medical application is considered, where sensitivity is concerned.

5. References

[1] J. H. Scatliff and P. J. Morris, “From Röntgen to Magnetic Resonance Imaging,” North Carolina Medical Journal, vol. 75, no. 2. North Carolina Institute of Medicine, pp. 111–113, Mar. 2014.

[2] M. M. Galloway, “Texture analysis using gray level run lengths,” Computer Graphics and Image Processing, vol. 4, no. 2. Elsevier BV, pp. 172–179, Jun. 1975.

[3] A. P. Pentland, “Fractal-Based Description of Natural Scenes,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. PAMI-6, no. 6. Institute of Electrical and Electronics Engineers (IEEE), pp. 661–674, Nov. 1984.

[4] M. Amadasun and R. King, “Textural features corresponding to textural properties,” IEEE Transactions on Systems, Man, and Cybernetics, vol. 19, no. 5. Institute of Electrical and Electronics Engineers (IEEE), pp. 1264–1274, 1989.

[5] G. Thibault, J. Angulo, and F. Meyer, “Advanced Statistical Matrices for Texture Characterization: Application to Cell Classification,” IEEE Transactions on Biomedical Engineering, vol. 61, no. 3. Institute of Electrical and Electronics Engineers (IEEE), pp. 630–637, Mar. 2014

[6] J. J. M. van Griethuysen et al., “Computational Radiomics System to Decode the Radiographic Phenotype,” Cancer Research, vol. 77, no. 21. American Association for Cancer Research (AACR), pp. e104–e107, Oct. 31, 2017

[7] “Welcome to pyradiomics documentation!” pyradiomics. https://pyradiomics.readthedocs.io/en/latest/index.html (accessed Jun. 18, 2023).

[8] H. Sung et al., “Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries,” CA: A Cancer Journal for Clinicians, vol. 71, no. 3. Wiley, pp. 209–249, Feb. 04, 2021.

[9] Adler I. Primary malignant growths of the lungs and bronchi. Longmans, Green, and Company; 1912.

[10] F. Bray, J. Ferlay, I. Soerjomataram, R. L. Siegel, L. A. Torre, and A. Jemal, “Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries,” CA: A Cancer Journal for Clinicians, vol. 68, no. 6. Wiley, pp. 394–424, Sep. 12, 2018.

[11] R. Doll, R. Peto, J. Boreham, and I. Sutherland, “Mortality in relation to smoking: 50 years’ observations on male British doctors,” BMJ, vol. 328, no. 7455. BMJ, p. 1519, Jun. 22, 2004

[12] H. Asamura et al., “IASLC Lung Cancer Staging Project: The New Database to Inform Revisions in the Ninth Edition of the TNM Classification of Lung Cancer,” Journal of Thoracic Oncology, vol. 18, no. 5. Elsevier BV, pp. 564–575, May 2023

[13] K. Zarogoulidis et al., “Treatment of non-small cell lung cancer (NSCLC),” Journal of Thoracic Disease, vol. 5, no. Suppl 4, pp. S389–S396, Sep. 2013

[14] C. Parmar, P. Grossmann, J. Bussink, P. Lambin, and H. J. W. L. Aerts, “Machine Learning methods for Quantitative Radiomic Biomarkers,” Scientific Reports, vol. 5, no. 1. Springer Science and Business Media LLC, Aug. 17, 2015

[15] C. Parmar et al., “Radiomic feature clusters and Prognostic Signatures specific for Lung and Head & Neck cancer,” Scientific Reports,

vol. 5, no. 1. Springer Science and Business Media LLC, Jun. 05, 2015

[16] W. Wu et al., "Exploratory Study to Identify Radiomics Classifiers for Lung Cancer Histology," Frontiers in Oncology, vol. 6. Frontiers Media SA, Mar. 30, 2016

[17] Lambrecht, J. Textural Analysis of Tumour Imaging: A Radiomics Approach, 2017

[18] A. Chaddad, C. Desrosiers, M. Toews, and B. Abdulkarim, "Predicting survival time of lung cancer patients using

[19] radiomic analysis," Oncotarget, vol. 8, no. 61. Impact Journals, LLC, pp. 104393–104407, Nov. 01, 2017

[20] C. Haarburger, P. Weitz, O. Rippel, and D. Merhof, "Image-Based Survival Prediction for Lung Cancer Patients Using CNNS," 2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019). IEEE, Apr. 2019

[21] Z. Shi et al., "Distributed radiomics as a signature validation study using the Personal Health Train infrastructure," Scientific Data, vol. 6, no. 1. Springer Science and Business Media LLC, Oct. 22, 2019.

[22] M. L. Welch et al., "Vulnerabilities of radiomic signature development: The need for safeguards," Radiotherapy and Oncology, vol. 130. Elsevier BV, pp. 2–9, Jan. 2019

[23] C. Haarburger et al., "Radiomic Feature Stability Analysis Based on Probabilistic Segmentations," 2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI). IEEE, Apr. 2020

[24] L. Ubaldi et al., "Strategies to develop radiomics and machine learning models for lung cancer stage and histology prediction using small data samples," Physica Medica, vol. 90. Elsevier BV, pp. 13–22, Oct. 2021

[25] S. Marinov et al., "Radiomics software for breast imaging optimization and simulation studies," Physica Medica, vol. 89. Elsevier BV, pp. 114–128, Sep. 2021

[26] G. Pasini, F. Bini, G. Russo, A. Comelli, F. Marinozzi, and A. Stefano, "matRadiomics: A Novel and Complete Radiomics Framework, from Image Visualization to Predictive Model," Journal of Imaging, vol. 8, no. 8. MDPI AG, p. 221, Aug. 18, 2022.

[27] M. Lei et al., "Benchmarking features from different radiomics toolkits / toolboxes using Image Biomarkers Standardization Initiative." arXiv, 2020.

[28] G. Pasini, F. Bini, G. Russo, A. Comelli, F. Marinozzi, and A. Stefano, "matRadiomics: A Novel and Complete Radiomics Framework, from Image Visualization to Predictive Model," Journal of Imaging, vol. 8, no. 8. MDPI AG, p. 221, Aug. 18, 2022.

[29] M. Lei et al., "Benchmarking Various Radiomic Toolkit Features While Applying the Image Biomarker Standardization Initiative toward Clinical Translation of Radiomic Analysis," Journal of Digital Imaging, vol. 34, no. 5. Springer Science and Business Media LLC, pp. 1156–1170, Sep. 20, 2021.

[30] A. Bettinelli et al., "A Novel Benchmarking Approach to Assess the Agreement among Radiomic Tools," Radiology, vol. 303, no. 3. Radiological Society of North America (RSNA), pp. 533–541, Jun. 2022.

[31] E. Pfaehler, A. Zwanenburg, J. R. de Jong, and R. Boellaard, "RaCaT: An open source and easy to use radiomics calculator tool," PLOS ONE, vol. 14, no. 2. Public Library of Science (PLoS), p. e0212223, Feb. 20, 2019.

[32] A. C. Lorena et al., "Comparing machine learning classifiers in potential distribution modelling," Expert Systems with Applications, vol. 38, no. 5. Elsevier BV, pp. 5268–5275, May 2011.

[33] H. J. W. L. Aerts et al., "Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach," Nature Communications, vol. 5, no. 1. Springer Science and Business Media LLC, Jun. 03, 2014.