Editorial, Radiation Protection.

Machine Learning and Radiomics in Nuclear Medicine and Molecular imaging: Part II.

Magdy M. Khalil, PhD[1, 2].

[1] School of Biotechnology, Badr University. [2] Medical Biophysics, Department of Physics, Faculty of Science, Helwan University, Cairo, Egypt.

ABSTRACT:

Diagnostic imaging modalities are undergoing a paradigm shift in technological advances and this has significantly impacted patient diagnosis and treatment. The introduction of machine learning and Radiomics in data analysis with capabilities of creating new clinical models has recently caught the attention of clinicians and scientists. Radiomics is a high throughput technology able to derive many imaging features from the diagnostic data while machine learning is a computer science discipline able to provide new forms of "electronic observer" able to mimic human tasks performed by radiologists and nuclear medicine physicians in daily routines. These technologies could be used individually or in combination to facilitate as well as solving issues associated with initial patient diagnosis, image processing, data analysis, stratification, prognosis and management. In the last decade, there was a rapidly growing interest in using Radiomics in nuclear medicine and molecular imaging providing several solutions in reducing the injected radio activities, reducing imaging time, lesion segmentation, diagnosis, and many other applications that could potentially serve or replace current practices. The goal of this part of the machine learning and Radiomics in nuclear medicine series is to introduce the reader to these new technologies and open avenues on current status, potential and future promises.

Key Words: Radiomics, supervised learning, feature extraction, model training, validation.

Corresponding Author: Magdy M. Khalil. **E-mail**: *magdykhalil@buc.edu.eg.*

INTRODUCTION:

The current status of many technologies in healthcare systems is the utilization of what is called “Big Data”. The analytical methods used to get most of this data are the topic of this interesting part of the series “Machine Learning and Radiomics in Nuclear Medicine and Molecular imaging”. There are several analytical tools that can be used to find out relationships between data derived from medical imaging and disease status including diagnosis, prognosis or response to treatments. Data analysis is carried out using statistical methods or machine learning approaches. However, it might be an interesting to define first these terms including Artificial Intelligence (AI), machine learning, neural networks and deep learning, ***(Figure 1)***.

AI means is to make a computer to mimic or simulate human behavior in some aspects. Machine learning is a subset of AI, and it consists of the techniques that enable computers to figure things out from the data and deliver AI applications. Neural networks or more specifically artificial neural networks (ANNs) - mimic the human brain through a set of algorithms. At a basic level, a neural network is comprised of four main components: inputs, weights, a bias and an output. Deep learning, meanwhile, is a subset of machine learning that enables computers to solve more complex problems. Within the framework of AI and machine learning, algorithms can be generally classified into supervised versus unsupervised methods. The former is based on training the algorithm using labeled datasets while the later depend on clustering data using some sort of similarity measures.

As outlined in the first part of this series, the field of Radiomics deals with extraction and mining of imaging data through different feature selection methods that vary in their hypothesis and then present them to machine learning methods that support inference. These models can then be used standalone or integrated within specific clinical practices. The convention in Radiomics is extraction of features, relating imaging features with molecular phenotypes and determining the relevance of radiologic features with pathological characteristics [(1)].

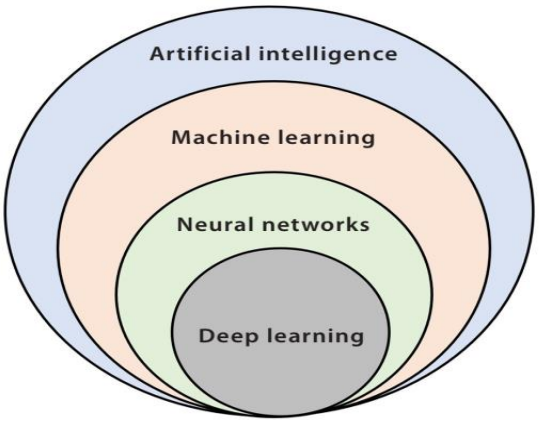

Figure (1): The relationship between artificial intelligence, machine learning, neural networks and deep learning.

The standard of care diagnostic modalities including x-ray computed tomography (CT), magnetic resonance imaging (MRI), positron emission tomography (PET) and single photon emission computed tomography (SPECT) are the main subject of using Radiomics as a growing powerful tool in decision making process. The values of radiomics stems from its strong complementarity to other clinic-pathological data obtained from patients and integrating these together to produce robust and potential clinical solutions.

Features selection: The two common approaches of data collection are handcrafted data extraction or radiomics based feature selection.

The hand-crafted approach employs region or volume of interest (ROI/VOI) where voxels data that belong to a certain lesion should be outlined.

However, this process is subject to bias due to the algorithm used in lesion segmentation. The calculation of radiomics feature is not only function of the algorithms but other variables can also influence this process including image preprocessing (i.e. discretization or data standardization), data filtration and/or smoothing levels, as well as reconstruction methods (e.g. iterative versus filtered back projection). How all these variables are integrated and implemented in routine clinical practices would be an enormous efforts placed on scientific community. There are also some variations that exist including mathematical definition, methodology, and implementation [(2)].

The other way around is to use the power of deep learning through the use of convolutional neural networks (CNN) that build up its own experience through several iterations over a labeled data set.

The labelled data set is collected beforehand with proper designation of clinical or clinic-pathological information that could be used as target for the learning algorithm. Once the model is built with high accuracy, it become then an easy process to serve as electronic observer in patient diagnosis, stratification, or prognosis. There are also some interests to correlate the imaging data with gene mutations or what is called radio genomics.

Challenges associated with deep learning are large data set required for model learning as well as incapability of the radiologist to interpret learning weights and parameters as opposed to conventional practice of using clinical or pathological information in diagnosis or treatments.

Modeling: The three basic steps of radiomics are feature extraction, data modeling and finally model validation, ***(Figure 2)***.

Model selection is not an easy task and many studies were reported using several algorithms investigating the relative performance using standard metrics for assessment. After model construction and identification of the fitting parameters, then the validation processes come into place to check the accuracy and verify model performance.

The proper standard is to validate the model using internal as well as external data to prove efficacy and escalate the situation to a more robust clinical trials.

Feature Extraction: The number of radiomics feature which could be extracted from imaging data are enormous and could exceed in some reports more than a thousand. There different approaches to extract features including filters and feature categories that are used to display the hidden information in medical imaging data. When the sample size is limited or the number of feature exceeds the number of observations, then overfitting is unavoidable data learning pitfalls that compromise generalizations, ***(Figure 3)***. Overfitting occurs when the model attempts to model the original variation as well as noise of the data and hence the use of those training parameters in predicting new cases would be of reduced performance.

To improve the overall performance of the model, researchers seek to incorporate as much information of the patient clinical data besides the imaging data. While this approach is data demanding, it provides a holistic version of a successful model that respects many aspects of patient clinical information.

It is then the power of the machine learning algorithm or the statistical model to prioritize which of those collective features would be better representative for final model formulation. Another maneuver to avoid overfitting is the exclusion of those feature that are less robust versus parameters of image acquisition and data processing.

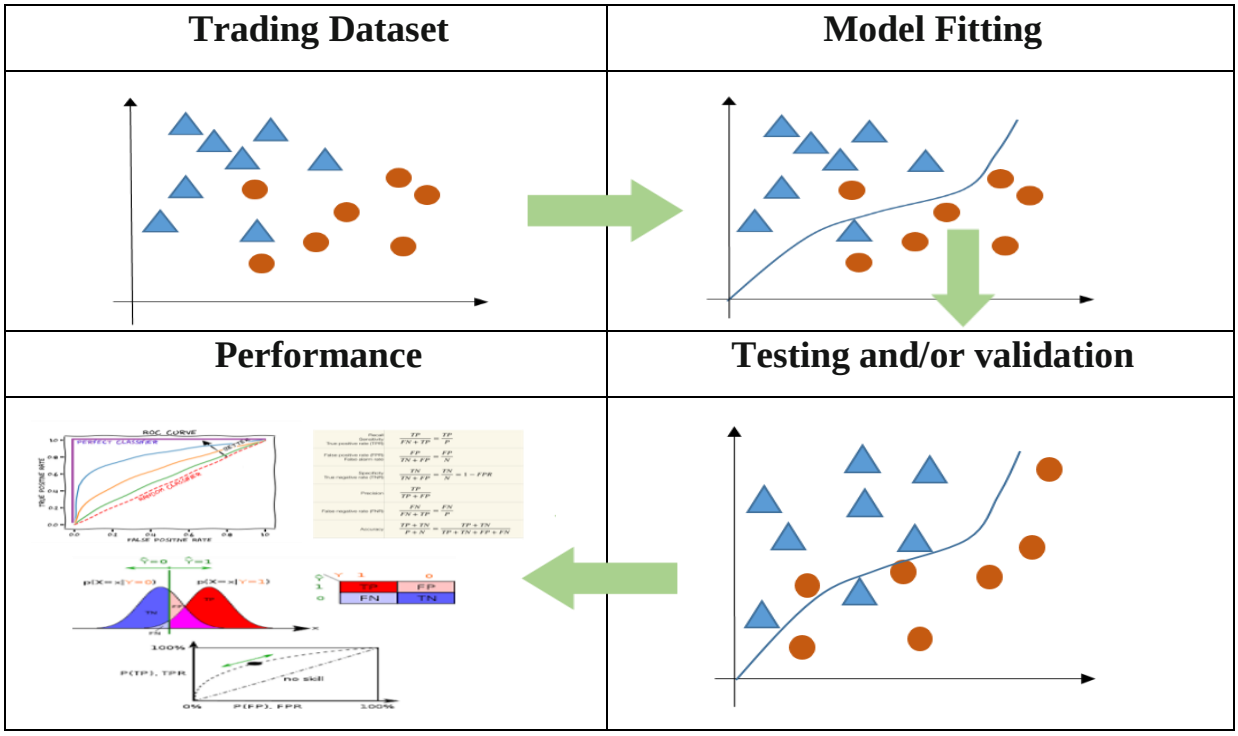

Figure (2): The flowchart of producing prediction models using statistical or machine learning.

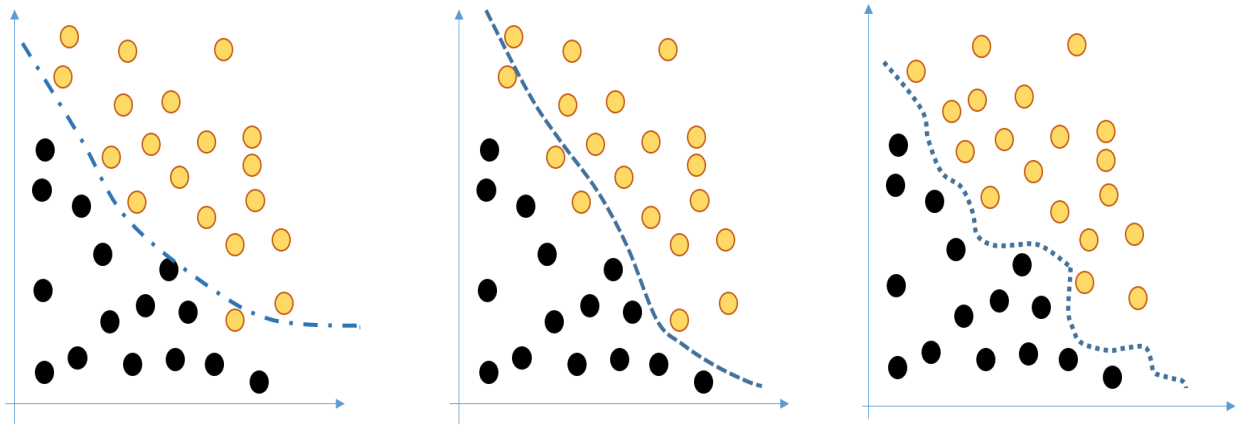

Figure (3): The example shows how overfitting, under fitting and optimal fitting can take place during model training and development.

One of the most successful approaches in reducing the number of features is the use of data reduction techniques such as principle component analysis (PCA) or unsupervised learning such as clustering. Statistical learning methods could also have some assumptions that should be followed including, for example, the linearity in the logit for continuous variables, independence of errors, and non-collinearity and lack of significant outliers. Bayesian networks also require feature discretization while deep learning should be coupled with proper network configuration.

Dimensionality reduction: As outlines above, the increasing numbers of features over the number of observation (i.e. sample size) pose numerical issues on model convergence and ultimately could result in overfitting.

Therefore, dimensionality reduction has become a common practice to reduce this undesired effect. The number of features is reduced to a smaller more representative feature that could improve classifier performance.

The dimension reduction can be done using different approaches such as collinearity analysis, collinearity analysis, algorithm-based feature selection, feature reproducibility analysis and cluster analysis [(3)].

Validation: Assessing the final performance of the model is indeed a critical component in radiomics implementation pipeline. Many studies were conducted focusing on internal validation using data that were collected under the same imaging scenario and same technical parameters.

This to some extent limits the validation and the full acceptance of the proposed model. Different and strong calls are being made to externally validate the radiomics models so that a more convincing and proofs are presented and documented. The most common internal validation techniques that can be encountered in the literature are k-fold, leave-one-out cross-validation, and holdout in addition to some more sophisticated techniques such as random subsampling, bootstrap cross-validation, and nested cross-validation [(3)].

Models

Supervised Learning models: The most common algorithms are k-nearest neighbors, naive Bayes, logistic regression, support vector machine, decision tree, random forest, neural networks and deep learning.

An overview of the models used in statistical machine learning would be simply presented for non-mathematical readers as well as new comers or beginners in the field of machine learning and radiomics data analysis.

Machine learning can be classified into two major categories including supervised and unsupervised learning. In supervised learning, there are two major types of problems: *regression and classification*. The former deals with continuous data for prediction while the latter handles as name implies two or more classes in the learning process. When the algorithm is used to distinct two groups it is then called binary classification but when the groups are more than two it is called multiclass classification.

Linear Methods: Linear regression and logistic regression are the most two famous algorithms in linear models such that the former assumes a linear relationship between the input variables and output while logistic regression transforms the output using the logistic function. An important assumption of linear regression is that the features should be uncorrelated and the true relationship should be linear otherwise an under fitting will take place.

The positive remarks of linear regression are simplicity, interpretability and less common to produce data fitting.

The drawbacks are their limitations to model nonlinear relationship which might exist in several real data problems. Logistic regression enjoys the feature of separating classes of linear behavior and fails when classes are overlapping.

k-nearest neighbors (KNN): k-nearest neighbors (KNN) is a non-parametric classification algorithm that can be used in classification and regression. The input consists of the k-closest training examples in a dataset. This machine learning algorithm classifies the unknown objects or instances by assigning them to the similar objects of the classes based on the number of neighbors. It can be intuitively seen as "similar birds fly together" since similar values are grouped together in the same class. In KNN, K is the number of nearest neighbors. The number of neighbors is the most critical to be chosen. K is generally an odd number if the number of classes is 2. When K=1, then the algorithm is known as the nearest neighbor algorithm. ***Figure (4)*** is a simplified explanation of how KNN works in data classification.

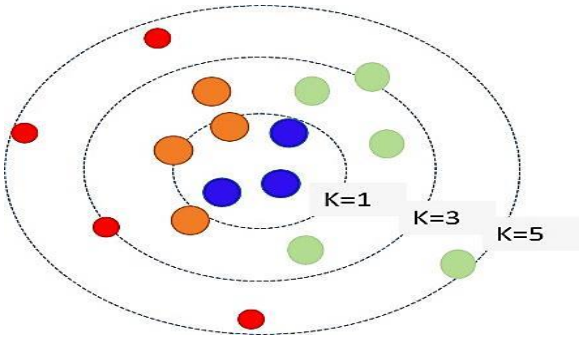

Figure (4): K-nearest neighbor in data classification.

In contrary to linear methods, KNN is highly flexible and adept at learning more complex, nonlinear relationships. The method performance gets slower at increasing number of features and sample size. Although KNN methods are simple and interpretable, a large number of K tends to under fit the data but few number of K tends to over fit the data.

Tree-based methods: Based on data labels, the algorithm classifies or stratifies all instances into many regions based on information contained within a given training data, ***(Figure 5)***.

Upon the end of classification, the data are then segmented into group of classes or range of predicted values based on whether the clinical question is classification or regression respectively. There is more than one variant based on the number of iterations or run over the entire or subset of data. Random forest is one more common variant that sampling not only the instances but also the predictors.

In this special tree based approach, random samples are selected and the prediction process is made based not on all but a subset of predictors, ***(Figure 6)***.

The number of predictors we consider for each split is usually the square root of the total number of predictors.

Boosting: In boosting, it is used to create multiple trees like in bagging but to build the trees sequentially, using what the AI learned from the previous tree to improve results on the subsequent tree. Each tree is kept pretty shallow, with only a few decision splits, and the learning occurs slowly, tree by tree.

The caveats of decision trees are their performance is weakened when many rules are needed to make predictions as well as the number of features are increased.

Support vector machines: Support vector machines look for the maximum boundary separation based on the data labels. It is not necessary that all points will belong to certain class but the separation between classes is maximized as much as possible. The algorithm uses linear as well as non-linear kernels to perform this task and create space hyperplanes among classes ***(Figure 7)***.

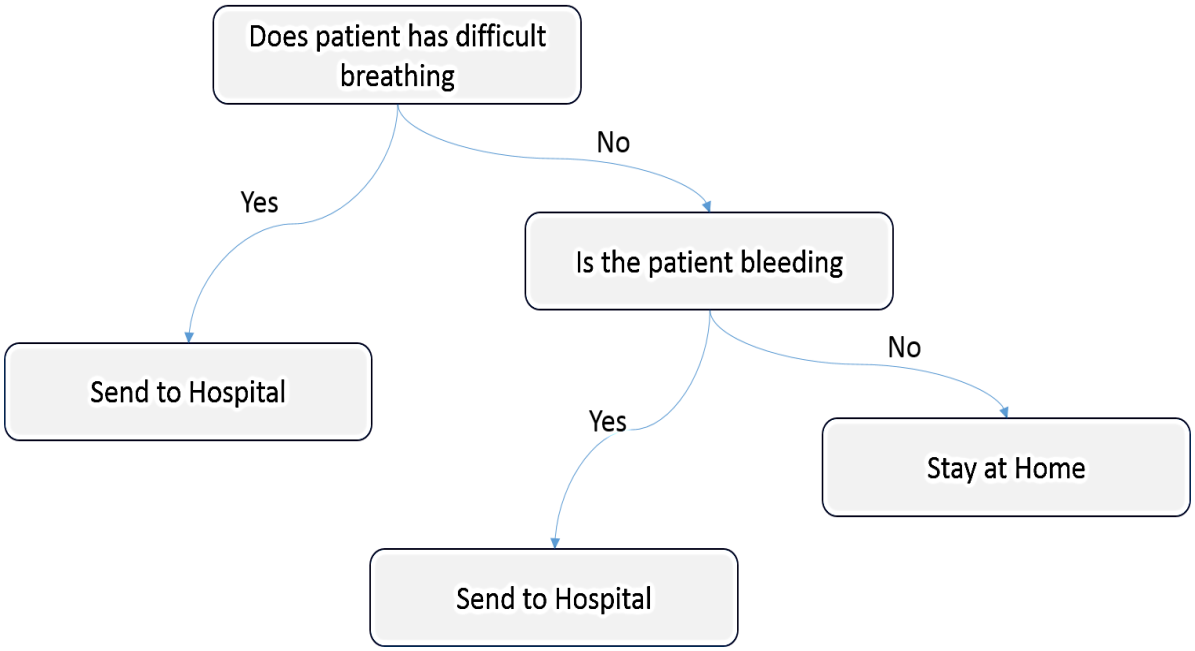

Figure (5): An example showing decision tree strategy in data classification and building up the decision classes or trees.

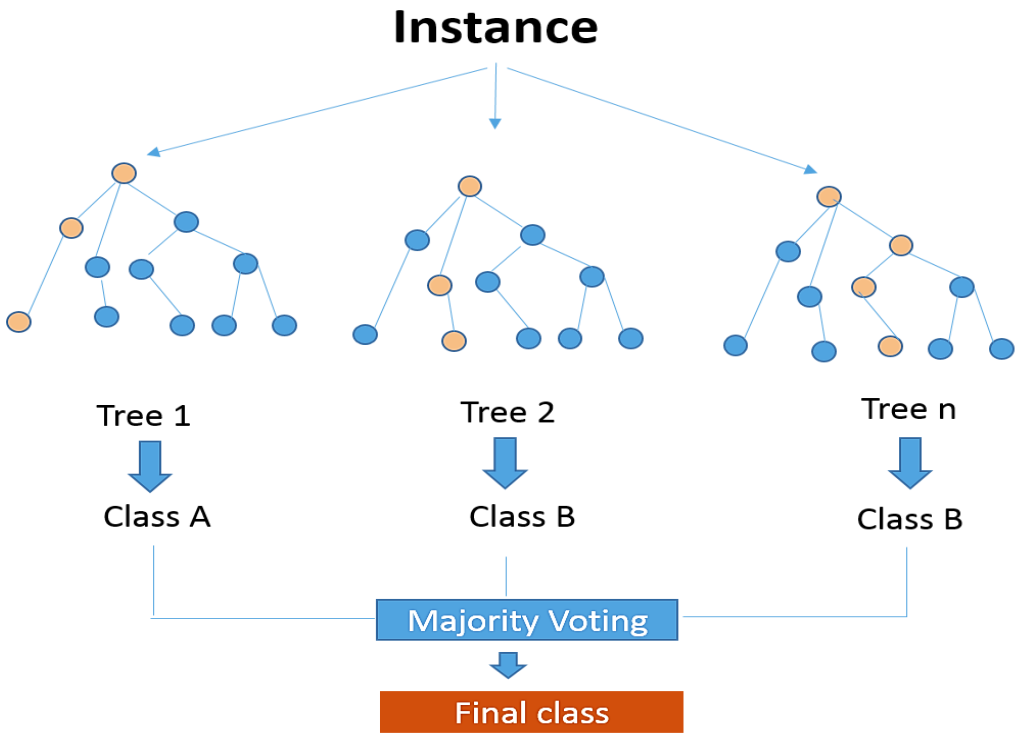

Figure (6): Random forest training scheme. In this special tree based approach, random samples are selected and the prediction process is made based not on all but a subset of predictors.

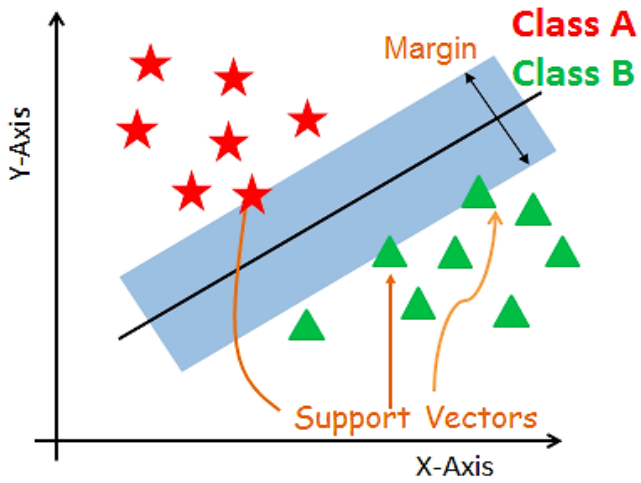

Figure (7): Demonstration of how support vector machines work and the process of creating hyperplanes between classes. From www.datacamp.com

Neural network: In analogous to human brain and the connectivity among neurons, neural network consists of input, hidden and output layers of nodes. The data "features" are initially fed into the input layer and through several steps of iteration the output layer is trained to predict the label. The hidden layers are tuned through this iterative process to successfully maximize the accuracy. In convolutional neural networks (CNN), the input is multidimensional with a shape that depends on number of inputs, height, width and channels. After passing through a convolutional layer, the image is transformed into a feature map, also called an activation map. Convolutional layers convolve the input and pass its result to the next layer. This is again similar to the activation and response of neurons of the visual cortex of the human brain.

Unsupervised learning: As outlined above, supervised learning needs labeled data but this is not necessarily found in all real life scenarios especially in medical diagnosis and clinical imaging. The other approaches that can be potentially used to treat data absent of labels or target output is unsupervised learning such as dimensionality reduction techniques, k-means clustering, hierarchical clustering and deep learning. I will deal with these methods in future editorial releases.

CONCLUSIONS:

Special attention must be carried out during feature extraction as well as model selection and training. Model validation is also a crucial step that must be carefully designed especially with inclusion of external data sets. There several supervised learning models available and it is the mission of the researchers to understand the weaknesses and strengths of the employed model. Supervised models require labelled data so that one can generate a predictive model for future application on unseen data. Hence the process of generalization must be ensured during model training and parameter tuning so that under – or overfitting is avoided.

All supervised learning methods have caveats and thus one should bear in mind model characteristics in the selection process and deployment. It has been common that comparative studies are conducted to look at the most robust and accurate approach for a given technical or clinical question. Unsupervised statistical learning is also interesting and potential candidate to understand data structure when data labels are not available. It is the topic of the next editorial issue.

Disclosure: The author discloses that he has not any particular interest or royalties associated with this article.

REFERENCES:

1. ***Capobianco E, Deng J. Radiomics at a Glance:*** A Few Lessons Learned from Learning Approaches. Cancers (Basel); 2020.

2. ***Guiot J, Vaidyanathan A, Deprez L, et al,.*** A review in radiomics: Making personalized medicine a reality via routine imaging. Med. Res. Rev; 2021.

3. ***Kocak B, Durmaz ES, Ates E, Kilickesmez O.*** Radiomics with artificial intelligence: a practical guide for beginners. Diagn Interv. Radiol. 25(6):485-95; 2019.